

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
20 February 2003 (20.02.2003)

PCT

(10) International Publication Number
WO 03/014297 A2

- (51) International Patent Classification⁷: **C12N** MA 01773 (US). **COEFFET-LEGAL, Marie Françoise** [FR/US]; 119 Lake Street, Arlington, MA 02474 (US).
- (21) International Application Number: PCT/US02/24310
- (22) International Filing Date: 31 July 2002 (31.07.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/310,385 6 August 2001 (06.08.2001) US
PCT/US01/32354 17 October 2001 (17.10.2001) US
60/379,866 10 May 2002 (10.05.2002) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US PCT/US01/32354 (CIP)
Filed on 17 October 2001 (17.10.2001)
- (71) Applicant (for all designated States except US): **CUBIST PHARMACEUTICALS, INC.** [US/US]; 65 Hayden Avenue, Lexington, MA 02421 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **MIAO, Vivian, Pak, Woon** [CA/CA]; 13750 31 Avenue, Surrey, British Columbia V4P 2B7 (CA). **BRIAN, Paul** [GB/US]; 1503 Stearns Hill Road, Waltham, MA 02451 (US). **BALTZ, Richard, H.** [US/US]; 115 Chestnut Circle, Lincoln,
- (74) Agents: **HALEY, James, F.** et al.; Fish & Neave, 1251 Avenue of the Americas, New York, NY 10020 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPOSITIONS AND METHODS RELATING TO THE DAPTOMYCIN BIOSYNTHETIC GENE CLUSTER

(57) Abstract: The invention provides nucleic acid molecules comprising all or a part of a daptomycin biosynthetic gene cluster. The daptomycin biosynthetic gene cluster may be derived from *Streptomyces*, preferably from *S. roseosporus*. The invention also provides other nucleic acid molecules from *S. roseosporus*. The invention further provides polypeptides encoded by the nucleic acid molecules, antibodies that specifically bind to the polypeptides, and methods of using the nucleic acid molecules, polypeptides and antibodies to produce daptomycin and other compounds.



WO 03/014297 A2

COMPOSITIONS AND METHODS RELATING
TO THE DAPTOMYCIN BIOSYNTHETIC GENE CLUSTER

Background of the Invention

[0001] Bacteria, including actinomycetes, and fungi
5 synthesize a diverse array of low molecular weight
peptide and polyketide compounds (approx. 2-48 residues
in length). The biosynthesis of these compounds is
catalyzed by non-ribosomal peptide synthetases (NRPSs)
and by polyketide syntheses (PKSs). The NRPS process,
10 which does not involve ribosome-mediated RNA translation
according to the genetic code, is capable of producing
peptides that exhibit enormous structural diversity,
compared to peptides translated from RNA templates by
ribosomes. These include the incorporation of D- and L-
15 amino acids and hydroxy acids; variations within the
peptide backbone which form linear, cyclic or branched
cyclic structures; and additional structural
modifications, including oxidation, acylation,
glycosylation, N-methylation and heterocyclic ring
20 formation. Many non-ribosomally synthesized peptides
have been found which have useful pharmacological (e.g.,
antibiotic, antiviral, antifungal, antiparasitic,
siderophore, cytostatic, immunosuppressive, anti-

- 2 -

cholesterolemic and anticancer), agrochemical or physicochemical (e.g., biosurfactant) properties.

[0002] Non-ribosomally synthesized peptides are assembled by large (e.g., about 200-2000 kDa),
5 multifunctional NRPS enzyme complexes comprising one or more subunits. Examples include daptomycin, vancomycin, echinocandin and cyclosporin. Likewise, polyketides are assembled by large multifunctional PKS enzyme complexes comprising one or more subunits. Examples include
10 erythromycin, tylosin, monensin and avermectin. In some cases, complex molecules can be synthesized by mixed PKS/NRPS systems. Examples include rapamycin, bleomycin and epothilone.

[0003] An NRPS usually consists of one or more open
15 reading frames that make up an NRPS complex. The NRPS complex acts as a protein template, comprising a series of protein biosynthetic units configured to bind and activate specific building block substrates and to catalyze peptide chain formation and elongation. (See,
20 e.g., Konz and Marahiel, *Chem. Biol.*, 6, pp. 39-48 (1999) and references cited therein; von Döhren et al., *Chem. Biol.*, 6, pp. 273-279, (1999) and references cited therein; and Cane and Walsh, *Chem. Biol.*, 6, pp. 319-325, (1999), and references cited therein - each hereby
25 incorporated by reference in its entirety). Each NRPS or NRPS subunit comprises one or modules. A "module" is defined as the catalytic unit that incorporates a single building block (e.g., an amino acid) into the growing peptide chain. The order and specificity of the
30 biosynthetic modules that form the NRPS protein template dictates the sequence and structure of the ultimate peptide products.

- 3 -

[0004] Each module of an NRPS acts as a semi-autonomous active site containing discrete, folded protein domains responsible for catalyzing specific reactions required for peptide chain elongation. A minimal module (in a single module complex) consists of at least two core domains: 1) an adenylation domain responsible for activating an amino acid (or, occasionally, a hydroxy acid); and 2) a thiolation or acyl carrier domain responsible for transferring activated intermediates to an enzyme-bound pantetheine cofactor. Most modules also contain 3) a condensation domain responsible for catalyzing peptide bond formation between activated intermediates. See Figure 9. Supplementing these three core domains are a variable number of additional domains which can mediate, e.g., N-methylation (M or methylation domain) and L- to D-conversion (E or epimerization domain) of a bound amino acid intermediate, and heterocyclic ring formation (Cy or cyclization domain). The domains are usually characterized by specific amino acid motifs or features. It is the combination of such auxiliary domains acting locally on tethered intermediates within nearby modules that contributes to the enormous structural and functional diversity of the mature peptide products assembled by NRPS and mixed NRPS/PKS enzyme complexes.

[0005] The adenylation domain of each minimal module catalyzes the specific recognition and activation of a cognate amino acid. In this early step of non-ribosomal peptide biosynthesis, the cognate amino acid of each NRPS module is bound to the adenylation domain and activated as an unstable acyl adenylate (with concomitant ATP-hydrolysis). See, e.g., Stachelhaus et al., *Chem. Biol.* 6: 493-505 (1999) and Challis et al., *Chem. Biol.* 7: 211-

- 4 -

224 (2000), each incorporated herein by reference in its entirety. In most NRPS modules, the acyl adenylate intermediate is next transferred to the T (thiolation) domain (also referred to as a peptidyl carrier protein or PCP domain) of the module where it is converted to a thioester intermediate and tethered via a transthioation reaction to a covalently bound enzyme cofactor (4'-phosphopantetheinyl (4'-PP) intermediate). Modules responsible for incorporating D-configured or N-methylated amino acids may have extra modifying domains which, in several NRPSs studied, are located between the A and T domains.

[0006] The enzyme-bound intermediates in each module are then assembled into the peptide product by stepwise condensation reactions involving transfer of the thioester-activated carboxyl group of one residue in one module to, e.g., the adjacent amino group of the next amino acid in the next module while the intermediates remain linked covalently to the NRPS. Each condensation reaction is catalyzed by a condensation (C) domain which is usually positioned between two minimal modules. The number of condensation domains in a NRPS generally corresponds to the number of peptide bonds present in the final (linear) peptide. An extra C domain has been found in several NRPSs (e.g., at the amino terminus of cyclosporin synthetase and the carboxyl terminus of rapamycin; see, e.g., Konz and Marahiel, *supra*) that has been proposed to be involved in peptide chain termination and cyclization reactions. Many other NRPS complexes, however, release the full length chain in a reaction catalyzed by a C-terminal thioesterase (Te) domain (of approximately 28K-35K relative molecular weight).

- 5 -

[0007] Thioesterase domains of most NRPS complexes use a catalytic triad (similar to that of the well-known chymotrypsin mechanism) which includes a conserved serine (less often a cysteine or aspartate) residue in a conserved three-dimensional configuration relative to a histidine and an acidic residue. See, e.g. V. De Crecy-Lagard in *Comprehensive Natural Products Chemistry*, Volume 4, ed. J.W. Kelly (New York: Elsevier), 1999, pp. 221-238, each incorporated herein by reference in its entirety. Thioester cleavage is a two step process. In the first (acylation) step, the full length peptide chain is transferred from the thiol tethered enzyme intermediate in the thiolation domain (see above) to the conserved serine residue in the Te domain, forming an acyl-O-Te ester intermediate. In the second (deacylation) step, the Te domain serine ester intermediate is either hydrolyzed (thereby releasing a linear, full length product) or undergoes cyclization, depending on whether the ester intermediate is attacked by water (hydrolysis) or by an activated intramolecular nucleophile (cyclization).

[0008] Sequence comparisons of C-terminal thioesterase domains from diverse members of the NRPS superfamily have revealed a conserved motif comprising the serine catalytic residue (GX SXG motif), often followed by an aspartic acid residue about 25 amino acids downstream from the conserved serine residue. A second type of thioesterase, a free thioesterase enzyme, is known to participate in the biosynthesis of some peptide and polyketide secondary metabolites. See e.g., Schneider and Marahiel, *Arch. Microbiol.*, 169, pp. 404-410 (1998), and Butler et al., *Chem. Biol.*, 6, pp. 87-292 (1999), each incorporated herein by reference in its entirety.

- 6 -

These thioesterases are often required for efficient natural product synthesis. Butler et al. have postulated that the free thioesterase found in the polyketide tylosin gene cluster -- which is required for efficient tylosin production -- may be involved in editing and proofreading functions.

[0009] The modular organization of the NRPS multienzyme complex is mirrored at the level of the genomic DNA encoding the modules. The organization and DNA sequences of the genes encoding several different NRPSs have been studied. (See, e.g., Marahiel, *Chem. Biol.*, 4, pp. 561-567 (1997), incorporated herein by reference in its entirety). Conserved sequences characterizing particular NRPS functional domains have been identified by comparing NRPS sequences derived from many diverse organisms and those conserved sequence motifs have been used to design probes useful for identifying and isolating new NRPS genes and modules.

[0010] The modular structures of PKS and NRPS enzyme complexes can be exploited to engineer novel enzymes having new specificities by changing the numbers and positions of the modules at the DNA level by genetic engineering and recombination *in vivo*. Functional hybrid NRPSs have been constructed, for example, based on whole-module fusions. See, e.g., Gokhale et al., *Science*, 284, pp. 482-485 (1999); Mootz et al., *Proc. Natl. Acad. Sci. U.S.A.*, 97, pp. 5848-5853 (2000), incorporated herein by reference in their entirety. Recombinant techniques may be used to successfully swap domains originating from a heterologous PKS or NRPS complex. See, e.g., Schneider et al., *Mol. Gen. Genet.*, 257, pp. 308-318 (1998); McDaniel et al., *Proc. Natl. Acad. Sci. U.S.A.*, 96, pp. 1846-1851 (1999); United States Patent Nos. 5,652,116 and

- 7 -

5,795,738; and International Publication WO 00/56896; incorporated herein by reference in their entirety.

[0011] Engineering a new substrate specificity within a module by altering residues which form the substrate binding pocket of the adenylation domain has also been described. See, e.g., Cane and Walsh, *Chem. Biol.*, 6, 319-325 (1999); Stachelhaus et al., *Chem. Biol.*, 6, 493-505 (1999); and WO 00/52152; each incorporated herein by reference in its entirety. By comparing the sequence of the *B. subtilis* peptide synthetase GrsA adenylation domain (PheA) (whose structure is known) with sequences of 160 other adenylation domains from pro- and eukaryotic NRPSs, for example, Stachelhaus et al. (*supra*) and Challis et al., *Chem. Biol.*, 7, pp. 211-224 (2000) defined adenylation (A) domain signature sequences (analogous to codons of the genetic code) for a variety of amino acid substrates. From the collection of those signature sequences, a putative NRPS selectivity-conferring code (with degeneracies like the genetic code) was formulated.

[0012] The ability to engineer NRPSs having new modular template structures and new substrate specificities by adding, deleting or exchanging modules (or by adding, deleting or exchanging domains within one or more modules) will enable the production of novel peptides having altered and potentially advantageous properties. A combinatorial library comprising over 50 novel polyketides, for example, was prepared by systematically modifying the PKS that synthesizes an erythromycin precursor (DEBS) by substituting counterpart sequences from the rapamycin PKS (which encodes alternative substrate specificities). See, e.g., WO

- 8 -

00/63361 and McDaniel et al., (1999), *supra*, each incorporated herein by reference in its entirety.

[0013] A number of bacteria that produce antibiotics and other potentially toxic compounds synthesize ATP-binding cassette (ABC) transporters. ABC transporters use proton-dependent transmembrane electrochemical potential to export toxic cellular metabolites such as antibiotics, and to import materials from the environment, e.g. iron or other metals. There are three types of ABC transporters and genes encoding pumps responsible for antibiotic resistance, and they are often linked to the biosynthetic cluster in antibiotic producer organisms (e.g. actinorhodin resistance in *Streptomyces coelicolor*). See, e.g., Mendez et al., *FEMS Microbiol. Lett.* 158: 1-8 (1998), herein incorporated by reference. All have ATP-binding regions that include Walker A and B motifs. *Id.* Type I systems involve separate genes for a hydrophilic ATP-binding domain and a hydrophobic integral membrane domain. Type III systems involve a single gene encoding a protein with a hydrophobic N-terminus and a hydrophilic, ATP-binding C-terminus. Type II transporters have no hydrophobic domain, and two sets of Walker motifs, in the order A:B:A:B.

[0014] The *Streptomyces glaucescens* genes, StrV (PIR Accession No. S57561) and StrW (PIR Accession No. S57562) encode type III transporters associated with resistance to streptomycin-related compounds. Both genes are within a 5'-hydroxystreptomycin antibiotic biosynthetic gene cluster. See, e.g., Beyer et al., *Mol. Gen. Genet.* 250: 775-84 (1996), herein incorporated by reference. Resistance to doxorubicin and related antibiotics is conferred by two type I transporters in *Streptomyces peucetius*, which are encoded by *drmA* and *drmB*. See,

- 9 -

e.g., Guifoile et al., *Proc. Natl. Acad. Sci. USA* 88:8553-57 (1991), herein incorporated by reference. Further, homologs of *drxAB* isolated from *Streptomyces rochei* confer multidrug resistance when expressed under control of the actinorhodin PKS promoter in *S. lividans*. See, e.g., Fernandez-Moreno et al., *J. Bacteriol.* 179: 6929-36 (1998), herein incorporated by reference.

[0015] Daptomycin (described by R.H. Baltz in *Biotechnology of Antibiotics*, 2nd Ed., ed. W.R. Strohl (New York: Marcel Dekker, Inc.), 1997, pp. 415-435) is an example of a non-ribosomally synthesized peptide made by a NRPS. Daptomycin, also known as LY146032, is a cyclic lipopeptide antibiotic that is produced by the fermentation of *Streptomyces roseosporus*. Daptomycin is a member of the factor A-21978C type antibiotics of *S. roseosporus* and comprises an n-decanoyl side chain linked via a three-amino acid chain to the N-terminal tryptophan of a cyclic 10-amino acid peptide. The compound is being developed in a variety of formulations to treat serious infections for which therapeutic options are limited, such as infections caused by bacteria including, but not limited to, methicillin resistant *Staphylococcus aureus*, vancomycin resistant enterococci, glycopeptide intermediary susceptible *Staphylococcus aureus*, coagulase-negative staphylococci, and penicillin-resistant *Streptococcus pneumoniae*. See, e.g., Tally et al., *Exp. Opin. Invest. Drugs* 8:1223-1238, 1999. The antibiotic action of daptomycin against Gram-positive bacteria has been attributed to its ability to interfere with membrane potential and to inhibit lipoteichoic acid synthesis.

[0016] Identification of the genes encoding the proteins involved in the daptomycin biosynthetic pathway,

- 10 -

including the daptomycin NRPS, will provide a first step in producing modified *Streptomyces roseosporus* as well as other host strains which can produce an improved antibiotic (for example, having greater potency); which
5 can produce natural or new antibiotics in increased quantities; or which can produce other peptide products having useful biological properties. Compositions and methods relating to the *Streptomyces roseosporus* daptomycin biosynthetic gene cluster, including isolated
10 nucleic acids and isolated proteins, are described in United States Provisional Applications 60/240,879, filed October 17, 2000; 60/272,207, filed February 28, 2001; and 60/310,385, filed August 8, 2001; all of which are hereby incorporated by reference in its entirety.

15 **[0017]** It would be advantageous, moreover, to identify the genetic and modular organization of the *Streptomyces roseosporus* daptomycin biosynthetic gene cluster in order to construct full length daptomycin NRPS templates for expression in *Streptomyces roseosporus* and in
20 heterologous hosts. In particular, it would be advantageous to know whether the daptomycin gene cluster comprises a thioesterase (Te) domain. If so, that Te domain could be isolated and used to catalyze peptide chain termination in new NRPS modules and templates by
25 expression as a fusion or as a free peptide. See, e.g., de Ferra et al., *J. Biol. Chem.*, 272, pp. 25304-25309 (1997); Guenzi et al., *J. Biol. Chem.*, 273, pp. 14403-14410 (1998); and Trauger et al., *Nature*, 407, pp. 215-218 (2000); each incorporated herein by reference in its
30 entirety. It would also be advantageous to identify other nucleic acid molecules that encode polypeptides involved in daptomycin biosynthesis. These include, without limitation, enzymes involved in attaching a lipid

- 11 -

tail to the peptide domain of daptomycin, polypeptides that regulate antibiotic resistance and ABC transporters. Polypeptides that regulate antibiotic resistance and ABC transporters could be used to confer resistance or
5 increase, modify or decrease resistance of a bacteria to daptomycin and related antibiotics. Polypeptides involved in antibiotic resistance would also be useful to determine bacterial mechanisms of resistance, so that daptomycin and related antibiotics can be modified to
10 make them more potent against resistant bacteria.

SUMMARY OF THE INVENTION

[0018] The instant invention addresses these problems by providing a nucleic acid molecule that comprises all or a part of a daptomycin biosynthetic gene cluster,
15 preferably one from *S. roseosporus*. The nucleic acid molecule may encode DptA, DptBC or DptD or may comprise one or more of the *dptA*, *dptBC* or *dptD* genes from the daptomycin biosynthetic gene cluster of *S. roseosporus*.

[0019] The instant invention also provides nucleic acid molecules encoding a free thioesterase and an
20 integral thioesterase from a daptomycin biosynthetic gene cluster. The nucleic acid molecule may encode DptH or the thioesterase domain from DptD, or may comprise the *dptH* or *dptD* gene from the daptomycin biosynthetic gene
25 cluster.

[0020] Another object of the invention is to provide a nucleic acid molecule comprising a DNA sequence from a bacterial artificial chromosome comprising a nucleic acid sequence from *S. roseosporus*. The nucleic acid molecule
30 preferably comprises a *S. roseosporus* nucleic acid sequence from any one of bacterial artificial chromosome (BAC) clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or B12:03A05. (Of these, only B12:03A05 has

- 12 -

been deposited; ATCC Deposit No. PTA-3141, deposited March 1, 2001). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide. In another preferred embodiment, the nucleic acid molecule encodes a polypeptide that is involved in daptomycin biosynthesis, such as a *dptA*, *dptBC*, *dptD*, *dptE*, *dptF*, *dptH*, an ABC transporter, or a polypeptide that regulates antibiotic resistance, as described herein.

[0021] The invention also provides selectively hybridizing or homologous nucleic acid molecules of the above-described nucleic acid molecules. The invention further provides allelic variants and parts thereof. The invention further provides nucleic acid molecules that comprise one or more expression control sequences controlling the transcription of the above-described nucleic acid molecules. The expression control sequence may be derived from the expression control sequences of the daptomycin biosynthetic gene cluster or may be derived from a heterologous nucleic acid sequence.

[0022] In another embodiment, the invention provides a nucleic acid molecule comprising one or more expression control sequences from a gene comprising a nucleic acid sequence that encodes a thioesterase and/or a daptomycin NRPS from the daptomycin biosynthetic gene cluster. Preferably, the nucleic acid molecule comprises a part or all of the expression control sequences of the daptomycin NRPS or *dptH*.

[0023] Another object of the invention is to provide a vector and/or host cell comprising one or more of the above-described nucleic acid molecules. In a preferred embodiment, the vector and/or host cell comprises a nucleic acid molecule encoding all or part of DptA, DptBC, DptD, DptE, DptF and/or DptH, or all or part of a

- 13 -

BAC clone described above. A host cell may comprise all or a part of an NRPS or PKS, such as a daptomycin NRPS. The host cell may further comprise one or more thioesterases.

5 **[0024]** Another object of the invention is to provide a polypeptide derived from the daptomycin biosynthetic gene cluster, preferably a polypeptide from the daptomycin biosynthetic gene cluster of *S. roseosporus*. The polypeptide may be DptA, DptBC or DptD.

10 **[0025]** The invention also provides a polypeptide derived from an integral or free thioesterase, preferably one derived from a daptomycin biosynthetic gene cluster of *S. roseosporus*. In a preferred embodiment, the polypeptide is derived from thioesterase. The
15 polypeptide may be derived from DptH or the thioesterase domain of DptD.

[0026] The invention also provides a polypeptide encoded by a nucleic acid molecule of any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04,
20 B12:20C09 or B12:03A05. These polypeptides include, among others, enzymes involved in attaching a lipid tail to the peptide domain of daptomycin, polypeptides that regulate antibiotic resistance and ABC transporters.

[0027] Another object of the invention is to provide
25 fragments of the polypeptides described above. In one embodiment, the fragment comprises at least one domain or module, as defined herein. In another embodiment, the fragment comprises at least one epitope of the polypeptide.

30 **[0028]** Another object of the invention is to provide polypeptides that are mutant proteins, fusion proteins, homologous proteins or allelic variants of the daptomycin NRPS polypeptides, thioesterases and polypeptides encoded

- 14 -

by the nucleic acid molecules of the BAC clones provided herein.

[0029] The invention also provides an antibody that specifically binds to a polypeptide of a daptomycin NRPS, a thioesterase polypeptide of a daptomycin biosynthetic gene cluster or a polypeptide encoded by a nucleic acid molecule from any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or B12:03A05. The invention also provides an antibody that can bind to a fragment, polypeptide mutant, a fusion protein, a polypeptide encoded by an allelic variant or a homologous protein of any one of the above-described polypeptides or proteins. The antibodies may be used to detect the presence or amount of a polypeptide of the instant invention or to inhibit or activate an activity of a polypeptide.

[0030] Another objective of the instant invention is to provide a method for recombinantly producing a polypeptide using a nucleic acid molecule described herein by introducing a nucleic acid molecule into a host cell and expressing the polypeptide.

[0031] The instant invention also provides a method for using the nucleic acid molecules of the instant invention to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein.

[0032] The nucleic acid molecules and polypeptides are useful for, for example, the biosynthesis and production of natural products and the engineered biosynthesis of new compounds. The daptomycin NRPS and/or thioesterases may be used to produce daptomycin and other lipopeptides, including both naturally-occurring and novel compounds. The polypeptides may be used *in vitro* for the production

- 15 -

of cyclic or non-cyclic lipopeptides, as well as other compounds produced by non-ribosomal peptide synthesis. Alternatively, a nucleic acid molecule of the invention may be introduced and expressed in a host cell, and the host cell may then be used to produce lipopeptides and other compounds produced by non-ribosomal peptide synthesis.

[0033] Another objective of the invention is to provide a novel gene cluster that can produce novel compounds by non-ribosomal peptide synthesis. A novel gene cluster may be obtained by altering nucleotides of the daptomycin biosynthetic gene cluster, particularly by altering nucleotides, domains or modules of the daptomycin NRPS, to make new polypeptides that are involved in non-ribosomal peptide synthesis. In this manner, different amino acids may be incorporated into a peptide produced by non-ribosomal peptide synthesis than the peptide produced by a naturally-occurring polypeptide. The invention also encompasses the compounds produced by the methods described herein.

[0034] Another objective of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the instant invention. The records of the computer readable means can be accessed for reading and display of sequences and for comparison, alignment and ordering of the sequences of the invention to other sequences.

BRIEF DESCRIPTION OF THE DRAWINGS

[0035] Figure 1 is a schematic diagram of methods in which daptomycin NRPS genes can be manipulated to alter gene expression or expression of the encoded proteins.

[0036] Figure 2A is a schematic diagram of BAC clone B12:03A05. The diagram shows a 90 kb region, referred to

- 16 -

as the 90 kb fragment, an approximately 13 kb region, referred to herein as the SP6 fragment, and an approximately 25-28 kb region, referred to herein as the GTC2 fragment. SEQ ID NO: 1 shows the nucleic acid sequence of the 90 kb fragment. SEQ ID NO: 103 shows the nucleic acid sequence of the SP6 fragment. The SP6 fragment flanks the 90 kb fragment at the left. The GTC2 fragment flanks the 90 kb fragment on the right. SEQ ID NO: 104 shows the nucleic acid sequence of the GTC2 fragment.

[0037] Figure 2B shows a schematic diagram of the 90 kb fragment. There are 38 open reading frames (ORFs), which are nucleic acid sequences that encode polypeptides, in the region of the daptomycin biosynthetic gene cluster.

[0038] Figure 2C shows a schematic diagram of the SP6 fragment. There are 9 ORFs in the SP6 fragment. See Table 5 for the amino acid and nucleic acid sequence identifiers for the ORFs of the 90 kb and the SP6 fragment.

[0039] Figure 2D shows a schematic diagram of the GTC2 fragment.

[0040] Figure 3 shows a comparison of the amino acid sequences of DptD (SEQ ID NO: 7) and the calcium dependent antibiotic (CDA) III protein of *Streptomyces coelicolor* (SEQ ID NO: 164) using the Clustal W program. See Example 3.

[0041] Figure 4 shows a comparison of the amino acid sequences of DptH (SEQ ID NO: 8) and the probable hydrolase (presumed thioesterase) associated with the CDA NRPS of *Streptomyces coelicolor* (SEQ ID NO: 165) using the Clustal W program. See Example 3.

- 17 -

[0042] Figures 5A-5C shows an analysis of daptomycin or A21978C lipopeptides produced from the *Streptomyces lividans* TK64 clone containing the daptomycin biosynthetic gene cluster CBUK138742 (ATCC Deposit PTA-3140, deposited March 1, 2001). Figure 5A shows an HPLC analysis of the broth of CBUK138742. The lower panel shows a trace plotting the maximum absorbance observed over the range of 200-600 nm for the HPLC eluate against time. The presence of three native lipopeptides, lipopeptides A21978C1 (the C1 lipopeptide), A21978C2 (the C2 lipopeptide) and A21978C3 (the C3 lipopeptide), is indicated by peaks with retention times of 5.61, 5.77 and 5.89 minutes, respectively. The upper panel shows the UV-visible spectra observed for these peaks. Figure 5B shows an ESI mass spectrum of daptomycin purified from decanoic acid-fed fermentation of *Streptomyces lividans* TK64 clone containing the daptomycin gene cluster. Figure 5C shows a ¹H NMR spectrum (400MHz, in d₆-DMSO) of daptomycin purified from decanoic acid-fed fermentation of CBUK138742.

[0043] Figure 6 is a diagram of the cloning vector pStreptoBAC V.

[0044] Figure 7 shows a *Hin*DIII digest of BAC clones from the daptomycin biosynthetic gene cluster. Lane 1 shows B12:01G05 (82 kb insert); Lane 2 shows B12:03A05 (120 kb insert); Lane 3 shows B12:06A12 (85 kb insert); Lane 3 shows B12:12FG06 (65 kb insert); Lane 5 shows B12:18H04 (46 kb insert) and Lane 6 shows B12:20C09 (65 kb insert).

[0045] Figure 8 shows a map of some BAC clones that cover approximately 180 to 200 kb of the daptomycin NPRS region in *Streptomyces roseosporus*.

- 18 -

[0046] Figure 9 is a schematic diagram of the gene structure of an NRPS.

[0047] Figure 10 is a dendrogram showing the adenylation (A) domain similarities for domains that specify Asn and Asp in the daptomycin NRPS and in the CDA NRPS from *Streptomyces coelicolor*. See Example 5.

[0048] Figure 11 shows the results of an HPLC analysis determining the stereochemistry of Asn. See Example 6.

[0049] Figure 12 is a schematic diagram showing the organization of the daptomycin NRPS.

[0050] Figure 13 shows a ¹H NMR spectrum the novel lipopeptide produced as described in Example 12C.

DETAILED DESCRIPTION OF THE INVENTION

15 Definitions and General Techniques

[0051] Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art.

20 Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as
30 described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. See, e.g.,

- 19 -

Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989); Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 3d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2000); Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel et al., *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, 4th ed., Wiley & Sons (1999); Harlow and Lane *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1990); Harlow and Lane *Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1998); and T. Kieser et al., *Practical Streptomyces Genetics*, John Innes Foundation, Norwich (2000); each of which is incorporated herein by reference in its entirety.

[0052] Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

[0053] The following terms, unless otherwise indicated, shall be understood to have the following meanings:

[0054] The term "thioesterase" refers to an enzyme that is capable of catalyzing the cleavage of a thioester

- 20 -

bond, which may result in the production of a cyclic or linear molecule.

[0055] The term "thioesterase activity" refers to an enzymatic activity of a thioesterase, or a mutein, homologous protein, analog, derivative, fusion protein or fragment thereof, that catalyzes cleavage of a thioester bond. A thioesterase activity includes, e.g., an association and/or dissociation constant, a catalytic rate and a substrate turnover rate. A thioesterase activity of a polypeptide may be the same as one of the thioesterase activities of DptH, the thioesterase domain of DptD, a polypeptide encoded by *dptH*, a polypeptide encoded by the thioesterase domain of *dptD*, a polypeptide having an amino acid sequence of the thioesterase domain of SEQ ID NO: 7 or a polypeptide having the amino acid sequence of SEQ ID NO: 8. The thioesterase activity may also differ from that of one of the above-described thioesterases; e.g., it may have an increased or decreased catalytic activity, a different association and/or dissociation constant or a different substrate for catalysis. A "decreased" or "increased" thioesterase activity refers to a decreased or increased catalytic activity of the thioesterase, respectively.

[0056] A "thioesterase derived from a daptomycin biosynthetic gene cluster" is a thioesterase or thioesterase domain that is encoded by one of the genes of a gene cluster that encodes polypeptides involved in the synthesis of daptomycin. Preferably, the thioesterase is derived from a daptomycin biosynthetic gene cluster from *Streptomyces*, preferably from a daptomycin biosynthetic gene cluster from *S. roseosporus*.

[0057] A "daptomycin biosynthetic gene cluster" is defined herein as a nucleic acid molecule that encodes a

- 21 -

number of polypeptides that are necessary for synthesis of daptomycin in an organism, preferably in a bacterial cell. A daptomycin biosynthetic gene cluster comprises a nucleic acid molecule that encodes at least DptA, DptBC, DptD and DptH, or that encode muteins, homologous proteins, allelic variants or fragments thereof, as well as other nucleic acid sequences that encode other polypeptides required for daptomycin synthesis.

Preferably, a daptomycin biosynthetic gene cluster comprises that part of BAC B12:03A05 that permits the synthesis of daptomycin when the part is introduced and expressed in a bacterial cell.

[0058] A "daptomycin NRPS" is defined herein as an NRPS that is capable of synthesizing daptomycin in an appropriate bacterial cell. A daptomycin NRPS comprises polypeptide subunits DptA, DptBC and DptD, or muteins, homologous proteins, allelic variants or fragments thereof, that are capable, when expressed in an appropriate cell, of directing the synthesis of daptomycin. A daptomycin NRPS may further comprise DptH and/or other polypeptide, such as DptE or DptF.

Preferably, the daptomycin NRPS is derived from the daptomycin biosynthetic gene cluster from *Streptomyces*, more preferably, the daptomycin NRPS is derived from *S. roseosporus*. The term "daptomycin NRPS" does not imply that the daptomycin NRPS can be used to synthesize only daptomycin. Rather, as used herein, the term is used solely for the purpose of describing that the NRPS was originally derived from a daptomycin biosynthetic gene cluster. The daptomycin NRPS may be used to synthesize molecules other than daptomycin, as described herein.

[0059] A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a

- 22 -

polypeptide and the expression control sequences that are operably linked to the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence
5 that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA.

[0060] A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid
10 molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

[0061] The terms "dptA", "dptBC" and "dptD" refer to
15 nucleic acid molecules that encode subunits of the daptomycin NRPS. In a preferred embodiment, the nucleic acid molecule is derived from *Streptomyces*, more preferably the nucleic acid molecule is derived from *S. roseosporus*. In a preferred embodiment, the *dptA*, *dptBC*
20 and *dptD* encode the polypeptides having the amino acid sequences of SEQ ID NOS: 9, 11 and 7, respectively. The terms "dptA", "dptBC" and "dptD" also refer to allelic variants of these genes, which may be obtained from other species of *Streptomyces* or from other *S. roseosporus*
25 strains.

[0062] The term "dptH" refers to a gene whose coding domain encodes a thioesterase from a daptomycin biosynthetic gene cluster of *S. roseosporus*, wherein the naturally-occurring thioesterase is a "free"
30 thioesterase. A free thioesterase is one that is not a functional domain of a larger polypeptide when it is naturally occurring. The *dptH* gene also encompasses the expression control sequences that are upstream of the

- 23 -

coding region of the gene, as discussed below. In one embodiment, the expression control sequences of *dptH* have the nucleic acid sequence of SEQ ID NO: 5. The term "dptH" also refers to the nucleic acid encoding the polypeptide defined by SEQ ID NO: 8. The term "dptH" also refers to allelic variants of this gene, which may be obtained from other species of *Streptomyces* or from other *S. roseosporus* strains.

[0063] The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each allele possesses a different nucleotide sequence. An allelic variant may encode the same polypeptide or a different one. As used herein, an allele is one that has at least 90% sequence identity, more preferably at least 95%, 96%, 97%, 98% or 99% sequence identity to the reference nucleic acid sequence, and encodes a polypeptide having similar or identical biological properties as the polypeptide encoded by the reference nucleic acid molecule.

[0064] The term "polynucleotide" or "nucleic acid molecule" refers to a polymeric form of nucleotides of at least 10 bases in length, either ribonucleotides or deoxynucleotides or a modified form of either type of nucleotide. The term includes single and double stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

[0065] An "isolated" or "substantially pure" nucleic acid or polynucleotide (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, e.g.,

- 24 -

ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, or (4) does not occur in nature as part of a larger sequence. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems.

[0066] A "part" of a nucleic acid molecule or polynucleotide refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 14 nucleotides of the reference nucleic acid molecule. Preferably, a part comprises at least 17 or 20 nucleotides of a reference nucleic acid molecule. More preferably, a part comprises at least 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 200, 300 400, 500 or 1000 nucleotides up to one nucleotide short of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

[0067] The term "oligonucleotide" refers to a polynucleotide generally comprising a length of 200 nucleotides or fewer. Preferably, oligonucleotides are 10 to 60 nucleotides in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50 or 60

- 25 -

nucleotides in length. Oligonucleotides may be single-stranded, e.g. for use as probes or primers, or may be double-stranded, e.g. for use in the construction of a mutant gene. Oligonucleotides of the invention can be
5 either sense or antisense oligonucleotides. An oligonucleotide can include a label for detection, if desired.

[0068] The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring
10 deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as
15 phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. See e.g., LaPlanche et al. *Nucl. Acids Res.* 14:9081 (1986); Stec et al. *J. Am. Chem. Soc.* 106:6077 (1984); Stein et al. *Nucl. Acids Res.* 16:3209 (1988); Zon et al. *Anti-Cancer Drug Design* 6:539 (1991); Zon et al. *Oligonucleotides and Analogues: A Practical Approach*, pp. 87-108 (F. Eckstein, Ed., Oxford University Press, Oxford England (1991)); Stec et al. U.S. Patent No.
20 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

[0069] Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the
30 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of

- 26 -

the polynucleotide sequence is referred to as the 3' direction.

[0070] The term "percent sequence identity" or "identical" in the context of nucleic acid sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. In one embodiment, polynucleotide sequences may be compared using Blast (Altschul et al., *J. Mol. Biol.* 215: 403-410, 1990). For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, 1990, (herein incorporated by reference). For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

[0071] The term "substantial homology" or "substantial similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with

- 27 -

another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%,
5 preferably at least about 90%, and more preferably at least about 95%, 96%, 97%, 98% or 99% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

10 **[0072]** Alternatively, substantial homology or similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions.

15 Typically, selective hybridization will occur when there is at least about 55% sequence identity -- preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% -- over a stretch of at least about 14 nucleotides. See, e.g., Kanehisa, 1984,
20 herein incorporated by reference.

[0073] Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the
25 number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend
30 upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One

- 28 -

having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization.

[0074] In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook et al., *supra*, page 9.51, hereby incorporated by reference.

[0075] The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^{\circ}\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l) \text{ where } l \text{ is the length of the hybrid in base pairs.}$$

[0076] The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$$T_m = 79.8^{\circ}\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35 (\% \text{ formamide}) - (820/l).$$

[0077] The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^{\circ}\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50 (\% \text{ formamide}) - (820/l).$$

[0078] In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter

hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

[0079] An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours, preferably 12-16 hours. Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours, preferably 12-16 hours. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or northern blot or for screening a library is 6X SSC at 42°C for at least ten hours, preferably 12-16 hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (e.g. 42°C and 6X SSC) and varying the formamide

- 30 -

concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. See Sambrook et al., *supra*, pages 8.46 and 9.46-9.58, herein incorporated by
5 reference.

[0080] Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (see Sambrook et al., *supra*, for SSC buffer). Often the high
10 stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In
15 general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

[0081] As defined herein, nucleic acids that do not
20 hybridize to each other under stringent conditions are still substantially homologous to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid is created synthetically or recombinantly using a
25 high codon degeneracy as permitted by the redundancy of the genetic code.

[0082] The polynucleotides of this invention may include both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of
30 the above. They may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for

- 31 -

example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.) Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

[0083] The term "mutated" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid sequence is the wild type nucleic acid sequence for a thioesterase. The nucleic acid sequence may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

[0084] The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire

length of the PCR product. See, e.g., Leung et al., *Technique*, 1, pp.11-15 (1989) and Caldwell and Joyce *PCR Methods Applic.*, 2, pp. 28-33 (1992).

5 **[0085]** The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. See, e.g., Reidhaar-Olson et al., *Science*, 241, pp. 53-57 (1988).

10 **[0086]** The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

15 **[0087]** The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence homology, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. See, e.g., Stemmer, *Proc. Natl. Acad. Sci. U.S.A.*, 91, pp. 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

25 **[0088]** The term "in vivo mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the

- 33 -

DNA in a mutator strain will eventually generate random mutations within the DNA.

[0089] The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

[0090] The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. See, e.g., Arkin and Youvan, *Proc. Natl. Acad. Sci. U.S.A.*, 89, pp. 7811-7815 (1992).

[0091] The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. See, e.g., Delegrave and Youvan, *Biotechnol. Res.*, 11, pp. 1548-1552 (1993); and random and site-directed mutagenesis, Arnold, *Curr. Opin. Biotechnol.*, 4, pp. 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

[0092] "Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

[0093] The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (e.g., ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

[0094] The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that

- 35 -

infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

[0095] The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

[0096] The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. As used herein, a

- 36 -

polypeptide comprises at least six amino acids, preferably at least 8, 10, 12, 15, 20, 25 or 30 amino acids, and more preferably the polypeptide is the full length of the naturally-occurring polypeptide. A
5 polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a thioesterase
10 derived from the daptomycin biosynthetic gene cluster, as well as a fragment, mutant, analog and derivative thereof.

[0097] The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue
15 of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in
20 nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially
25 free of naturally associated components by isolation, using protein purification techniques well known in the art.

[0098] A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially
30 purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically

comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art for purification.

[0099] The term "polypeptide fragment" as used herein refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

[0100] A "derivative" refers to polypeptides or fragments thereof that are substantially homologous in primary structural sequence but which include, e.g., *in vivo* or *in vitro* chemical and biochemical modifications or which incorporate amino acids that are not found in the native polypeptide. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides, and various enzymatic modifications, as

will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes
5 such as ^{125}I , ^{32}P , ^{35}S , and ^3H , ligands which bind to labeled antiligands (e.g., antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity
10 required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well known in the art. See Ausubel et al., 1992, hereby incorporated by reference.

[0101] The term "fusion protein" refers to
15 polypeptides comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least
20 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a
25 nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the
30 polypeptide or a fragment thereof to another protein.

[0102] The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may also

- 39 -

be termed a "peptide mimetic" or a "peptidomimetic."
See, e.g., Fauchere, *J. Adv. Drug Res.* 15:29 (1986);
Veber and Freidinger *Trends Neurosci.* p.392 (1985); and
Evans et al. *J. Med. Chem.* 30:1229 (1987), which are
5 incorporated herein by reference. Such compounds are
often developed with the aid of computerized molecular
modeling. Peptide mimetics that are structurally similar
to useful peptides may be used to produce an equivalent
effect. Generally, peptidomimetics are structurally
10 similar to a paradigm polypeptide (i.e., a polypeptide
that has a desired biochemical property or
pharmacological activity), such as a thioesterase, but
have one or more peptide linkages optionally replaced by
a linkage selected from the group consisting of:
15 --CH₂NH--, --CH₂S--, --CH₂-CH₂--, --CH=CH--(cis and
trans), --COCH₂--, --CH(OH)CH₂--, and -CH₂SO--, by methods
well known in the art. Systematic substitution of one or
more amino acids of a consensus sequence with a D-amino
acid of the same type (e.g., D-lysine in place of
20 L-lysine) may also be used to generate more stable
peptides. In addition, constrained peptides comprising a
consensus sequence or a substantially identical consensus
sequence variation may be generated by methods known in
the art (Rizo and Gierasch, *Annu. Rev. Biochem.* 61:387
25 (1992), incorporated herein by reference); for example,
by adding internal cysteine residues capable of forming
intramolecular disulfide bridges which cyclize the
peptide.

[0103] A "polypeptide mutant" or "mutein" refers to a
30 polypeptide whose sequence contains substitutions,
insertions or deletions of one or more amino acids
compared to the amino acid sequence of a native or wild
type protein. A mutein may have one or more amino acid

- 40 -

point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. In a preferred embodiment of the present invention, a mutein has the same or increased thioesterase activity as a naturally-occurring thioesterase. A mutein has at least 50%, 60% or 70% sequence homology to the wild type protein, more preferred are muteins having at least 80%, 85% or 90% sequence homology to the wild type protein, even more preferred are muteins exhibiting at least 95%, 96%, 97%, 98% or 99% sequence identity. Sequence homology may be measured by any common sequence analysis algorithm, such as Gap or Bestfit, using default parameters.

[0104] Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such derivatives, analogs, fusion proteins and muteins. Single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. A conservative amino acid

- 41 -

substitution should not substantially change the structural characteristics of the parent sequence (e.g., a replacement amino acid should not tend to break a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in *Proteins, Structures and Molecular Principles* (Creighton, Ed., W. H. Freeman and Company, New York (1984)); *Introduction to Protein Structure* (C. Branden and J. Tooze, eds., Garland Publishing, New York, N.Y. (1991)); and Thornton et al. *Nature* 354:105 (1991), which are each incorporated herein by reference.

[0105] As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See *Immunology - A Synthesis* (2nd Edition, E.S. Golub and D.R. Gren, Eds., Sinauer Associates, Sunderland, Mass. (1991)), which is incorporated herein by reference. Stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α -, α -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: γ -carboxyglutamate, ϵ -N,N,N-trimethyllysine, ϵ -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

- 42 -

[0106] A protein has "homology" or is "homologous" to a protein from another organism if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences. Although two proteins are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences. In a preferred embodiment, a homologous protein is one that exhibits at least 50%, 60% or 70% sequence identity to the wild type protein, preferred are homologous proteins that exhibit at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% sequence identity. In addition, although in many cases proteins with similar amino acid sequences will have similar functions, the term "homologous" does not imply that the proteins must be functionally similar to each other.

[0107] When "homologous" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of homology may be

- 43 -

adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art (see, e.g., Pearson et al., 1994, herein incorporated by reference).

5 **[0108]** The following six groups each contain amino acids that are conservative substitutions for one another:

1. Serine (S), Threonine (T);
2. Aspartic Acid (D), Glutamic Acid (E);
- 10 3. Asparagine (N), Glutamine (Q);
4. Arginine (R), Lysine (K);
5. Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
6. Phenylalanine (F), Tyrosine (Y),
- 15 Tryptophan (W).

[0109] Sequence homology for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics

20 Computer Group (GCG), University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions and other modifications,

25 including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to

determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. See,
5 e.g., GCG Version 6.1.

[0110] A preferred algorithm when comparing a polypeptide sequence to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp, tblastn or
10 BlastX. See Altschul et al. Nucleic Acids Res. 25:3389-3402 (1997), herein incorporated by reference. BlastX, which compares a translated nucleotide sequence to a protein database, may be performed online through the servers located at the National Center for Biotechnology
15 Information. Preferred parameters for blastp, which compares a protein sequence to a protein database are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)
Cost to extend a gap:	1 (default)
Max. alignments:	100 (default)
Word size:	11 (default)
No. of descriptions:	100 (default)
Penalty Matrix:	BLOSUM62

[0111] The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more
20 usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

[0112] Database searching using amino acid sequences can be measured by algorithms other than blastp known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1.

5 FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, 1990, herein incorporated by reference). For example, percent sequence identity between amino acid sequences can be determined using
10 FASTA with its default parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

[0113] An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof
15 that competes with the intact antibody for antigen-specific binding. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')₂, Fv, dAb,
20 and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab
25 fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; a F(ab')₂ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; a Fd fragment consists of the VH and CH1 domains; an Fv fragment
30 consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment (Ward et al., *Nature* 341:544-546, 1989) consists of a VH domain.

[0114] A single-chain antibody (scFv) is an antibody in which a VL and VH regions are paired to form a monovalent molecules via a synthetic linker that enables them to be made as a single protein chain (Bird et al.,
5 Science 242:423-426, 1988 and Huston et al., *Proc. Natl. Acad. Sci. USA* 85:5879-5883, 1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing
10 between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites (see e.g., Holliger et al., *Proc. Natl. Acad. Sci. USA* 90:6444-6448, 1993, and Poljak et al., *Structure*
15 2:1121-1123, 1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another
20 polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions
25 from one or more other antibodies.

[0115] An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two
30 identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

[0116] An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature.

[0117] A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. For example, a neutralizing anti-thioesterase antibody may be one that blocks the activity of the thioesterase. An "activating antibody" is an antibody that increases the activity of a polypeptide. For example, an activating anti-thioesterase antibody is one that increases the activity of a thioesterase.

[0118] The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is $\leq 1 \mu\text{M}$, preferably $\leq 100 \text{ nM}$ and most preferably $\leq 10 \text{ nM}$.

[0119] The term patient includes human and veterinary subjects.

[0120] Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion

- 48 -

of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

5 *Nucleic Acid Molecules*

[0121] In one aspect, the present invention provides a nucleic acid molecule encoding a thioesterase or a daptomycin NRPS or a subunit thereof. In one embodiment, the nucleic acid molecule encodes one or more of DptA, 10 DptBC or DptD. In a preferred embodiment, the nucleic acid molecules encodes a polypeptide comprising any one of the amino acid sequences of SEQ ID NOS: 9, 11 or 7. In another preferred embodiment, the nucleic acid molecule comprises *dptA*, *dptBC* and/or *dptD*. In a further 15 preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence comprising any one of SEQ ID NOS: 10, 12 or 3.

[0122] In another embodiment, the nucleic acid molecule encodes a thioesterase that is derived from a 20 daptomycin biosynthetic gene cluster. In a preferred embodiment, the nucleic acid molecule encodes a thioesterase derived from a daptomycin biosynthetic gene cluster that is a free thioesterase or is an integral thioesterase. In another preferred embodiment, the 25 nucleic acid molecule encodes DptH or the thioesterase domain of DptD. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide comprising an amino acid sequence of the thioesterase domain of SEQ ID NO: 7 or has the amino acid sequence of SEQ ID NO: 8. In 30 another embodiment, the nucleic acid molecule comprises the thioesterase-encoding domain of *dptD* or *dptH* from the daptomycin biosynthetic gene cluster. In another preferred embodiment, the nucleic acid molecule comprises

a nucleic acid sequence of SEQ ID NO: 6 or of SEQ ID NO: 3, or the region comprising the thioesterase-encoding portion thereof. In another embodiment, the nucleic acid molecule also encodes a daptomycin NRPS or a subunit thereof. See Examples 1-6 regarding the isolation and identification of *dptA*, *dptBC*, *dptD* and *dptH* and other genes of the daptomycin biosynthetic gene cluster.

[0123] In another embodiment, the nucleic acid molecule encodes an acyl CoA ligase. In a preferred embodiment, the nucleic acid molecule encodes DptE, preferably a nucleic acid molecule encoding SEQ ID NO: 15. In a more preferred embodiment, the nucleic acid molecule comprises *dptE*. In an even more preferred embodiment, the nucleic acid molecule comprises SEQ ID NO: 16. In another embodiment, the nucleic acid molecule encodes an acyl transferase. In a preferred embodiment, the nucleic acid molecule encodes DptF, preferably a nucleic acid molecule encoding SEQ ID NO: 17. In a more preferred embodiment, the nucleic acid molecule comprises *dptF*. In an even more preferred embodiment, the nucleic acid molecule comprises SEQ ID NO: 18.

[0124] Another embodiment of the invention provides a nucleic acid molecule comprising a DNA sequence from a bacterial artificial chromosome (BAC) comprising nucleic acid sequences from *S. roseosporus*. In a preferred embodiment, the nucleic acid molecule comprises a *S. roseosporus* nucleic acid sequence from any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or B12:03A05. In a preferred embodiment, the nucleic acid molecule comprises a *S. roseosporus* nucleic acid sequence from B12:03A05 (ATCC Deposit PTA-3140, deposited March 1, 2001). The nucleic acid molecule may comprise the entire *S. roseosporus* nucleic acid sequence

in the BAC clone or may comprise a part thereof. In a preferred embodiment, the part is a nucleic acid molecule that comprises at least one nucleic acid sequence that can encode a polypeptide, preferably a full-length polypeptide, i.e., a nucleic acid molecule that encodes a polypeptide from its start codon to its stop codon. In one preferred embodiment, the part comprises a nucleic acid molecule encoding a polypeptide involved in daptomycin biosynthesis, such as, without limitation, *dptA*, *dptBC*, *dptD*, *dptE*, *dptF* or *dptH*.

[0125] In another embodiment, a part from the BAC clone is a nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide selected from SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136. In another embodiment, the part from the BAC clone is a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133 or 135.

[0126] The polypeptides having the amino acid sequences of SEQ ID NOS: 110 and 112 are regulators of daptomycin biosynthesis. Multiple regulators are needed for the function of biosynthetic pathways in *Streptomyces* (Bate et al. *Chem. Biol.*, 6, 617-624, 1999; Baltz, *Bioprocess Technol.* 22, 308-381, 1995). For example, the biosynthetic pathway for bialaphos in *S. hygroscopicus* contains a gene, *bprA*, which has a positive regulatory

- 51 -

role in antibiotic production (Raibaud et al., *J. Bacteriol.* 173, 4454-4463, 1991). It has also been shown that increases of antibiotic production can be achieved by increasing copy number of positive regulator genes in a variety of producing strains (e.g. *S. lividans*, Vogtli et al., *Mol. Microbiol.* 14, 643-653, 1994; in *S. argillaceus*, Lombo et al., *J. Bacteriol.* 181, 642-647, 1999 and in *S. peucetius*, Otten et al., *Microbiology* 146, 1457-1468, 2000). The regulatory activator polypeptide of SEQ ID NO: 110, which is encoded by the nucleic acid molecule having the nucleotide sequence of SEQ ID NO: 109, shares identities and similarities not only with brpA and with other regulatory proteins found in streptomyces but also with the luxR-family proteins involved in quorum sensing. All of these are DNA-binding proteins in the family of two-component transcriptional activators (Kenney, *Curr. Opin. Microbiol.* 5, 135-141, 2002). Thus, the regulatory activator polypeptide of SEQ ID NO: 110 can be used to augment the yield of daptomycin in *S. roseosporus*. The regulatory activator gene or a biologically active portion thereof can be cloned into an integrative or autonomously replicating expression vector and reintroduced into one or more neutral sites in one or more copies in *S. roseosporus*. The transgenic strain may be fermented and analyzed for daptomycin production as described in Example 9 and could be used to produce a larger amount of daptomycin than the wildtype strain.

[0127] The polypeptide having the amino acid sequence of SEQ ID NO: 112, which is encoded by the nucleotide sequence of SEQ ID NO: 111, shares significant amounts of identities and similarities with a putative DeoR-family transcriptional regulator from *Streptomyces coelicolor* as well as a variety of catabolite repressors (LacI, rbsR,

- 52 -

malR, REG1). These proteins bind to the promoter regions to prevent the transcription (Zeng and Saxild, J. Bacteriol. 181, 1719-1729, 1999; Oskouian and Stewart, J. Bacteriol. 172, 3804-3812, 1990). Thus, this gene is a negative regulator of daptomycin biosynthesis.

Therefore, disruption or deletion of this negative regulator gene or inhibition of its protein product should lead to constitutive expression of daptomycin and/or enhanced yield. In another embodiment, one may delete the negative regulatory gene and insert multiple copies of the positive regulatory gene to increase daptomycin production even more.

[0128] The polypeptides having amino acids sequences of SEQ ID NOS: 19, 21, 29, 45, 47, 49, 63, 67, 75 and 77 (nucleic acid sequences of SEQ ID NOS: 20, 22, 30, 46, 48, 50, 64, 68, 76 or 78) are ABC transporters. Some of the polypeptides are pump-like polypeptides with Walker motifs while others are polypeptides that have a role in metal scavenging, e.g., iron or manganese transport (see Tables 6 and 7). The nucleic acid molecule comprising SEQ ID NO: 76 encodes an ATP-binding component of an ABC transporter system, as determined by its sequence similarity to ORF1 of (AAD44229.1) of *S. rochei* and the *S. peucetius* DrrA (P32010) genes. The encoded polypeptide has both a Walker A and a Walker B motif. Further, its synthesis appears to be translationally coupled to that of a nucleic acid molecule comprising SEQ ID NO: 78, which encodes a DrrB-like polypeptide, as determined by its sequence similar to the *S. peuticeus* DrrB product (AAA74718.1), encoding the integral membrane component. The polypeptide having an amino acid sequence of SEQ ID NO: 21 is a *StrV* homolog, while the polypeptide having an amino acid sequence of SEQ ID NO: 19 is a *StrW*

- 53 -

homolog. See, e.g., Beyer et al., 1996, *supra*. The *StrV* homolog has both Walker motifs, while the *StrW* homolog has only a Walker B motif. Both nucleic acid sequences encoding the polypeptide are on the complementary strand and appear to be translationally regulated. They have *S. coelicolor* homologs, G8A.01 and G8A.02 (emb| CAB88931, CAB88932). See Tables 6 and 7.

[0129] In another aspect, a part of the BAC clone is a nucleic acid molecule comprising a nucleic acid sequence encoding an oxidoreductase; a dehydrogenase; a transcriptional regulator involved in antibiotic resistance; NovABC-related polypeptides, which are involved in the biosynthesis of novobiocin, an antimicrobial agent; a monooxygenase; an acyl CoA thioesterase; a DNA helicase; a DNA ligase; a hydrolase; a thermostable neutral protease; ABC transporters that may be useful in the transport of daptomycin; a spo VK-like protein involved in endospore formation; a serine protease; and an FtsK/SpoIIIE-like protein involved in DNA segregation during septation and spore formation. These nucleic acid molecules and encoded polypeptides may be useful in daptomycin biosynthesis; e.g., the acyl CoA thioesterase may be useful for the reasons provided above for thioesterases and may also be important in the addition of the lipid tail to the peptide domain of daptomycin. These nucleic acid molecules encoding enzymes are also useful because they may be used in the same way as other oxidoreductases, dehydrogenases, monooxygenases, hydrolases, serine or neutral proteases, DNA helicases or DNA ligases are used in the art. Notably, the transcriptional regulator can be mutated using well-known methods to increase or decrease daptomycin or other antibiotic resistance. The nucleic

acid molecules encoding NovABC-related polypeptides may be used in the same way as NovABC is used in the art, e.g., to produce novobiocin or related antimicrobial agents. The polypeptides having the above-described activity comprise the amino acid sequences of SEQ ID NOS: 23, 25, 27, 29, 33, 35, 37, 91, 93, 97, 99, 104, 108, 114, 116, 118, 120, 130, 132, 134 and 136 and are encoded by nucleic acid sequences of SEQ ID NOS: 24, 26, 28, 30, 34, 36, 38, 92, 94, 98, 100, 105, 107, 113, 115, 117, 119, 129, 131, 133 and 135.

[0130] In another aspect, a part of the BAC clone is a nucleic acid molecule that encodes a polypeptide that does not have a defined function but which is highly homologous to nucleic acid molecules and polypeptides from other *Streptomyces*. These nucleic acid molecules (SEQ ID NOS: 62, 66, 70, 80, 82, 84, 86, 88, 96, 102, 121, 123, 125 and 127), the polypeptides they encode (SEQ ID NOS: 61, 65, 69, 79, 81, 83, 85, 87, 95, 101, 122, 124, 126 and 128) and antibodies to the polypeptides may be used to identify other *Streptomyces* species using standard molecular biological and protein chemistry techniques (e.g., PCR, RT-PCR, Southern blotting, northern blotting, ELISAs, radioimmunoassays or western blotting), which is useful, e.g., in microbiological testing or forensics. In another embodiment, a part of the BAC clone is a nucleic acid molecule that encodes a polypeptide that does not have a defined function and is not highly homologous to a nucleic acid molecule or polypeptide from another species. These nucleic acid molecules (SEQ ID NOS: 32, 40, 42, 44, 52, 54, 56, 58, 60, 72 and 74) are nevertheless useful because they are close to the daptomycin biosynthetic gene cluster, and as such, they can be used to identify nucleic acid molecules

that encode all or a part of the daptomycin biosynthetic gene cluster. Parts of the BAC clone that do not encode a polypeptide are useful for the same reasons. Further, the polypeptides having the amino acid sequence of SEQ ID NOS: 31, 39, 41, 43, 51, 53, 55, 57, 59, 71 and 73 can be used to make antibodies that can be used to identify *S. roseosporus*. Because the polypeptides are not highly homologous to any other species, the antibodies would likely be highly specific for *S. roseosporus*.

10 [0131] In another aspect, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule as described above. In a preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes DptA, DptBC, DptD or DptH. In another preferred embodiment, the invention provides a nucleic acid molecules that selectively hybridizes to a nucleic acid molecule that encodes SEQ ID NOS: 9, 11, 7 or 8. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of *dptA*, *dptBC*, *dptD* or *dptH*. In another preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence SEQ ID NOS: 10, 12, 3 or 6. The invention also provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising an *S. roseosporus* nucleic acid sequence from any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or B12:03A05, preferably that from B12:03A05. In a preferred embodiment, the invention provides a nucleic acid molecule that selectively

- 56 -

hybridizes to a nucleic acid molecule encoding SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136 or to a nucleic acid molecule comprising the nucleic acid sequence SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133 or 135. The selective hybridization of any of the above-described nucleic acid sequences may be performed under low stringency hybridization conditions. In a preferred embodiment, the selective hybridization is performed under high stringency hybridization conditions. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to recombinantly express a polypeptide of the invention.

[0132] In another aspect, the invention provides a nucleic acid molecule that is homologous to a nucleic acid encoding a daptomycin NRPS or subunit thereof, a thioesterase from a daptomycin biosynthetic gene cluster, or a nucleic acid molecule comprising an *S. roseosporus* nucleic acid sequence from any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or, preferably, B12:03A05. The invention provides a nucleic acid molecule homologous to a nucleic acid molecule encoding DptA, DptBC, DptD or DptH. In one embodiment, the nucleic acid molecule is homologous to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NOS: 9, 11, 7 or 8. In a preferred

- 57 -

embodiment, the nucleic acid molecule is homologous to any one or more of *dptA*, *dptBC* or *dptD*. In another embodiment, the nucleic acid molecule is homologous to a thioesterase encoded by the thioesterase domain of *dptD* or by *dptH*. In a more preferred embodiment, the nucleic acid molecule is homologous to a nucleic acid molecule having a nucleic acid sequence of SEQ ID NOS: 10, 12, 3 or 6. In another preferred embodiment, the invention provides a nucleic acid molecule that is homologous to a nucleic acid molecule encoding SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136 or to a nucleic acid molecule comprising the nucleic acid sequence SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133 or 135. In a preferred embodiment, a homologous nucleic acid molecule is one that has at least 60%, 70%, 80% or 85% sequence identity with a nucleic acid molecule described herein. In a more preferred embodiment, the homologous nucleic acid molecule is one that has at least 90%, 95%, 97%, 98% or 99% sequence identity with a nucleic acid molecule described herein. Further, in one embodiment, a homologous nucleic acid molecule is homologous over its entire length to a nucleic acid molecule encoding a daptomycin NRPS or subunit thereof, a thioesterase, or nucleic acid molecule that encodes a polypeptide as described herein. In another embodiment, a homologous nucleic acid molecule is homologous over

- 58 -

only a part of its length to a nucleic acid molecule described herein, wherein the part is at least 50 nucleotides of the nucleic acid molecule, preferably at least 100 nucleotides, more preferably at least 200
5 nucleotides, even more preferably at least 300 nucleotides.

[0133] In another embodiment, the invention provides a nucleic acid that is an allelic variant of a gene encoding a daptomycin NRPS or subunit thereof, a
10 thioesterase from a daptomycin biosynthetic gene cluster, or a nucleic acid molecule comprising an *S. roseosporus* nucleic acid sequence from any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or B12:03A05. In a preferred embodiment, the invention
15 provides a nucleic acid that is an allelic variant of *dptA*, *dptBC*, *dptD* or *dptH*. In an even more preferred embodiment, the allelic variant is a variant of a gene, wherein the gene encodes DptA, DptBC, DptD or DptH. In another preferred embodiment, the allelic variant is a
20 variant of a gene that encodes a polypeptide comprising an amino acid sequence of SEQ ID NOS: 9, 11, 7 or 8. In a yet more preferred embodiment, the allelic variant is a variant of a gene, wherein the gene has the nucleic acid sequence of SEQ ID NOS: 10, 12, 3 or 6. An allelic
25 variant of *dptH* or the thioesterase of *dptD* preferably encodes a thioesterase with the same or similar enzymatic activity compared to that of the polypeptide having the amino acid sequence of the thioesterase domain of SEQ ID NO: 7 or has the amino acid sequence of SEQ ID NO: 8. An
30 allelic variant of *dptA*, *dptBC* or *dptD* preferably encodes a polypeptide having the same activity as the daptomycin NRPS having the amino acid sequences of SEQ ID NOS: 9, 11 or 7, respectively. In another embodiment, the invention

provides an allelic variant of a nucleic acid molecule that encodes SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 5 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136 or to a nucleic acid molecule comprising the nucleic acid sequence SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 10 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133 or 135. In a preferred embodiment, the allelic variant encodes a polypeptide having the same biological activity of the 15 polypeptide; e.g., it encodes a polypeptide having ABC-transporter activity.

[0134] A further object of the invention is to provide a nucleic acid molecule that comprises a part of a nucleic acid sequence of the instant invention. The 20 invention provides a part of a nucleic acid molecule encoding a daptomycin NRPS, a subunit thereof, a thioesterase from a daptomycin biosynthetic gene cluster, or a part of a nucleic acid molecule that comprises an *S. roseosporus* nucleic acid sequence from any one of BAC 25 clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or, preferably, B12:03A05. The invention also provides a part of a selectively-hybridizing or homologous nucleic acid molecule, as described above. The invention provides a part of an allelic variant of a 30 nucleic acid molecule, as described above. A part comprises at least 10 nucleotides, more preferably at least 15, 20, 25, 30, 35, 40, 50, 100, 150, 200, 250 or 300 nucleotides. The maximum size of a nucleic acid part

- 60 -

is one nucleotide shorter than the entire nucleic acid molecule, if the nucleic acid molecule encodes more than one gene, or is one nucleotide shorter than the nucleic acid molecule encoding the full-length protein, if the
5 nucleic acid molecule encodes a single polypeptide.

[0135] In another aspect, the hybridizing or homologous nucleic acid molecule, the allelic variant, or the part of the nucleic acid molecule encodes a polypeptide that has the same biological activity as the
10 native (wild-type) polypeptide.

[0136] In another aspect, the invention provides a nucleic acid molecule that encodes a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

15 [0137] A nucleic acid molecule of this invention may encode a single polypeptide or multiple polypeptides. In one embodiment, the invention provides a nucleic acid molecule that encodes multiple, translationally coupled polypeptides, e.g., a nucleic acid molecule that encodes
20 DptA, DptBC and DptD. The invention also provides a nucleic acid molecule that encodes a single polypeptide derived from *S. roseosporus*, e.g., DptA, DptBC or DptD, or a polypeptide fragment, mutein, fusion protein, polypeptide analog or homologous protein thereof. The
25 invention also provides nucleic acid sequences, such as expression control sequences, that are not associated with other *S. roseosporus* sequences.

[0138] In certain embodiments, the nucleic acid molecules of this invention may not include any one or
30 more of the plasmids, cosmids designated, pRHB153, pRHB157, pRHB159, pRHB160, pRHB161, pRHB162, pRHB166, pRHB168, pRHB169, pRHB170, pRHB172, pRHB173, pRHB174, pRHB599, pRHB602, pRHB603, pRHB613, pRHB614, pRHB680,

pRHB678 or pRHB588 by McHenney et al., *J. Bacteriol.* 180: 143-151 (1998), herein incorporated by reference in its entirety to the extent any of those plasmids or cosmids are part of the prior art and fall within the scope of
5 any specific claim made in this application. Further analysis performed has indicated that the location and orientation of some of the daptomycin inserts in plasmids or cosmids recited in McHenney et al., supra, are incorrect.

10 *Expression Control Sequences*

[0139] In another embodiment, the invention provides a nucleic acid molecule comprising one or more expression control sequences from a gene comprising a nucleic acid sequence that encodes a thioesterase or daptomycin NRPS
15 from the daptomycin biosynthetic gene cluster. In a preferred embodiment, the nucleic acid molecule comprises a part or all of the expression control sequences of the daptomycin NRPS or *dptH*. In a yet more preferred embodiment, the nucleic acid molecule comprises all or a
20 part of SEQ ID NO: 2 or SEQ ID NO: 5. In another preferred embodiment, the nucleic acid molecule comprises an expression control sequence from an *S. roseosporus* nucleic acid sequence from any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or,
25 preferably, B12:03A05. Without wishing to be bound by any theory, it is thought that the nucleic acid sequence upstream of *dptA* in the daptomycin biosynthetic gene cluster (SEQ ID NO: 2) comprises the native expression control sequences for *dptA*, *dptBC* and *dptD*. Further, it
30 is thought that a single transcript for *dptA*, *dptBC* and *dptD* is generated and that expression of DptA, DptBC and DptD are translationally coupled.

- 62 -

[0140] In a preferred embodiment, the entire expression control sequence of a gene comprising a nucleic acid sequence that encodes a daptomycin NRPS and/or a thioesterase from the daptomycin biosynthetic gene cluster is used to control transcription. In another embodiment, only a part of the expression control sequence of a gene comprising a nucleic acid sequence that encodes a daptomycin NRPS and/or a thioesterase from the daptomycin biosynthetic gene cluster is used to control transcription. One having ordinary skill in the art may determine which part(s) of the gene to use to control transcription using methods known in the art. For instance, one may ligate a nucleic acid sequence comprising all or a part of an expression control sequence of a daptomycin NRPS and/or a thioesterase gene into a vector comprising a reporter gene. Examples of such reporter genes include, without limitation, chloramphenicol acetyltransferase (CAT), luciferase, green fluorescent protein, β -galactosidase and the like. The nucleic acid molecule comprising the expression control sequence is ligated into the vector such that it can act as a promoter or enhancer of the reporter gene. The vector is introduced into a host cell and expression is induced. Then, one may assay for the production of the reporter gene product to determine if the part(s) of the expression control sequence is sufficient to activate or regulate transcription. Methods of determining whether a nucleic acid sequence is sufficient to regulate transcription are routine and well-known in the art. See, e.g., Ausubel et al., *supra*.

[0141] A nucleic acid molecule comprising all or a part of an expression control sequence described herein, or multiple copies of these expression control sequences

- 63 -

or parts thereof, may be operatively linked to a second nucleic acid molecule to regulate the transcription of the second nucleic acid molecule. In one embodiment, the invention provides a nucleic acid molecule comprising the expression control sequences operatively linked to a heterologous nucleic acid molecule, such as a nucleic acid molecule that encodes a polypeptide not usually expressed by *S. roseosporus*. In another preferred embodiment, the nucleic acid molecule comprising the expression control sequences is inserted into a vector, preferably a bacterial vector. In a more preferred embodiment, the vector is introduced into a bacterial host cell, more preferably into a *Streptomyces* or *E. coli*, and even more preferably into a *S. roseosporus*, *S. lividans* or *S. fradiae* host cell.

[0142] The invention also provides a nucleic acid sequence comprising the expression control sequence from *S. roseosporus* as described herein operatively linked to a nucleic acid sequence encoding a polypeptide involved in a daptomycin NRPS, a thioesterase derived from the daptomycin biosynthetic gene cluster, or a nucleic acid molecule from a BAC clone or part thereof as described herein. The expression control sequence may be operatively linked to a nucleic acid molecule encoding DptA, DptBC, DptD or DptH, to a nucleic acid molecule encoding a polypeptide derived from the *S. roseosporus* sequences from a BAC clone of the invention, preferably B12:03A05, or to a nucleic acid molecule encoding a fragment, homologous protein, mutein, analog, derivative or fusion protein thereof. The expression control sequence may be operatively linked to a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence of SEQ ID NOS: 9, 11, 7 or 8, or to a fragment

- 64 -

thereof. Preferably, the expression control sequence is operatively linked to the coding region of one or more of *dptA*, *dptBC*, *dptD* or *dptH*. In a more preferred embodiment, the expression control sequence is

5 operatively linked to a nucleic acid sequence selected from SEQ ID NOS: 10, 12, 3 or 6, or to a part thereof. The invention also provides an expression control sequence operatively linked to the coding region of a polypeptide comprising an amino acid sequence SEQ ID NOS:

10 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136 or to a nucleic acid molecule

15 comprising the nucleic acid sequence SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127,

20 129, 131, 133 or 135.

[0143] In another embodiment, the invention provides a nucleic acid molecule comprising one or more expression control sequences that directs the transcription of a nucleic acid molecule encoding a daptomycin NRPS, a

25 subunit, module or domain thereof, a thioesterase, or a nucleic acid molecule encoding a polypeptide derived from the *S. roseosporus* sequences from a BAC clone of the invention, wherein the expression control sequence(s) are not derived from a daptomycin biosynthetic gene cluster.

30 Examples of suitable expression control sequences are provided *infra*.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

- 65 -

[0144] Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic acid sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

[0145] A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

[0146] In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, e.g., the numerous derivatives of phage lambda, e.g., NM989, λ GT10 and λ GT11, and other phages, e.g., M13 and filamentous single stranded phage DNA. A preferred vector is a bacterial artificial chromosome (BAC). A more preferred vector is pStreptoBAC, as described in Example 2.

- 66 -

[0147] In other embodiments, eukaryotic host cells, such as yeast, insect or mammalian cells, may be used. Yeast vectors include Yeast Integrating plasmids (e.g., YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast centromere plasmids (the YCp series plasmids), pGPD-2, 2 μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz and Sugino, Gene, 74, pp. 527-34 (1988) (YIplac, YEplac and YCplac). Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (e.g., vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (e.g., bovine papillomavirus), and retroviral vectors (e.g., murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

[0148] In addition, any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, e.g., promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, e.g., sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in

- 67 -

the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

[0149] Examples of useful expression control sequences include, for example, the early and late promoters of SV40 or adenovirus, the lac system, the trp system, the TAC or TRC system, the T3 and T7 promoters, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast α -mating system, the GAL1 or GAL10 promoters, and other constitutive and inducible promoter sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. Other expression control sequences include those from the daptomycin biosynthetic gene cluster, such as those described *supra*.

[0150] Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (e.g., ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. Preferred marker genes and stabilizing sequences are disclosed in pStreptoBAC, which is described in Example 2. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well known to those of skill in

- 68 -

the art and are described in an assortment of laboratory manuals, including Sambrook et al., *supra*, 1989; and Ausubel et al. Product information from manufacturers of biological, chemical and immunological reagents also
5 provide useful information. Example 2 provides preferred nucleic acid cloning and sequencing methods.

[0151] Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither
10 will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For
15 example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the
20 vector, such as antibiotic or other selection markers, should also be considered.

[0152] In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the
25 sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures.

Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity
30 of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of

purification from them of the products coded for by the nucleic acid sequences of this invention.

[0153] The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention
5 may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to
10 this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity. In a preferred embodiment, the polypeptides are expressed in a heterologous bacterial host cell. In a more preferred embodiment, the
15 polypeptides are expressed in a heterologous *Streptomyces* host cell, still more preferably a *S. lividans* or *S. fradiae* host cell. See, e.g., Example 7, *infra*.

[0154] Transformation and other methods of introducing nucleic acids into a host cell (e.g., conjugation,
20 protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well known in the art (see,
25 for instance, Ausubel, *supra*, and Sambrook et al., *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest.
30 Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable

- 70 -

expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or in a stable manner, and whether to express the protein constitutively or inducibly.

[0155] A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well known eukaryotic and prokaryotic hosts, such as strains of *E. coli*,
10 *Pseudomonas*, *Bacillus*, *Streptomyces*, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, BHK, MDCK and various murine cells, e.g., 3T3 and WEHI cells, African green monkey cells such as COS 1, COS 7, BSC 1, BSC 40, and BMT 10, and human cells
15 such as VERO, WI38, and HeLa cells, as well as plant cells in tissue culture. In a preferred embodiment, the host cell is *Streptomyces*. In a more preferred embodiment, the host cell is *S. roseosporus*, *S. lividans* or *S. fradiae*.

20 [0156] Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant
25 production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. See, e.g., Ausubel et al., *supra*, and Sambrook et al., *supra*, and Kieser et al., *supra*, herein incorporated by reference.

30 Polypeptides

Thioesterases and Fragments Thereof

- 71 -

[0157] Another object of the invention is to provide a polypeptide derived from a thioesterase involved in daptomycin synthesis. In one embodiment, the polypeptide is derived from a daptomycin biosynthetic gene cluster.

5 In a preferred embodiment, the polypeptide is derived from an integral or free thioesterase. In a more preferred embodiment, the polypeptide comprises the thioesterase domain of DptD or the amino acid sequence of DptH. In an even more preferred embodiment, the
10 polypeptide comprises the amino acid sequence of the thioesterase domain of SEQ ID NO: 7 or the amino acid sequence of SEQ ID NO: 8. The polypeptide derived from a thioesterase may also be encoded by an *S. roseosporus* nucleic acid sequence from any one of BAC clones
15 B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or B12:03A05, preferably from B12:03A05. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically
20 synthesized following the teachings of the specification and using methods well known to those having ordinary skill in the art. See, e.g., Examples 3-6.

[0158] The polypeptide may comprise a fragment of a thioesterase as defined herein. A polypeptide that
25 comprises only a part or fragment of the entire thioesterase may or may not encode a polypeptide that has thioesterase activity. A polypeptide that does not have thioesterase activity, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless
30 useful, especially for immunizing animals to prepare anti-thioesterase antibodies. However, in a preferred embodiment, the part or fragment encodes a polypeptide having thioesterase activity. Methods of determining

- 72 -

whether a polypeptide has thioesterase activity are described *infra*. Further, in a preferred embodiment, the fragment comprises an amino acid sequence comprising the GX SXG thioesterase motif (see Example 3). In a more preferred embodiment, the fragment comprises an amino acid sequence comprising the thioesterase motif GWSFG or GTSLG, which are derived from the thioesterase domain of SEQ ID NO: 7 or the amino acid sequence of SEQ ID NO: 8, respectively.

- 10 [0159] One can produce fragments of a polypeptide encoding a thioesterase by truncating the DNA encoding the thioesterase and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide.
- 15 One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art (see, e.g., Sambrook et al. and Ausubel et al., *supra*). In one
- 20 embodiment, a polypeptide comprising only a part or fragment of a thioesterase may be produced by chemical or enzymatic cleavage of a thioesterase. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of
- 25 the thioesterase in a host cell.

Daptomycin NRPS Polypeptides, and Subunits and Fragments Thereof

- [0160] Another object of the invention is to provide a polypeptide derived from a daptomycin NRPS or subunit thereof. The daptomycin NRPS comprises the subunits DptA, DptBC and DptD. As discussed in greater detail in Examples 3-6 below, each subunit comprises a number of modules that bind and activate specific building block
- 30

- 73 -

substrates and to catalyze peptide chain formation and elongation. Further, each module comprises a number of domains that participate in condensation, adenylation and thiolation. In addition, some modules comprise a
5 epimerization domain, discussed in greater detail in Example 6. DptD also comprises a thioesterase domain, as discussed *supra* and in Example 5.

[0161] In one embodiment, the polypeptide comprises an amino acid sequence from DptA, DptBC and/or DptD. In an
10 even more preferred embodiment, the polypeptide comprises an amino acid sequence SEQ ID NOS: 9, 11 or 7. A daptomycin NRPS polypeptide may also be encoded by an *S. roseosporus* nucleic acid sequence from any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04,
15 B12:20C09 or B12:03A05, preferably from B12:03A05. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the
20 specification and using methods well known to those having ordinary skill in the art. See, e.g., Examples 3-6 regarding amino acid sequences as well as modules and domains of DptA, DptBC and DptD.

[0162] The polypeptide may comprise a fragment of a
25 daptomycin NRPS as defined herein. In one embodiment, a fragment comprises one or more complete modules of a daptomycin NRPS subunit. In another embodiment, a fragment comprises one or more domains of a daptomycin NRPS subunit. In yet another embodiment, a fragment may
30 not comprise a complete domain or module but may comprise only a part of one or more domains or modules. A polypeptide that does not comprise a full domain or module of a daptomycin NRPS, whether it is a fragment,

- 74 -

analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-thioesterase antibodies. In a more preferred embodiment, the fragment comprises an amino acid sequence comprising at least that part of an adenylation domain that is required for binding to an amino acid. This part of the domain is delimited by the amino acid pocket code of a particular adenylation domain, as discussed below in Example 5.

10 **[0163]** As discussed above, one can produce fragments of a polypeptide of the invention recombinantly, by chemical synthesis or by enzymatic cleavage.

Polypeptides from S. roseosporus BAC Clones

[0164] Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule or part thereof from a *S. roseosporus* BAC clone of the invention. In one embodiment, the invention provides a polypeptide encoded by a nucleic acid molecule or part thereof from 1G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or, preferably, B12:03A05. In a preferred embodiment, the invention provides a polypeptide comprising an amino acid sequence SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136 or encoded by a nucleic acid molecule comprising the nucleic acid sequence SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133 or 135. In another preferred embodiment, the invention provides a

- 75 -

polypeptide that is DptE or DptF, a polypeptide having an amino acid sequence of SEQ ID NO: 15 or SEQ ID NO: 17, or encoded by *dptE* or *dptF*, or encoded by a nucleic acid sequence of SEQ ID NO: 16 or SEQ ID NO: 18. In another preferred embodiment, the invention provides an ABC transporter comprising an amino acid sequence SEQ ID NOS: 19, 21, 29, 45, 47, 49, 63, 67, 75 and 77, or encoded by a nucleic acid sequence of SEQ ID NOS: 20, 22, 30, 46, 48, 50, 64, 68, 76 or 78. In another preferred embodiment, the invention provides a polypeptide that is an oxidoreductase, such as a dehydrogenase; a transcriptional regulator involved in antibiotic resistance; NovABC-related polypeptides, which are involved in the biosynthesis of novobiocin, an antimicrobial agent; a monooxygenase; an acyl CoA thioesterase; a DNA helicase; or a DNA ligase, such as provided by a polypeptide having an amino acid sequence selected from SEQ ID NOS: 23, 25, 27, 29, 33, 35, 37, 91, 93, 97 and 99. In another preferred embodiment, the invention provides a polypeptide that is highly homologous to a *Streptomyces* polypeptide, such as provided by a polypeptide having an amino acid sequence selected from SEQ ID NOS: 61, 65, 69, 79, 81, 83, 85, 87, 95 and 101. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well known to those having ordinary skill in the art. The invention also provides a polypeptide that comprises a fragment of a nucleic acid molecule that encodes a polypeptide from a BAC clone, as defined herein. As discussed above, one can produce fragments of a

polypeptide of the invention recombinantly, by chemical synthesis or by enzymatic cleavage.

Muteins, Homologous Proteins, Allelic Variants, Analogs and Derivatives

5 **[0165]** Another object of the invention is to provide polypeptides that are mutant proteins (muteins), fusion proteins, homologous proteins or allelic variants of the daptomycin NRPS, subunits thereof, thioesterases or the polypeptides encoded by the *S. roseosporus* BAC nucleic
10 acid molecules or parts thereof provided herein. A mutant thioesterase may have the same or different enzymatic activity compared to a naturally-occurring thioesterase and comprises at least one amino acid insertion, duplication, deletion, rearrangement or
15 substitution compared to the amino acid sequence of a native protein. In one embodiment, the mutein has the same or a decreased thioesterase activity compared to a naturally-occurring thioesterase. In another embodiment, the mutant thioesterase has an increased thioesterase
20 activity compared to a naturally-occurring thioesterase. In a preferred embodiment, muteins of thioesterases of a daptomycin biosynthetic gene cluster may be used to alter thioesterase activity. See, e.g., Examples 12 and 13. In another embodiment, a mutant daptomycin NRPS or
25 subunit thereof may have the same or different amino acid specificity, thiolation activity, condensation activity, or, if present, epimerization activity, as a naturally-occurring daptomycin NRPS. Daptomycin NRPS muteins may be used to alter amino acid recognition, binding,
30 epimerization or other catalytic properties of an NRPS. See, e.g., Examples 12 and 16. Similarly, a mutein of a polypeptide encoded by the *S. roseosporus* BAC nucleic acid molecule of the invention may have a similar

- 77 -

biological activity or a different one, but preferably has a similar biological activity.

[0166] A mutein of the invention may be produced by isolation from a naturally-occurring mutant microorganism or from a microorganism that has been experimentally mutagenized, may be produced by chemical manipulation of a polypeptide, or may be produced from a host cell comprising an altered nucleic acid molecule. In a preferred embodiment, the mutein is produced from a host cell comprising an altered nucleic acid molecule. Muteins may also be produced chemically by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. One may produce muteins of a polypeptide by introducing mutations into the nucleic acid sequence encoding a daptomycin NRPS, subunit thereof or a thioesterase, or into a *S. roseosporus* BAC nucleic acid molecule, and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity, such as thioesterase activity, amino acid specificity, thiolation activity, epimerization activity, or condensation activity, as described below. Muteins may also be screened, e.g., for oxidoreductase activity, ABC transporter activity, monooxygenase activity, or DNA ligase or helicase activity using methods known in the art. Multiple random mutations can be introduced into the gene by methods well-known to the art, e.g., by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis,

cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well known in the art. See, e.g., Sambrook et al., *supra*, Ausubel et al., *supra*, U.S. Pat. No. 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

[0167] The invention also provides a polypeptide that is homologous to a daptomycin NRPS, subunit thereof, a thioesterase from a daptomycin biosynthetic gene cluster, or to a polypeptide encoded by a *S. roseosporus* BAC nucleic acid molecule as described herein. In one embodiment, the polypeptide is homologous to the thioesterase domain of DptD or to DptH, or to a polypeptide encoded by the thioesterase domain of *dptD* or by *dptH*. In a preferred embodiment, the polypeptide is homologous to a thioesterase having the amino acid sequence of the thioesterase domain of SEQ ID NO: 7 or having the amino acid sequence of SEQ ID NO: 8. In another embodiment, the polypeptide is homologous to DptA, DptBC or DptD, or to a polypeptide encoded by *dptA*, *dptBC* or *dptD*. In a more preferred embodiment, the polypeptide is homologous to a polypeptide having the amino acid sequence of SEQ ID NO: 9, 11 or 3. The invention also provides a polypeptide that is homologous to a polypeptide encoded by a nucleic acid molecule from a *S. roseosporus* BAC clone described herein, e.g., 1G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or, preferably, B12:03A05. In a preferred embodiment, the invention provides a polypeptide homologous to a polypeptide comprising an amino acid sequence of SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71,

- 79 -

73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136 or encoded by a nucleic acid molecule comprising a nucleic acid sequence selected

5 from SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133 or 135.

10 **[0168]** In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a polypeptide of the invention. In a more preferred embodiment, the homologous polypeptide is one that exhibits at least 50%, 60%, 70%, or 80% sequence

15 identity to a polypeptide comprising an amino acid sequence of SEQ ID NOS: 9, 11, 7 or 8 or SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104,

20 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, 90%, 95%, 96%, 97%, 98% or 99% sequence identity to a polypeptide comprising an amino

25 acid sequence of SEQ ID NOS: 9, 11, 7 or 8 or SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126,

30 128, 130, 132, 134 or 136.

[0169] The homologous protein may be a naturally-occurring one that is derived from another species, especially one derived from another *Streptomyces* species,

or one derived from another *Streptomyces roseosporus* strain, wherein the homologous protein comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NOS: 9, 11, 7 or 8 or SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136. The naturally-occurring homologous protein may be isolated directly from the other species or strain. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous protein may be isolated and used to express the homologous protein recombinantly. In another embodiment, the homologous protein may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous protein may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of the polypeptide.

[0170] In another embodiment, the invention provides a polypeptide encoded by an allelic variant of a gene encoding a thioesterase from a daptomycin biosynthetic gene cluster, or a daptomycin NRPS or subunit thereof. In a preferred embodiment, the invention provides a polypeptide encoded by an allelic variant of *dptA*, *dptBC*, *dptD* or *dptH*. In an even more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence of SEQ ID NOS: 9, 11, 7 or 8. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene, wherein the gene has the nucleic acid

- 81 -

sequence of SEQ ID NOS: 10, 12, 3 or 6. An allelic variant may have the same or different biological activity as the thioesterase, daptomycin NRPS or subunit thereof, described herein. In a preferred embodiment, an allelic variant is derived from another species of *Streptomyces*, even more preferably from a strain of *Streptomyces roseosporus*. In another embodiment, the invention provides a polypeptide encoded by an allelic variant of an *S. roseosporus* nucleic acid sequence from any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or B12:03A05, preferably from B12:03A05. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence of SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136, or that is encoded by an allelic variant of a gene, wherein the gene has a nucleic acid sequence of SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133 or 135.

[0171] In another embodiment, the invention provides a derivative of a polypeptide of the invention. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, e.g., radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H . In another preferred embodiment, the derivative has been labeled

- 82 -

with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand. In a preferred embodiment, the polypeptide is a thioesterase involved in the biosynthesis of daptomycin. In an even more preferred embodiment, the polypeptide comprises the thioesterase domain of DptD or comprises the amino acid sequence of DptH, or is a thioesterase encoded by the thioesterase-encoding domain of *dptD* or by *dptH*. In another preferred embodiment, the polypeptide is a daptomycin NRPS or subunit thereof, more preferably DptA, DptBC or DptD, even more preferably a polypeptide encoded by *dptA*, *dptBC* or *dptD*. In a yet more preferred embodiment, the polypeptide has an amino acid sequence of SEQ ID NOS: 9, 11, 7 or 8 or is a mutein, allelic variant, homologous protein or fragment thereof. Preferably, a thioesterase derivative has a thioesterase activity that is the same or similar to a thioesterase involved in the biosynthesis of daptomycin, more preferably, the derivative has a thioesterase activity that is the same or similar to a thioesterase having an amino acid sequence of the thioesterase domain of SEQ ID NO: 7 or having the amino acid sequence of SEQ ID NO: 8. In another preferred embodiment, a daptomycin NRPS or NRPS subunit derivative has the same or similar activity as a naturally-occurring daptomycin NRPS or subunit thereof. In yet another embodiment, the derivative is derived from a polypeptide encoded by a nucleic acid molecule from a *S. roseosporus* nucleic acid sequence from any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or, preferably, B12:03A05. In a preferred embodiment, the derivative is derived from a polypeptide having an amino acid sequence of SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31,

- 83 -

33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59,
61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87,
89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114,
116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136,
5 or that is encoded by a gene having a nucleic acid
sequence of SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34,
36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62,
64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90,
92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115,
10 117, 119, 121, 123, 125, 127, 129, 131, 133 or 135.

[0172] The invention also provides non-peptide
analogs. In a preferred embodiment, the non-peptide
analog is structurally similar to a thioesterase involved
in daptomycin synthesis, to a daptomycin NRPS or subunit
15 thereof, or to a polypeptide encoded by a nucleic acid
molecule from an *S. roseosporus* BAC clone, but in which
one or more peptide linkages is replaced by a linkage
selected from the group consisting of --CH₂NH--, --CH₂S--,
--CH₂-CH₂--, --CH=CH--(cis and trans), --COCH₂--,
20 --CH(OH)CH₂-- and -CH₂SO--. In another embodiment, the
non-peptide analog comprises substitution of one or more
amino acids of a thioesterase or daptomycin NRPS or
subunit thereof with a D-amino acid of the same type in
order to generate more stable peptides. Preferably, both
25 a non-peptide and a peptide analog has a biological
activity that is the same or similar to the naturally-
occurring polypeptide involved in the biosynthesis of
daptomycin, more preferably, the analog has a biological
activity that is the same or similar to the polypeptide
30 having an amino acid sequence of SEQ ID NOS: 9, 11, 7 or
8. The invention also provides analogs of polypeptides
encoded by an *S. roseosporus* nucleic acid sequence from

any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or B12:03A05, preferably from B12:03A05. The invention provides an analog of a polypeptide having an amino acid sequence of SEQ ID NOS:

5 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136, or that is encoded by a gene

10 having a nucleic acid sequence of SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129,

15 131, 133 or 135.

Fusion Proteins

[0173] The polypeptides of this invention may be fused to other molecules, such as genetic, enzymatic or chemical or immunological markers such as epitope tags.

20 Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β -galactosidase, biotin *trpE*, protein A, β -lactamase, α -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl

25 terminus of the polypeptide), *lacZ*, green fluorescent protein (GFP), yeast α mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. See, e.g., Godowski et al., 1988, and

30 Ausubel et al., *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art.

- 85 -

Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques such as those described in Merrifield, 1963, herein incorporated by
5 reference, or produced by chemical cross-linking.

[0174] Tagged fusion proteins permit easy localization, screening and specific binding via the epitope or enzyme tag. See Ausubel, 1991, Chapter 16. Some tags allow the protein of interest to be displayed
10 on the surface of a phagemid, such as M13, which is useful for panning agents that may bind to the desired protein targets. Another advantage of fusion proteins is that an epitope or enzyme tag can simplify purification. These fusion proteins may be purified, often in a single
15 step, by affinity chromatography. For example, a His⁶ tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or
20 Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the nucleic acid molecule of the
25 invention by an enzymatic cleavage site that can be cleaved after purification.

[0175] A second advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for
30 screening targets.

[0176] Therefore, in another aspect, the invention provides a fusion protein comprising all or a part of a thioesterase derived from a daptomycin biosynthetic gene

- 86 -

cluster and provides a nucleic acid molecule that encodes such a fusion protein. Another aspect provides a fusion protein comprising all or a part of a daptomycin NRPS or subunit thereof and provides a nucleic acid molecule

5 encoding such a protein. See, e.g., Examples 11-16. The invention also provides a fusion protein comprising all or part of a polypeptide encoded by a nucleic acid molecule from any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or B12:03A05. In a

10 preferred embodiment, the fusion protein comprises all or a part of a polypeptide encoded by one or more of *dptA*, *dptBC*, *dptD* or *dptH*. In another preferred embodiment, the fusion protein comprises a polypeptide encoded by a nucleic acid molecule that selectively hybridizes to

15 *dptA*, *dptBC*, *dptD* or *dptH*. In a more preferred embodiment, the fusion protein comprises a polypeptide having an amino acid sequence of SEQ ID NOS: 9, 11, 7 or 8, or comprises a polypeptide that is a fragment, mutein, homologous protein, derivative or analog thereof. In an

20 even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NOS: 10, 12, 3 or 6, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid

25 molecule comprising said nucleic acid sequence. The invention also provides fusion proteins comprising polypeptide sequences encoded by an *S. roseosporus* nucleic acid sequence from any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or

30 B12:03A05, preferably from B12:03A05. The invention provides a fusion protein comprising a polypeptide having an amino acid sequence of SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55,

- 87 -

57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136, or comprising a polypeptide that is a fragment, 5 mutein, homologous protein, derivative or analog thereof. The invention also provides a fusion protein comprising a polypeptide encoded by SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 10 86, 88, 90, 92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133 or 135, or comprising all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising said nucleic acid sequence.

15 **[0177]** In one aspect of the invention, the fusion protein that comprises all or a part of a thioesterase derived from a daptomycin biosynthetic gene cluster comprises other modules (including heterologous or hybrid modules) from a polypeptide involved in non-ribosomal 20 protein synthesis. See, e.g., Examples 12E, G and H and Example 13. In another preferred embodiment, the fusion protein comprises one or more amino acid sequences that encode thioesterases, wherein the thioesterases may be identical to one another or may be different. See, e.g., 25 Examples 11E-G (duplication of daptomycin thioesterase genes), Example 12 (producing modified NRPS thioesterase fusion proteins) and Example 13 (producing free thioesterase fusion proteins).

[0178] In another embodiment, the invention provides a 30 fusion protein that is a hybrid of amino acid sequences from two or more different thioesterases and a nucleic acid molecule that encodes such a fusion protein. The hybrid fusion protein may consist of two, three or more

portions of different thioesterases. The hybrid thioesterase may have a different or the same specificity.

5 *Methods to Assay Thioesterase and Daptomycin NRPS Activity*

[0179] There are a number of methods known in the art to determine whether a fragment, mutein, homologous protein, analog, derivative or fusion protein of a thioesterase has the same, enhanced or decreased
10 biological activity as a wild-type thioesterase polypeptide. In one embodiment, a thioesterase assay which monitors cleavage of a suitable thioester bond and/or release of a corresponding product is performed *in vitro*. Any of a number of thioesterase assays well-known
15 in the art may be used, including those which use photo- or radio-labeled substrates.

[0180] In a preferred embodiment, thioesterase activity associated with peptide synthesis by a NRPS is determined using cellular assays. For example, a nucleic
20 acid molecule encoding a fragment, mutein, homologous protein or fusion protein may be introduced into a bacterial cell comprising a daptomycin biosynthetic gene cluster absent one or both of the thioesterase domains of *dptD* or *dptH*. Alternatively, the nucleic acid molecule
25 may be introduced into a bacterial cell comprising a different biosynthetic gene cluster that produces a different compound, e.g., a different lipopeptide. In a preferred embodiment, the bacterial cell may be *S. lividans*. The nucleic acid molecule may be introduced
30 into the bacterial cell by any method known in the art, including conjugation, transformation, electroporation, protoplast fusion or the like. The bacterial cell comprising the nucleic acid molecule is incubated under

conditions in which the polypeptide encoded by the nucleic acid molecule is expressed. After incubation, the bacterial cells may be analyzed by, e.g., HPLC and/or LC/MS, to determine if the bacterial cells produce the
5 desired lipopeptide. See, e.g., the method of expressing daptomycin described in Examples 7- 9, *infra*. When the thioesterase activity is associated with synthesis of a peptide having an anti-cell growth property (e.g., an antibiotic, antifungal, antiviral or antimitotic agent) a
10 desired assay known to those of skill in the art may be used.

[0181] Alternatively, a fragment, mutein, homologous protein, analog, derivative or fusion protein of a thioesterase may be introduced into a cell, particularly
15 a bacterial cell, comprising a daptomycin biosynthetic gene cluster absent one or both of the thioesterase domain of *dptD* or *dptH*. After incubation, the bacterial cells may be analyzed by, e.g., HPLC and/or LC/MS, as described in Example 7, to determine if the bacterial
20 cells produce the desired lipopeptide. The same method can be used with a cell comprising a different biosynthetic gene cluster that produces a different compound, e.g., a different lipopeptide.

[0182] In a preferred embodiment, a fragment, mutein, homologous protein, analog, derivative or fusion protein
25 comprises an amino acid sequence comprising the GX SXG thioesterase motif (see Example 3). In a more preferred embodiment, a fragment, mutein, homologous protein, analog or derivative comprises an amino acid sequence
30 comprising the thioesterase motif GWSFG or GTSLG, which are derived from SEQ ID NO: 7 and SEQ ID NO: 8, respectively.

[0183] Similar methods known in the art may be used to determine whether a fragment, mutein, homologous protein, analog, derivative or fusion protein of a daptomycin NRPS or subunit thereof has the same or different biological activity as a wild-type NRPS or subunit thereof.

Antibodies

[0184] The polypeptides encoded by the genes of this invention may be used to elicit polyclonal or monoclonal antibodies that bind to a polypeptide of this invention, as well as a fragment, mutein, homologous protein, analog, derivative or fusion protein thereof, using a variety of techniques well known to those of skill in the art. Antibodies directed against the polypeptides of this invention are immunoglobulin molecules or portions thereof that are immunologically reactive with the polypeptide of the present invention.

[0185] Antibodies directed against a polypeptide of the invention may be generated by immunization of a mammalian host. Such antibodies may be polyclonal or monoclonal. Preferably they are monoclonal. Methods to produce polyclonal and monoclonal antibodies are well known to those of skill in the art. For a review of such methods, see Harlow and Lane, Antibodies: A Laboratory Manual (1988) and Ausubel et al. *supra*, herein incorporated by reference. Determination of immunoreactivity with a polypeptide of the invention may be made by any of several methods well known in the art, including by immunoblot assay and ELISA.

[0186] Monoclonal antibodies with affinities of 10^{-8} M⁻¹ or preferably 10^{-9} to 10^{-10} M⁻¹ or stronger are typically made by standard procedures as described, e.g., in Harlow and Lane, 1988. Briefly, appropriate animals are selected and the desired immunization protocol followed.

- 91 -

After the appropriate period of time, the spleens of such animals are excised and individual spleen cells fused, typically, to immortalized myeloma cells under appropriate selection conditions. Thereafter, the cells
5 are clonally separated and the supernatants of each clone tested for their production of an appropriate antibody specific for the desired region of the antigen.

[0187] Other suitable techniques involve *in vitro* exposure of lymphocytes to the antigenic polypeptides, or
10 alternatively, to selection of libraries of antibodies in phage or similar vectors. See Huse et al., 1989. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, polypeptides and antibodies will be labeled by joining,
15 either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides,
20 enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like. Patents teaching the use of such labels include U.S. Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241, herein
25 incorporated by reference. Also, recombinant immunoglobulins may be produced (see U.S. Patent 4,816,567, herein incorporated by reference).

[0188] An antibody of this invention may also be a hybrid molecule formed from immunoglobulin sequences from
30 different species (e.g., mouse and human) or from portions of immunoglobulin light and heavy chain sequences from the same species. An antibody may be a single-chain antibody or a humanized antibody. It may be

a molecule that has multiple binding specificities, such as a bifunctional antibody prepared by any one of a number of techniques known to those of skill in the art including the production of hybrid hybridomas, disulfide exchange, chemical cross-linking, addition of peptide linkers between two monoclonal antibodies, the introduction of two sets of immunoglobulin heavy and light chains into a particular cell line, and so forth.

5 [0189] The antibodies of this invention may also be human monoclonal antibodies, for example those produced by immortalized human cells, by SCID-hu mice or other non-human animals capable of producing "human" antibodies, or by the expression of cloned human immunoglobulin genes. The preparation of humanized
10 antibodies is taught by U.S. Pat. Nos. 5,777,085 and 5,789,554, herein incorporated by reference.

[0190] In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological
15 properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular
20 application.

[0191] In a preferred embodiment, an antibody of the present invention binds to a thioesterase involved in daptomycin synthesis or to a daptomycin NRPS or subunit thereof. In a more preferred embodiment, the antibody
25 binds to a polypeptide encoded by *dptA*, *dptBC*, *dptD* or *dptH*, or to a fragment thereof. In another preferred embodiment, the antibody binds to a polypeptide encoded by a nucleic acid molecule that selectively hybridizes to
30

- 93 -

dptA, *dptBC*, *dptD* or *dptH*. In a more preferred embodiment, the antibody binds to a polypeptide having an amino acid sequence of SEQ ID NOS: 9, 11, 7 or 8, or binds to a polypeptide that is fragment, mutein, homologous protein, derivative, analog or fusion protein thereof. In an even more preferred embodiment, the antibody binds to a polypeptide encoded by a nucleic acid molecule comprising all or part of the nucleic acid sequence of SEQ ID NOS: 10, 12, 3 or 6. In another embodiment, the antibody binds to a polypeptide encoded by a nucleic acid molecule that comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NOS: 10, 12, 3 or 6.

15 **[0192]** The invention provides an antibody that selectively binds to a polypeptide encoded by an *S. roseosporus* nucleic acid sequence from any one of BAC clones B12:01G05, B12:06A12, B12:12F06, B12:18H04, B12:20C09 or B12:03A05, preferably from B12:03A05. The polypeptide may comprise an amino acid sequence selected from SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136 or is encoded by a nucleic acid sequence SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133 or 135. Preferably, the antibody selectively binds to a polypeptide comprising an amino acid sequence selected from SEQ ID NOS: 23, 25, 27, 29, 33, 35, 37, 91, 93, 97,

- 94 -

99, 110 or 112 or from SEQ ID NOS: 61, 65, 69, 79, 81, 83, 85, 87, 95 and 101. The invention also provides an antibody that selectively binds to a fragment, mutein, homologous protein, derivative, analog or fusion protein thereof.

Computer Readable Means

[0193] A further aspect of the invention is a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing all of the nucleic acid and amino acid sequences described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

[0194] The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

[0195] This invention provides computer readable media having stored thereon sequences of the invention. A

- 95 -

computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

[0196] Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology

- 96 -

analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

5 [0197] A computer-based method is provided for performing nucleic acid homology identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence a nucleic acid of the invention in a computer readable medium; and comparing
10 said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify homology.

[0198] A computer-based method is also provided for performing amino acid homology identification, said
15 method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

20 [0199] A computer based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid
25 of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

30 Methods of Using Nucleic Acid Molecules as Probes and Primers

[0200] In one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a nucleic acid molecule that

- 97 -

selectively hybridizes to the nucleic acid molecule. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a daptomycin NRPS, subunit thereof or thioesterase from a daptomycin biosynthetic gene cluster. The probe or primer may also be derived from an expression control sequence derived from a daptomycin NRPS or thioesterase gene of a daptomycin biosynthetic gene cluster. In a preferred embodiment, the probe or primer is derived from *dptA*, *dptBC*, *dptD* or *dptH*. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule that encodes a polypeptide having an amino acid sequence of SEQ ID NOS: 9, 11, 7 or 8. In a yet more preferred embodiment, the probe or primer is derived from a nucleic acid molecule that has a nucleic acid sequence of SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133 or 135. In another embodiment, the probe or primer is derived from a nucleic acid sequence that encodes SEQ ID NOS: 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134 or 136.

[0201] In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22

- 98 -

nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. See, e.g., Sambrook et al., *supra*. See, e.g., Chapter 11 and pages 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pages 11.45-11.53, which describes hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pages 11.50-11.51). Methods of performing PCR using primers are also well-known in the art. See, e.g., Sambrook et al., *supra* and Ausubel et al., *supra*. PCR methods may be used to identify and/or isolate allelic variants and fragments of the nucleic acid molecules of the invention; PCR may also be used to identify and/or isolate nucleic acid molecules that hybridize to the primers and that may be amplified, and may be used to isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention.

Methods of Using Thioesterases for Biosynthesis of Compounds -Manipulations of *Dpt* Genes

[0202] Genes of the daptomycin biosynthetic gene cluster of the invention may be manipulated in a variety of ways to produce new biosynthetic peptide products or to alter the regulation of one or more genes expressed from the gene cluster. See, e.g., Figure 1.

Disruption of a Gene Encoding a Thioesterase

[0203] In one aspect, the invention provides a method of disrupting or deleting a gene encoding a thioesterase that is involved in a NRPS or PKS pathway in a bacterial cell. Preferably, the method comprises the step of

- 99 -

disrupting or deleting a gene or portion thereof that encodes a thioesterase in a daptomycin biosynthetic gene cluster. Disruption or deletion of a gene encoding an integral thioesterase would be likely to result in the production of compounds that are intermediates to the final product. In one aspect, a gene or portion thereof encoding an integral thioesterase may be disrupted or deleted. In a preferred embodiment, disruption or deletion of a gene encoding an integral thioesterase of the daptomycin biosynthetic gene cluster in *S. roseosporus* would produce a linear lipopeptide compound. The linear lipopeptide compound may be used directly if its release from the NRPS were to be catalyzed by a different endogenous or exogenously provided thioesterase activity within the host cell. Such linear lipopeptide compounds, if not released from the NRPS by an endogenous thioesterase activity, may be useful intermediates for testing potential but as yet unidentified thioesterase polypeptides or for testing thioesterase fusion, fragment, mutein, derivative, analog or homolog polypeptides for activity. The linear lipopeptide compound may alternatively be used as an intermediate for production of novel lipopeptides.

[0204] In another aspect, a gene encoding a free thioesterase may be disrupted or deleted in a bacterial cell comprising an NRPS. Because free thioesterases are thought to be involved in proofreading of the peptide compounds produced in NRPS, disruption or deletion of a gene encoding a free thioesterase may lead to the production of compounds that have mutations compared to the compound produced in the presence of the free thioesterase. These mutated compounds may be used to generate novel lipopeptides. See, e.g., Example 13.

- 100 -

[0205] In a preferred embodiment, the method comprises the step of disrupting or deleting the thioesterase-encoding portion of *dptD* or disrupting or deleting *dptH* in a daptomycin biosynthetic gene cluster. In an even
5 more preferred embodiment, the method comprises the step of disrupting or deleting a gene encoding a thioesterase having an amino acid sequence of the thioesterase domain of SEQ ID NO: 7 or having the amino acid sequence of SEQ ID NO: 8. The invention also comprises a method of
10 disrupting or deleting a gene encoding a thioesterase wherein the gene is one that selectively hybridizes or is homologous to a gene encoding a thioesterase having an amino acid sequence of the thioesterase domain of SEQ ID NO: 7 or the amino acid sequence of SEQ ID NO: 8. In
15 another preferred embodiment, disruption or deletion of a thioesterase may be combined with the methods of altering the gene cluster involved in non-ribosomal peptide synthesis, as described below.

[0206] Disruption of a gene encoding a thioesterase
20 may be accomplished by any method known to one having ordinary skill in the art following the teachings of the instant specification. In a preferred embodiment, disruption of a gene encoding a thioesterase may be accomplished by targeted gene disruption using methods
25 taught, e.g., in Hosted and Baltz, *J. Bacteriol.*, 179, pp. 180-186 (1997); Butler et al., *Chem. Biol.*, 6, pp. 287-292 (1999); and Xue et al., *Proc. Natl. Acad. Sci. U.S.A.*, 95, pp. 12111-12116 (1998), each of which is incorporated herein by reference in its entirety. See,
30 e.g., Example 11.

Alteration of Site of Cyclization and Cyclic Peptide Produced Using Thioesterases

- 101 -

[0207] In a naturally-occurring polypeptide involved in NRPS, an integral thioesterase is located at the carboxy-terminus of the polypeptide, where it is involved in product cyclization. In one aspect, the invention provides a method to alter the site of cyclization of a cyclic peptide (or release of a linear peptide) by changing the location of a module encoding a thioesterase. In one embodiment, the site of cyclization may be altered by inserting the module encoding the thioesterase into the gene encoding the polypeptide involved in NRPS in a region that is upstream of the region in which the thioesterase module naturally occurs. In this embodiment, the cyclic peptide that is produced will be smaller than the naturally-occurring cyclic peptide. See, e.g., Example 12.

[0208] In a preferred embodiment, the module encodes an integral thioesterase from a daptomycin biosynthetic gene cluster. In a more preferred embodiment, the module comprises the thioesterase domain of DptD. In an even more preferred embodiment, the module encodes a polypeptide having all or a portion of the amino acid sequence of SEQ ID NO: 7, preferably a portion of SEQ ID NO: 7 that comprises the thioesterase domain. In another preferred embodiment, the module comprises a nucleic acid molecule that is homologous to or selectively hybridizes to a nucleic acid molecule encoding all or a portion of the thioesterase domain of SEQ ID NO: 7 or to a nucleic acid molecule encoding the thioesterase domain that comprises all or a portion of the nucleic acid sequence of SEQ ID NO: 3.

[0209] Alternatively, other modules that are involved in adding amino acids to the peptide (or otherwise modifying amino acids within the peptide) may be inserted

- 102 -

upstream of the module encoding the thioesterase. See, e.g., Example 12. Such modules include a minimal module comprising at least an adenylation domain and a thiolation or acyl carrier domain. In a preferred embodiment, the inserted module would also include a condensation domain. Additional domains may also be inserted upstream of the thioesterase module including an M domain, an E domain and/or a Cy domain. The type of module(s) that would be inserted upstream of the thioesterase domain would depend upon the type of amino acid residues that were desired. Methods of inserting modules that will add and/or modify a specific amino acid are well known in the art. See, e.g., Mootz et al., *Curr. Opin. Biotechnol.*, 10, pp. 341-348 (1999), herein incorporated by reference in its entirety. Addition of one or more modules upstream of the thioesterase will produce a polypeptide involved in NRPS that is capable of synthesizing a cyclic peptide that is larger and that may contain different amino acid residues than the naturally-occurring cyclic peptide.

In Vivo Use of Thioesterases

[0210] Another use of the genes of the present invention is to improve the yield of a product in a cell expressing an NRPS. See, e.g., Example 11. Nucleic acid molecules that may be used to increase yield include nucleic acid molecules that encode positive regulatory factors, acyl CoA thioesterase, ABC transporters, NovABC-related polypeptides, DptA, DptBC, DptD, polypeptides that encode daptomycin resistance and daptomycin thioesterases, including DptD and DptH. The complete

- 103 -

daptomycin biosynthetic gene cluster, daptomycin NRPS or any domain or subunit thereof may also be duplicated. In a preferred embodiment, a free and/or an integral thioesterase from a daptomycin biosynthetic gene cluster are introduced into a cell to improve production of daptomycin. In another preferred embodiment, the additional copies of a thioesterase may be introduced into a cell comprising altered NRPS polypeptides, as described *supra*. Without wishing to be bound by any theory, additional copies of a free and/or an integral thioesterase may improve the NRPS processing of the peptide by increasing the proofreading capacity (e.g., the free thioesterase) or the cyclization and/or peptide release capacity (e.g., the integral thioesterase) of the bacterial cell.

[0211] In a preferred embodiment, additional copies of a nucleic acid molecule encoding thioesterase may be introduced into a cell. See, e.g., Example 11. Introduction of the thioesterase may be performed by any method known in the art. In a more preferred embodiment, the additional copies of the gene are under the regulatory control of strong expression control sequences. These sequences may be derived from another thioesterase gene or may be derived from heterologous sequences, as described *supra*. Further, a nucleic acid molecule encoding a thioesterase may be introduced into a cell such that it is expressed as a separate polypeptide. This may be especially useful for a free thioesterase. Alternatively, a nucleic acid molecule encoding a thioesterase may be introduced into a cell such that it forms part of a multi-domain protein. This can be accomplished, e.g., by homologous recombination into a polypeptide which forms or interacts with an NRPS. This

- 104 -

may be especially useful, although not required, for an integral thioesterase.

[0212] In another embodiment, copies of a free and/or an integral thioesterase may be introduced into a cell that expresses a NRPS complex that is other than a daptomycin biosynthetic gene cluster. See, e.g., Example 13. In one preferred embodiment, the complex is a NRPS complex. In another preferred embodiment, the complex is a PKS complex or a mixed PKS/NRPS complex. Numerous PKS and NRPS complexes are known in the art. See, e.g., complexes that produce vancomycin, bleomycin, A54145, CDA, amphomycin, echinocandin, cyclosporin, erythromycin, tylosin, monensin, avermectin, penicillin, cephalosporin, pristnamycins, erythromycin, rapamycin, spinosyn, didemnin, discobahamian, and epothilone. As described above, addition of a free and/or an integral thioesterase may improve the NRPS or PKS processing of a peptide by increasing the proofreading capacity (the free thioesterase) or the cyclization capacity (the integral thioesterase) of the bacterial cell. Addition of a free and/or integral thioesterase may be achieved by the methods described above.

[0213] In a preferred embodiment, a nucleic acid molecule encoding a thioesterase that is introduced into a cell is a thioesterase from a daptomycin biosynthetic gene cluster. In a preferred embodiment, the gene is the thioesterase-encoding domain of *dptD* or is *dptH*. More preferably, the nucleic acid molecule encodes a thioesterase having an amino acid sequence of the thioesterase domain of SEQ ID NO: 7 or SEQ ID NO: 8, or is a homologous protein, fusion protein, mutein, derivative, analog or fragment thereof having thioesterase activity.

Methods of Altering Gene Clusters for Production of Novel Compounds by NRPS

Alteration of NRPS Polypeptide Modules and Domains

- 5 **[0214]** In another aspect, the invention provides a method of altering the number or position of the modules in an NRPS. In one embodiment, one or more modules may be deleted from the NRPS. These deletions will result in synthesis by the NRPS of a peptide product that is shorter than the naturally-occurring one. In another
- 10 embodiment, one or more domains may be deleted from the NRPS. In this case, the product produced by the NRPS will have a chemical change relative to the peptide produced in the absence of the deletion, e.g., if an epimerization and/or methylation domain is deleted.
- 15 **[0215]** In another embodiment, one or more modules or domains may be added to the NRPS. In this case, the peptide synthesized by the NRPS will be longer than the naturally-occurring one or will have an additional chemical change, respectively. For instance, if an
- 20 epimerization domain or a methylation domain is added, the resultant peptide will contain an extra D-amino acid or will contain a methylated amino acid, respectively. In a yet further embodiment, one or more modules may be mutated, e.g., an adenylation domain may be mutated such
- 25 that it has a different amino acid specificity than the naturally-occurring adenylation domain. The amino acid pocket code for the daptomycin NRPS - which determines which amino acid will bind within each adenylation domain of modules 1-13 - is described in Example 5; see also
- 30 Table 2. With the amino acid code in hand, one of skill in the art can perform mutagenesis, by a variety of well known techniques, to exchange the code in one module for another code, thus altering the ultimate amino acid

- 106 -

composition and/or sequence of the resulting peptide synthesized by the altered NRPS. See, e.g., Example 12A. In another embodiment, one or more subunits may be added or deleted to the NRPS.

- 5 **[0216]** In a still further embodiment, one or more domains, modules or subunits may be substituted with another domain, module or subunit in order to produce novel peptides by complementation. In this case, the peptide produced by the altered NRPS will have, e.g., one
10 or more different amino acids compared to the naturally-occurring peptide. In addition, different combinations of insertions, deletions, substitutions and mutations of domains, modules or subunits may be used to produce a peptide of interest. For instance, one may substitute a
15 modified module, domain or subunit for a naturally-occurring one, or may substitute a naturally-occurring module, domain or subunit from the NRPS from one organism for a module, domain or subunit of an NRPS from another organism. See, e.g., Example 12C. Modifications of the
20 modules, domains and subunits may be performed by site-directed mutagenesis, domain exchange (for module or subunit modification), deletion, insertion or substitution of a domain in a module or subunit, or deletion, insertion or substitution of a module in a
25 subunit. Further, a domain, module or subunit may be disrupted such that it does not function using any method known in the art. These disruptions include, e.g., such techniques as a single crossover disruptant or replacement through homologous recombination by another
30 gene (e.g., a gene that permits selection or screening).
[0217] The products produced by the modified NRPS complexes will have different incorporated amino acids, different chemical alterations of the amino acids (e.g.,

- 107 -

methylation and epimerization) and may be shorter or longer than the native lipopeptides. The domains, modules or subunits may be derived from any number of NRPS desired, including two, three or four NRPS.

- 5 Further, the invention contemplates these altered NRPS complexes with and without an integral thioesterase domain. See, e.g., Example 12B-J.

[0218] The source of the modules, domains and/or subunits may be derived from the daptomycin biosynthetic gene cluster NRPS or may be derived from the NRPS that
10 encodes another lipopeptide or other peptide source. These peptide sources include glycopeptide gene clusters, mixed pathway gene clusters and siderophore gene clusters. Further, the source of the modules, domains
15 and/or subunits may be obtained from any appropriate source, including both streptomycete and non-streptomycete sources. Non-streptomycete sources include actinomycetes, e.g., Amycolatopsis; prokaryotic non-actinomycetes, e.g., Bacillus and cyanobacteria; and non-
20 bacterial sources, e.g., fungi.

[0219] An NRPS or portion thereof may be heterologous to a host cell of interest or may be endogenous to the host cell. In one embodiment, the daptomycin NRPS or a portion thereof (e.g., a domain, module or subunit
25 thereof) is introduced into the host cell on any vector known to one having ordinary skill in the art, e.g., a plasmid, a cosmid, bacteriophage or BAC. The host cell into which the daptomycin NRPS or portion thereof is introduced may contain an endogenous NRPS or portion
30 thereof (e.g., a domain, module or subunit thereof). Alternatively, a heterologous NRPS or portion thereof may be introduced into the host cell containing the heterologous daptomycin NRPS or portion thereof. The

- 108 -

daptomycin NRPS, other NRPS, or domain, module or subunit of an NRPS may have either a naturally-occurring sequence or a modified sequence. In another embodiment, the
daptomycin NRPS or portion thereof is endogenous to the
5 host cell, e.g., the host cell is *S. roseosporus*. A naturally-occurring or modified NRPS, or a domain, module or subunit thereof may be introduced into the host cell comprising the daptomycin NRPS or portion thereof. The heterologous domains, modules, subunits or NRPS may
10 comprise a constitutive or regulatable promoter, which are known to those having ordinary skill in the art. The promoter can be either homologous or heterologous to the nucleic acid molecule being introduced into the cell. In one embodiment, the promoter may be from the daptomycin
15 biosynthetic gene cluster, as described above.

[0220] The nucleic acid molecule comprising the NRPS or portion thereof (e.g., a domain, module or subunit) may be maintained episomally or integrated into the genome. The nucleic acid molecule may be introduced into
20 the genome at, e.g., phage integration sites. Further, the nucleic acid molecule may be introduced into the genome at the site of an endogenous or heterologous NRPS or portion thereof or elsewhere in the genome. The nucleic acid molecule may be introduced in such a way to
25 disrupt all or part of the function of a domain, module or subunit of an NRPS already present in the genome, or may be introduced in a manner that does not disturb the function of the NRPS or portion thereof.

[0221] The peptides produced by these NRPS may be
30 useful as new compounds or may be useful in producing new compounds. In a preferred embodiment, the new compounds are useful as or may be used to produce antibiotic compounds. In another preferred embodiment, the new

- 109 -

compounds are useful as or may be used to produce other peptides having useful activities, including but not limited to antibiotic, antifungal, antiviral, antiparasitic, antimitotic, cytostatic, antitumor, immuno-modulatory, anti-cholesterolemic, siderophore, agrochemical (e.g., insecticidal) or physicochemical (e.g., surfactant) properties. In a more preferred embodiment, the compounds produced using an altered NRPS polypeptide may be used in the synthesis of daptomycin-related compounds, including those described in United States Application Nos. 09/738,742, 09/737,908 and 09/739,535, filed December 15, 2000.

[0222] In addition, diverse variants of non-ribosomally synthesized peptides and polyketides may be achieved by altering the pools of available substrates during host cell cultivation. Commercial production of daptomycin, for example, is the result of cultivating the daptomycin producer *Streptomyces roseosporus* in the presence of decanoic acid, which alters the lipopeptide profile of the final products. See, e.g., United States Patent 4,885,243. The feeding of N-acetyl cysteamine (SNAC) analogs of polyketide intermediates resulted in substantial increases in incorporation of the intermediates into the polyketide, when compared to the free carboxylic acid or ester analogs. See, e.g., Yue et al., *J. Am. Chem. Soc.*, 109, pp. 1253-1255 (1987); Cane and Yang, *J. Am. Chem. Soc.*, 109, 1255-1257 (1987); Cane et al., *J. Am. Chem. Soc.*, 115, pp. 522-526 and 527-535 (1993); Cane et al., *J. Am. Chem. Soc.*, 117, pp. 633-634 (1995); Pieder et al., *J. Am. Chem. Soc.*, 117, pp. 11373-11374 (1995); each of which is incorporated herein by reference in its entirety. SNAC analogs of amino acids have been incorporated into a NRPS *in vitro*.

- 110 -

Ehmann et al., *Chem. Biol.*, 7, pp. 765-772 (2000). Thus it should be possible to feed SNAC or other pantetheine mimics to incorporate unnatural substrates into a NRPS-produced peptide.

5 **[0223]** Further diversity of non-ribosomally synthesized peptides and polyketides may also be achieved by expressing one or more NRPS and PKS genes (encoding natural, hybrid or otherwise altered modules or domains) in heterologous host cells, i.e., in host cells other
10 than those from which the NRPS and PKS genes or modules originated.

[0224] In addition, one may express an ABC transporter or other polypeptide involved in antibiotic resistance in order to increase the resistance of a bacterial cell to
15 daptomycin or a related compound. The ABC transporter may be overexpressed in an autologous cell (i.e., a cell that comprises the gene) or may be expressed in a heterologous cell (i.e., a cell that normally does not have the gene). Further, one may express an ABC
20 transporter gene of the invention or another polypeptide involved in antibiotic resistance described herein in order to be able to select cells that are resistant to daptomycin. This selection may be useful for determining mechanisms of daptomycin resistance or may be used in
25 standard molecular biological techniques in which antibody resistance is selected for.

Compounds Of The Invention, Pharmaceutical Compositions
Thereof And Methods Of Treating Using Compounds And
Compositions

30 **[0225]** Another object of the instant invention is to provide peptides or lipopeptides that may be produced by using the thioesterases, an NRPS or subunits thereof of the instant invention, as well as salts, esters, amides,

- 111 -

ethers and protected forms thereof, and pharmaceutical formulations comprising these peptides, lipopeptides or their salts. In a preferred embodiment, the lipopeptide is daptomycin or a daptomycin-related lipopeptide, as
5 described *supra*.

[0226] One may determine whether a peptide, lipopeptide or other compound of this invention has antibiotic activity using any of a variety of routine and well-known protocols in the art. One may use either an
10 isolated or purified compound or may use an unpurified compound that is present in, e.g., fermentation culture broth or in a cell lysate. One may use either or both a gram-positive or a gram-negative bacterial test strain, and may use a variety of test strains to determine
15 efficacy. In a preferred embodiment, the bacterial test strain will be a gram-positive test strain. In a more preferred embodiment, the bacterial test strain will be a *Staphylococcus*, more preferably *S. aureus*. An example of methods that can be used to determine antibiotic activity
20 are provided in United States Patents 4,208,408 and 4,537,717. One having ordinary skill in the art will recognize that other potential antibiotics and other test strains may be used.

[0227] Peptides, lipopeptides or pharmaceutically
25 acceptable salts thereof can be formulated for oral, intravenous, intramuscular, subcutaneous, aerosol, topical or parenteral administration for the therapeutic or prophylactic treatment of diseases, particularly bacterial infections. In a preferred embodiment, the
30 lipopeptide is daptomycin or a daptomycin-related lipopeptide. Reference herein to "daptomycin," "daptomycin-related lipopeptide" or "lipopeptide" includes pharmaceutically acceptable salts thereof.

- 112 -

Peptides, including daptomycin or daptomycin-related lipopeptides, can be formulated using any pharmaceutically acceptable carrier or excipient that is compatible with the peptide or with the lipopeptide of interest. See, e.g., Handbook of Pharmaceutical Additives: An International Guide to More than 6000 Products by Trade Name, Chemical, Function, and Manufacturer, Ashgate Publishing Co., eds., M. Ash and I. Ash, 1996; The Merck Index: An Encyclopedia of Chemicals, Drugs and Biologicals, ed. S. Budavari, annual; Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, PA; Martindale: The Complete Drug Reference, ed. K. Parfitt, 1999; and Goodman & Gilman's The Pharmaceutical Basis of Therapeutics, Pergamon Press, New York, NY, ed. L. S. Goodman et al.; the contents of which are incorporated herein by reference, for a general description of the methods for administering various antimicrobial agents for human therapy. Peptides or lipopeptides of this invention can be mixed with conventional pharmaceutical carriers and excipients and used in the form of tablets, capsules, elixirs, suspensions, syrups, wafers, creams and the like. Peptides or lipopeptides may be mixed with other therapeutic agents and antibiotics, such as discussed herein. The compositions comprising a compound of this invention will contain from about 0.1 to about 90% by weight of the active compound, and more generally from about 10 to about 30%.

[0228] The compositions of the invention can be delivered using controlled (e.g., capsules) or sustained release delivery systems (e.g., bioerodable matrices). Exemplary delayed release delivery systems for drug delivery that are suitable for administration of the

- 113 -

compositions of the invention are described in U.S. Patent Nos. 4,452,775 (issued to Kent), 5,239,660 (issued to Leonard), 3,854,480 (issued to Zaffaroni).

[0229] The compositions may contain common carriers and excipients, such as corn starch or gelatin, lactose, sucrose, microcrystalline cellulose, kaolin, mannitol, dicalcium phosphate, sodium chloride and alginic acid. The compositions may contain croscarmellose sodium, microcrystalline cellulose, corn starch, sodium starch glycolate and alginic acid.

[0230] Tablet binders that can be included are acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

[0231] Lubricants that can be used include magnesium stearate or other metallic stearates, stearic acid, silicone fluid, talc, waxes, oils and colloidal silica.

[0232] Flavoring agents such as peppermint, oil of wintergreen, cherry flavoring or the like can also be used. It may also be desirable to add a coloring agent to make the dosage form more aesthetic in appearance or to help identify the product.

[0233] For oral use, solid formulations such as tablets and capsules are particularly useful. Sustained release or enterically coated preparations may also be devised. For pediatric and geriatric applications, suspensions, syrups and chewable tablets are especially suitable. For oral administration, the pharmaceutical compositions are in the form of, for example, a tablet, capsule, suspension or liquid. The pharmaceutical composition is preferably made in the form of a dosage unit containing a therapeutically-effective amount of the active ingredient. Examples of such dosage units are

- 114 -

tablets and capsules. For therapeutic purposes, the tablets and capsules which can contain, in addition to the active ingredient, conventional carriers such as binding agents, for example, acacia gum, gelatin, polyvinylpyrrolidone, sorbitol, or tragacanth; fillers, for example, calcium phosphate, glycine, lactose, maize-starch, sorbitol, or sucrose; lubricants, for example, magnesium stearate, polyethylene glycol, silica, or talc; disintegrants, for example, potato starch, flavoring or coloring agents, or acceptable wetting agents. Oral liquid preparations generally are in the form of aqueous or oily solutions, suspensions, emulsions, syrups or elixirs may contain conventional additives such as suspending agents, emulsifying agents, non-aqueous agents, preservatives, coloring agents and flavoring agents. Oral liquid preparations may comprise lipopeptide micelles or monomeric forms of the lipopeptide. Examples of additives for liquid preparations include acacia, almond oil, ethyl alcohol, fractionated coconut oil, gelatin, glucose syrup, glycerin, hydrogenated edible fats, lecithin, methyl cellulose, methyl or propyl *para*-hydroxybenzoate, propylene glycol, sorbitol, or sorbic acid.

[0234] For intravenous (IV) use, a water soluble form of the peptide or lipopeptide can be dissolved in any of the commonly used intravenous fluids and administered by infusion. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts. Intravenous fluids include, without limitation, physiological saline or Ringer's solution. Peptides or lipopeptides also may be placed in injectors, cannulae, catheters and lines.

- 115 -

[0235] Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions. These solutions or suspensions can be prepared from sterile powders or granules having one or more of the carriers mentioned for use in the formulations for oral administration. Lipopeptide micelles may be particularly desirable for parenteral administration. The compounds can be dissolved in polyethylene glycol, propylene glycol, ethanol, corn oil, benzyl alcohol, sodium chloride, and/or various buffers. For intramuscular preparations, a sterile formulation of a lipopeptide compound or a suitable soluble salt form of the compound, for example the hydrochloride salt, can be dissolved and administered in a pharmaceutical diluent such as Water-for-Injection (WFI), physiological saline or 5% glucose.

[0236] Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

[0237] For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of creams, ointments, liquid sprays or inhalants, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate

- 116 -

surface penetration of the active ingredient. For topical preparations, a sterile formulation of daptomycin, daptomycin-related lipopeptide or suitable salt forms thereof, may be administered in a cream, ointment, spray or other topical dressing. Topical preparations may also be in the form of bandages that have been impregnated with daptomycin or a daptomycin-related lipopeptide composition.

[0238] For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

[0239] For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glycerides.

[0240] For aerosol preparations, a sterile formulation of the peptide or lipopeptide or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. A sterile formulation of a lipopeptide micelle may also be used for aerosol preparation. Aerosolized forms may be especially useful for treating respiratory infections, such as pneumonia and sinus-based infections.

[0241] Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery. In one embodiment, the unit dosage form of the compound can be a solution of the compound or a salt thereof, in a suitable diluent in sterile, hermetically sealed ampules. The concentration of the compound in the unit dosage may vary, e.g. from about 1

- 117 -

percent to about 50 percent, depending on the compound used and its solubility and the dose desired by the physician. If the compositions contain dosage units, each dosage unit preferably contains from 50-500 mg of
5 the active material. For adult human treatment, the dosage employed preferably ranges from 100 mg to 3 g, per day, depending on the route and frequency of administration.

[0242] In a further aspect, this invention provides a
10 method for treating an infection, especially those caused by gram-positive bacteria, in humans and other animals. The term "treating" is used to denote both the prevention of an infection and the control of an established infection after the host animal has become infected. An
15 established infection may be one that is acute or chronic. The method comprises administering to the human or other animal an effective dose of a compound of this invention. An effective dose of daptomycin, for example, is generally between about 0.1 and about 25 mg/kg
20 daptomycin, daptomycin-related lipopeptide or pharmaceutically acceptable salts thereof. The daptomycin or daptomycin-related lipopeptide may be monomeric or may be part of a lipopeptide micelle. A preferred dose is from about 1 to about 25 mg/kg of
25 daptomycin or daptomycin-related lipopeptide or pharmaceutically acceptable salts thereof. A more preferred dose is from about 1 to 12 mg/kg daptomycin or a pharmaceutically acceptable salt thereof. These dosages for daptomycin may be used as a starting point by
30 one of skill in the art to determine and optimize effective dosages of other linear and cyclic peptides produced by the modified NRPS complexes of the invention.

- 118 -

[0243] In one embodiment, the invention provides a method for treating an infection, especially those caused by gram-positive bacteria, in a subject with a therapeutically-effective amount of modified daptomycin or other antibacterial peptide or lipopeptide produced by a modified NRPS of the invention. The daptomycin or antibacterial peptide or lipopeptide may be monomeric or in a lipopeptide micelle. Exemplary procedures for delivering an antibacterial agent are described in U.S. Patent No. 5,041,567, issued to Rogers and in PCT patent application number EP94/02552 (publication no. WO 95/05384), the entire contents of which documents are incorporated in their entirety herein by reference. As used herein the phrase "therapeutically-effective amount" means an amount of modified daptomycin or other antibacterial peptide or lipopeptide produced by a modified NRPS according to the present invention, that prevents the onset, alleviates the symptoms, or stops the progression of a bacterial infection. The term "treating" is defined as administering, to a subject, a therapeutically-effective amount of a compound of the invention, both to prevent the occurrence of an infection and to control or eliminate an infection. The term "subject", as described herein, is defined as a mammal, a plant or a cell culture. In a preferred embodiment, a subject is a human or other animal patient in need of peptide or lipopeptide compound treatment.

[0244] The peptide or lipopeptide antibiotic compound can be administered as a single daily dose or in multiple doses per day. The treatment regime may require administration over extended periods of time, e.g., for several days or for from two to four weeks. The amount per administered dose or the total amount administered

- 119 -

will depend on such factors as the nature and severity of the infection, the age and general health of the patient, the tolerance of the patient to the antibiotic and the microorganism or microorganisms involved in the

infection. A method of administration is disclosed in United States Serial No. 09/406,568, filed September 24, 1999, herein incorporated by reference, which claims the benefit of U.S. Provisional Application Nos. 60/101,828, filed September 25, 1998, and 60/125,750, filed March 24, 1999.

[0245] The methods of the present invention comprise administering modified daptomycin or other peptide or lipopeptide antibiotics, or pharmaceutical compositions thereof to a patient in need thereof in an amount that is efficacious in reducing or eliminating the gram-positive bacterial infection. The antibiotic may be administered orally, parenterally, by inhalation, topically, rectally, nasally, buccally, vaginally, or by an implanted reservoir, external pump or catheter. The antibiotic may be prepared for ophthalmic or aerosolized uses. Modified daptomycin, a peptide or lipopeptide antibiotic produced by a modified NRPS of the invention, or a pharmaceutical compositions thereof, also may be directly injected or administered into an abscess, ventricle or joint.

Parenteral administration includes subcutaneous, intravenous, intramuscular, intra-articular, intra-synovial, cisternal, intrathecal, intrahepatic, intralesional and intracranial injection or infusion. In a preferred embodiment, daptomycin or another peptide or lipopeptide is administered intravenously, subcutaneously or orally.

[0246] The method of the instant invention may be used to treat a patient having a bacterial infection in which

- 120 -

the infection is caused or exacerbated by any type of gram-positive bacteria. In a preferred embodiment, modified daptomycin, daptomycin-related lipopeptide, or another peptide or lipopeptide antibiotic produced by a modified NRPS of the invention, or pharmaceutical compositions thereof, are administered to a patient according to the methods of this invention. In another preferred embodiment, the bacterial infection may be caused or exacerbated by bacteria including, but not limited to, methicillin-susceptible and methicillin-resistant staphylococci (including *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Staphylococcus haemolyticus*, *Staphylococcus hominis*, *Staphylococcus saprophyticus*, and coagulase-negative staphylococci), glycopeptide intermediary-susceptible *Staphylococcus aureus* (GISA), penicillin-susceptible and penicillin-resistant streptococci (including *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus avium*, *Streptococcus bovis*, *Streptococcus lactis*, *Streptococcus sanguis* and Streptococci Group C, Streptococci Group G and viridans streptococci), enterococci (including vancomycin-susceptible and vancomycin-resistant strains such as *Enterococcus faecalis* and *Enterococcus faecium*), *Clostridium difficile*, *Clostridium clostridiiforme*, *Clostridium innocuum*, *Clostridium perfringens*, *Clostridium ramosum*, *Haemophilus influenzae*, *Listeria monocytogenes*, *Corynebacterium jeikeium*, *Bifidobacterium* spp., *Eubacterium aerofaciens*, *Eubacterium lentum*, *Lactobacillus acidophilus*, *Lactobacillus casei*, *Lactobacillus plantarum*, *Lactococcus* spp., *Leuconostoc* spp., *Pediococcus*, *Peptostreptococcus anaerobius*, *Peptostreptococcus asaccarolyticus*, *Peptostreptococcus*

- 121 -

magnus, *Peptostreptococcus micros*, *Peptostreptococcus prevotii*, *Peptostreptococcus productus*, *Propionibacterium acnes*, and *Actinomyces* spp.

[0247] The antibacterial activity of daptomycin
5 against classically "resistant" strains is comparable to that against classically "susceptible" strains in *in vitro* experiments. In addition, the minimum inhibitory concentration (MIC) value for daptomycin against susceptible strains is typically 4-fold lower than that
10 of vancomycin. Thus, in a preferred embodiment, modified daptomycin, daptomycin-related lipopeptide antibiotic, a peptide or lipopeptide antibiotic produced by the modified NRPS of the invention, or pharmaceutical compositions thereof, are administered according to the
15 methods of this invention to a patient who exhibits a bacterial infection that is resistant to other antibiotics, including vancomycin. In addition, unlike glycopeptide antibiotics, daptomycin exhibits rapid, concentration-dependent bactericidal activity against
20 gram-positive organisms. Thus, in a preferred embodiment, daptomycin, a lipopeptide antibiotic, or pharmaceutical compositions thereof are administered according to the methods of this invention to a patient in need of rapidly acting antibiotic therapy.

25 [0248] The method of the instant invention may be used for a gram-positive bacterial infection of any organ or tissue in the body. These organs or tissue include, without limitation, skeletal muscle, skin, bloodstream, kidneys, heart, lung and bone. The method of the
30 invention may be used to treat, without limitation, skin and soft tissue infections, bacteremia and urinary tract infections. The method of the invention may be used to treat community acquired respiratory infections,

- 122 -

including, without limitation, otitis media, sinusitis, chronic bronchitis and pneumonia, including pneumonia caused by drug-resistant *Streptococcus pneumoniae* or *Haemophilus influenzae*. The method of the invention also

5 may be used to treat mixed infections that comprise different types of gram-positive bacteria, or which comprise both gram-positive and gram-negative bacteria, including aerobic, caprophilic or anaerobic bacteria. These types of infections include intra-abdominal

10 infections and obstetrical/gynecological infections. The methods of the invention may be used in step-down therapy for hospital infections, including, without limitation, pneumonia, intra-abdominal sepsis, skin and soft tissue infections and bone and joint infections. The method of

15 the invention also may be used to treat an infection including, without limitation, endocarditis, nephritis, septic arthritis and osteomyelitis. In a preferred embodiment, any of the above-described diseases may be treated using daptomycin, lipopeptide antibiotic, or

20 pharmaceutical compositions thereof. Further, the diseases may be treated using daptomycin or lipopeptide antibiotic in either a monomeric or micellar form.

[0249] Modified daptomycin, daptomycin-related lipopeptide, or another peptide or lipopeptide produced

25 by a modified NRPS according to the invention, may also be administered in the diet or feed of a patient or animal. If administered as part of a total dietary intake, the amount of modified daptomycin or other peptide or lipopeptide can be less than 1% by weight of

30 the diet and preferably no more than 0.5% by weight. The diet for animals can be normal foodstuffs to which modified daptomycin or the other peptide or lipopeptide can be added or it can be added to a premix.

- 123 -

[0250] The method of the instant invention may also be practiced while concurrently administering one or more antifungal agents and/or one or more antibiotics other than modified daptomycin or other peptide or lipopeptide antibiotic. Co-administration of an antifungal agent and an antibiotic other than modified daptomycin or another peptide or lipopeptide antibiotic may be useful for mixed infections such as those caused by different types of gram-positive bacteria, those caused by both gram-positive and gram-negative bacteria, or those that caused by both bacteria and fungus. Furthermore, modified daptomycin or other peptide or lipopeptide antibiotic may improve the toxicity profile of one or more co-administered antibiotics. It has been shown that administration of daptomycin and an aminoglycoside may ameliorate renal toxicity caused by the aminoglycoside. In a preferred embodiment, an antibiotic and/or antifungal agent may be administered concurrently with modified daptomycin, other peptide or lipopeptide antibiotic, or in pharmaceutical compositions comprising modified daptomycin or another peptide or lipopeptide antibiotic.

[0251] Antibacterial agents and classes thereof that may be co-administered with modified daptomycin or other peptide or lipopeptide antibiotics include, without limitation, penicillins and related drugs, carbapenems, cephalosporins and related drugs, aminoglycosides, bacitracin, gramicidin, mupirocin, chloramphenicol, thiamphenicol, fusidate sodium, lincomycin, clindamycin, macrolides, novobiocin, polymyxins, rifamycins, spectinomycin, tetracyclines, vancomycin, teicoplanin, streptogramins, anti-folate agents including sulfonamides, trimethoprim and its combinations and

- 124 -

pyrimethamine, synthetic antibacterials including nitrofurans, methenamine mandelate and methenamine hippurate, nitroimidazoles, quinolones, fluoroquinolones, isoniazid, ethambutol, pyrazinamide, para-aminosalicylic acid (PAS), cycloserine, capreomycin, ethionamide, 5 prothionamide, thiacetazone, viomycin, eveminomycin, glycopeptide, glycylcycline, ketolides, oxazolidinone; imipenen, amikacin, netilmicin, fosfomycin, gentamicin, ceftriaxone, Zircin, LY 333328, CL 331002, HMR 3647, 10 Linezolid, Synercid, Aztreonam, and Metronidazole, Epiroprim, OCA-983, GV-143253, Sanfetrinem sodium, CS-834, Biapenem, A-99058.1, A-165600, A-179796, KA 159, Dynemicin A, DX8739, DU 6681; Cefluprenam, ER 35786, Cefoselis, Sanfetrinem celexetil, HGP-31, Cefpirome, HMR- 15 3647, RU-59863, Mersacidin, KP 736, Rifalazil; Kosan, AM 1732, MEN 10700, Lenapenem, BO 2502A, NE-1530, PR 39, K130, OPC 20000, OPC 2045, Venepriam, PD 138312, PD 140248, CP 111905, Sulopenem, ritipenam acoxyl, RO-65-5788, Cyclothialidine, Sch-40832, SEP-132613, micacocidin 20 A, SB-275833, SR-15402, SUN A0026, TOC 39, carumonam, Cefozopran, Cefetamet pivoxil, and T 3811.

[0252] In a preferred embodiment, antibacterial agents that may be co-administered with modified daptomycin or peptide or lipopeptide antibiotic produced by a modified 25 NRPS according to this invention include, without limitation, imipenen, amikacin, netilmicin, fosfomycin, gentamicin, ceftriaxone, teicoplanin, Zircin, LY 333328, CL 331002, HMR 3647, Linezolid, Synercid, Aztreonam, and Metronidazole.

30 **[0253]** Antifungal agents that may be co-administered with modified daptomycin or other peptide or lipopeptide antibiotic include, without limitation, Caspofungen, Voriconazole, Sertaconazole, IB-367, FK-463, LY-303366,

- 125 -

Sch-56592, Sitafloracin, DB-289 polyenes, such as Amphotericin, Nystatin, Primaricin; azoles, such as Fluconazole, Itraconazole, and Ketoconazole; allylamines, such as Naftifine and Terbinafine; and anti-metabolites
5 such as Flucytosine. Other antifungal agents include without limitation, those disclosed in Foster et al., Drug Discovery Today 5:25-32 (2000), herein incorporated by reference. Foster et al. disclose antifungal compounds including Corynecandin, Mer-WF3010,
10 Fusacandins, Artrichitin/LL 15G256γ, Sordarins, Cispentacin, Azoxybacillin, Aureobasidin and Khafrefungin.

[0254] Modified daptomycin or other peptide or lipopeptide antibiotics, including daptomycin-related
15 lipopeptides, may be administered according to this method until the bacterial infection is eradicated or reduced. In one embodiment, modified daptomycin, or other peptide or lipopeptide produced by a modified NRPS according to the invention, is administered for a period
20 of time from 3 days to 6 months. In a preferred embodiment, modified daptomycin, or other peptide or lipopeptide, is administered for 7 to 56 days. In a more preferred embodiment, modified daptomycin, or other peptide or lipopeptide is administered for 7 to 28 days.
25 In an even more preferred embodiment, modified daptomycin or other peptide or lipopeptide antibiotic is administered for 7 to 14 days. In another embodiment, the antibiotic is administered for 3 to 7 days. Modified daptomycin, or other peptide or lipopeptide produced by a
30 modified NRPS according to the invention, according to the invention may be administered for a longer or shorter time period if it is so desired.

- 126 -

[0255] In order that this invention may be more fully understood, the following examples are set forth. These examples are for the purpose of illustration only and are not to be construed as limiting the scope of the invention in any way.

EXAMPLE 1: Initial sequencing of the *Streptomyces roseosporus* daptomycin biosynthetic gene cluster

[0256] *Streptomyces roseosporus* strain A21978.6 (American Type Culture Collection Accession No. 31568) was used for the construction of a cosmid library. Genomic DNA was digested partially with *Sau3A1* and alkaline phosphatase (Boehringer Mannheim Biochemicals). DNA of approximately 40 kb in length was isolated and ligated to *Bam*HI-digested cosmid pKC1471 and packaged with a Gigapack packaging extract (Stratagene, Inc.) as described in Hosted and Baltz, J. Bacteriol., 179, pp. 180-186 (1997). Packaged DNA was introduced into *E. coli* XL1-Blue-MFR' (Stratagene, Inc.) and individual clones containing cosmid DNA were stored as an ordered array in a 96-well dot blot apparatus. Twelve cultures from a row of microtiter wells were pooled, and screened by hybridization to a 2.1-kB *SphI* fragment of DNA from plasmid pRHB153 and to a 5.2-kB *DraI-KpnI* fragment from pRHB157, both containing NRPS sequences cloned from *S. roseosporus* (see McHenney et al., *supra*). Individual cosmids from the hybridizing pools were identified by hybridization to the same probes.

[0257] Cosmid and plasmid DNA was hydrodynamically sheared and then separated by electrophoresis on a standard 1% agarose gel. The separated DNA fragments 2500-3000 bp in length were excised from the gel and

- 127 -

purified by the GeneClean™ procedure (BIO 101, Inc.). The ends of the gel-purified DNA fragments were then filled in or made blunt using T4 DNA polymerase. The DNA fragments were ligated to unique *Bst*XI-linker adapters
5 (5'-GTCTTCACCCACGGGG-3'-SEQ ID NO: 13, and 5'GTGGTGAAGAC-3'-SEQ ID NO: 14, in 100-1000 fold molar excess). These linkers are complementary to the *Bst*XI-cut pGTC vector (Genome Therapeutics Corp., Waltham, MA), while the overhang is not self-complementary. Therefore,
10 the linkers will not concatemerize, nor will the open vector self-ligate easily. The linker-adapted inserts were separated from the unincorporated linkers by electrophoresis on a 1% agarose gel and purified using GeneClean™. The purified linker-adapted inserts were
15 ligated to *Bst*XI-cut pGTC vector to construct "shotgun" subclone libraries.

[0258] The pGTC library was then transformed into DH5α competent cells (Gibco/BRL, DH5α transformation protocol). Transformation was assessed by plating onto
20 antibiotic plates containing ampicillin and IPTG/Xgal (IPTG = isopropyl-b-D-thiogalactopyranoside; Xgal=5-bromo-4-chloro-3-indoyl-b-D-thiogalactopyranoside.) The plates were incubated overnight at 37°C. Transformants were plate purified and the purified clones containing
25 the following plasmids were picked for further analysis.

[0259] Plasmids pRHB160, containing an insert of approximately 50 kb of *S. roseosporus* DNA, pRHB613, containing an insert of approximately 15 kb, pRHB614, containing an insert of approximately 13 kb, and pRHB159,
30 containing an insert of approximately 51 kb, were chosen for DNA sequencing. (See McHenney *et al.*, *supra*).

[0260] Individual cultures of strains transformed with the above plasmids were grown overnight at 37°C. DNA was

- 128 -

purified using a silica bead DNA preparation method (Engelstein et al., *Microb. Comp. Genomics* 3(4):237-241, 1998). In this manner, 25 mg of DNA were obtained per clone. These purified DNA samples were then sequenced
5 using primarily ABI dye-terminator chemistry. All subsequent steps were based on sequencing by ABI377 or Amersham automated DNA sequencing methods according to the manufacturer's instructions. The ABI dye terminator sequence reads were run on either ABI377 or Amersham
10 MegaBace™ capillary machines. The data were transferred to UNIX machines following lane tracking of the gels. Base calls and quality scores were determined using the program PHRED (Ewing et al., *Genome Res.* 8:175-185, 1998). Reads were assembled using PHRAP (P. Green,
15 Abstracts of DOE Human Genome Program Contractor-Grantee Workshop V, Jan. 1996, p.157) with default program parameters and quality scores. The initial assembly was done at 6x coverage.

20 EXAMPLE 2: Isolation and analysis of additional DNA
molecules of the *Streptomyces roseosporus*
biosynthetic gene cluster

[0261] Mycelium for preparation of megabase DNA was obtained from overnight cultures of *Streptomyces roseosporus* (NRRL11379) (ATCC No.31568) shaken in F10A
25 broth (2% agar, 25% soluble starch, 0.2% dextrose, 0.5% yeast extract, 0.5% peptone and 0.3% calcium carbonate) at 30°C. Washed cells were embedded in Seakem™ GTG agarose (FMC Bioproducts, 1% final concentration), incubated in lysozyme (2mg/mL TE) at 37°C for 3h, then
30 lysed in 0.1x NLS + 0.2mg/mL proteinase K at 50°C overnight to release DNA into the gel matrix. Agarose containing DNA was washed with 1 mM EDTA (pH 8) before

- 129 -

treatment with *Bam*HI at 37°C. Partially digested DNA was then subjected to a two-step size selection process in 0.6% agarose gels (in 0.5X TBE) by pulsed-field electrophoresis using a CHEF Mapper DRIII (Biorad) set at 6V/cm, 120° angle, 12°C. The first selection consisted of a 14 h run with a 22-44 sec linearly ramped switch time. Gel containing DNA co-migrating with 100-200 kb lambda concatamer size markers was excised and cast in a second gel for an 18 h run with a 3-5 sec linear ramp. DNA estimated at 75-145 kb relative to size markers was electroeluted (MiniProtean II Cell model, Biorad) in TAE.

[0262] The single-copy BAC library cloning vector pStreptoBAC V is derived from pBACe3.6 (Frengen et al., A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites, *Genomics*, 58: 250-253 (1999)). The pBACe3.6 was modified to contain two markers, Amp^R for selection in *E. coli* and Apra^R for selection in *Streptomyces*, as well as oriT and attP sequences from the phage jC31 for conjugation and site specific integration in *Streptomyces*. See Figure 6. To prepare the pStreptoBAC V vector for ligation with the *S. roseosporus* DNA, the vector was first digested with *Bam*HI and the reaction was inactivated by heat (65°C for 1h). DNA was then dephosphorylated with Shrimp Alkaline Phosphatase for 30min. The two bands (13 kb and 3kb corresponding to the pUC fragment) were separated on 0.6% agarose gel and the 13 kb band was purified using GeneClean spin columns.

[0263] 200 ng of the *S. roseosporus* DNA was ligated to 75 ng of *Bam*HI cut and phosphatased pStreptoBAC V vector DNA using 9 U of T4 DNA ligase (Promega) in a 150 µl reaction. After 16 h at 16°C, the ligations were heated at 65° C for 30 min, dialyzed against 10% polyethylene

- 130 -

glycol 8000, and transformed into 10 µl of DH10B electrocompetent cells (Gibco/BRL) using a cell porator with voltage booster (Gibco/BRL) at 300 V and 4 kΩ. Cells were plated on media (LB agar) containing 100mg/mL apramycin and 5% sucrose. Analysis of sample clones showed a range of inserts from 39 kb to 105 kb. The mean insert size was 71.4 kb, with a standard deviation of 14.7 kb. Approximately 2,000 clones were archived at -80°C in 96-well microtiter plates.

[0264] This BAC library was screened using the polymerase chain reaction (PCR) using primer pairs P61/P62, P72/P73 and P74/P75, shown below. Nucleotide positions refer to the numbering of SEQ ID NO: 1, and "C" indicates that the primer sequence corresponds to the complementary strand of SEQ ID NO: 1:

<u>Primer</u>	<u>Sequence</u>	<u>SEQ ID NO:</u>	<u>Nucleotide Position</u>
P61	GCTCGTCCCCCTCCCCGCACT	137	41305-41325
P62	CGAACAGGTGGGCTTTGAGTGG	138	41993-42014 (C)
P72	CTTCGTGAACACCCTCGTCC	139	82103-82122
P73	GTTTCGTCGAGGTCCAGTACG	140	83009-83028 (C)
P74	GCACCAGCGTGTGCGGATCG	141	92-111
P75	CACGTACGTGACGATCCTCG	142	800-819 (C)

[0265] PCR was performed under the following conditions: 94° C, 45 sec., 54° C, 30 sec., 72° C, 1 min. for 32 cycles. Taq polymerase, as well as the accessory reagents, were supplied by Gibco BRL (Bethesda); all reactions included 5% DMSO.

[0266] Clone B12:03A05 was initially detected with primer pair P61/P62 (see above), and subsequently confirmed as a positive hit with the other two primer pairs. DNA of clone B12:03A05 was obtained by standard

- 131 -

alkaline lysis procedures and used for DNA sequencing (see below).

[0267] A number of other clones that encompass parts of the daptomycin gene cluster (*dpt*-related clones) were isolated from the BAC library. These clones include B12:01G05 (insert size 82 kb), B12:06A12 (insert size 85 kb), B12:12F06 (insert size 65 kb), B12:18H04 (insert size 46 kb) and B12:20C09 (insert size 65 kb). See Figure 7, which shows a *Hin*DIII digest of these BAC clones. Other BACs that were isolated in the daptomycin gene cluster region include B12:09D02, B12:17F08, B12:05D08, B12:15H07, B12:21F10 and B12:16D12. These BACS cover 180 to 200 kb. Figure 8 shows the approximate location of the BAC clones relative to the daptomycin gene cluster.

[0268] Extension of the daptomycin biosynthetic gene cluster sequence determined in Example 2 was accomplished by sequencing 1 µg aliquots of BAC DNA from clone B12:03A05 using the ABI Prism Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin Elmer), the manufacturer's recommended reaction mix and conditions, and the following primers (C indicates that the primer sequence corresponds to the complementary strand of SEQ ID NO: 1):

<u>Primer</u>	<u>Sequence</u>	<u>SEQ ID NO:</u>	<u>Nucleotide Position</u>
P76	CGTACTGGACCTCGACGACC	143	83009-83028
P78	CGACCAGCGTGTGTACGTCC	144	83609-83628
P92	AGTCCTCAGCCATCTCCTCG	145	84584-84603 (C)
P84	GAGACCGTCGGCGTGGACG	146	84222-84240
P95	AGGGCCACACCGTCGAACCTCC	147	84709-84729
P86	ATCGTCGCCGACTACCTCGC	148	84795-84814
P96	GGCAGCTACCTCGTACTGG	149	85297-85315
P97	TGTACGACAGCGGCGTCGAAC	150	85959-85979
P101	CGATTCTCGGCATGTTCCGC	151	86636-86655
P105	TCGTCTCCTACATGACCTCG	152	87194-87213

- 132 -

P107	TTCACGGAAACCGAACGTCG	153	87864-87883
P111	GGTTCAGGCCGCGAGCCAACG	154	88468-88487
P117	CGCTGACCTTGGTCAGAAGCC	155	89176-89196

[0269] Electropherograms were inspected and corrected as appropriate, and the sequences were aligned using the AssemblyLign Module of MacVector™. The aligned sequence (contig) was saved as a MacVector™ file for analysis and annotation. Identification of potential ORFs and potential stops/starts was performed using the open reading frames option in MacVector™.

[0270] Analysis of the 90kb sequence showed a total of 38 open reading frames in the daptomycin biosynthetic gene cluster region. See Figure 2. The ORFs range in size from 228 basepairs (bp) to 22 kb. The three largest ORFs are NRPS genes, as discussed below. One of the NRPS genes were predicted to have thioesterase activity based on the presence of conserved motifs, GX SXG (see Example 3). Another predicted open reading frames also encodes a protein with thioesterase activity (see Example 3). A number of potential ABC transporters were also identified.

[0271] The sequence of the daptomycin biosynthetic gene cluster is shown in SEQ ID NO: 1. See also Figure 2. The genes encoding the daptomycin non-ribosomal peptide synthetase (NRPS) are designated *dptA*, *dptBC* and *dptD*. We designate as a promoter region all sequences upstream from the start of an ORF of interest that are not part of an upstream ORF. Because *dptA*, *dptBC* and *dptD* have overlapping start and stop codons and apparently are translationally coupled (e.g., the TGA stop codon of *dptBC* overlaps with the ATG start codon of *dptD*, which is presumably associated with its own ribosome binding site), we thus indicate the promoter of

- 133 -

the whole cluster (comprising *dptE*, *dptF*, *dptA*, *dptBC* and *dptD*) as the daptomycin NPRS promoter.

[0272] The DNA sequence of the ORF of the daptomycin NRPS *dptA* gene (nucleotides 38555-56047 of SEQ ID NO: 1) is shown in SEQ ID NO: 10. The ORF is 17493 nucleotides in length. The amino acid sequence of the encoded DptA protein is shown in SEQ ID NO: 9. The protein is 5830 amino acid residues in length.

[0273] The DNA sequence of the ORF of the daptomycin NRPS *dptBC* gene (nucleotides 56044-78060 of SEQ ID NO: 1) is shown in SEQ ID NO: 12. The ORF is 22017 nucleotides in length. The amino acid sequence of the encoded DptBC protein is shown in SEQ ID NO: 11. The protein is 7338 amino acid residues in length.

[0274] The DNA sequence of the ORF of the daptomycin NRPS *dptD* gene (nucleotides 78057-85196 of SEQ ID NO: 1) is shown in SEQ ID NO: 3. The ORF is 7140 nucleotides. The *dptD* gene ORF encodes a type I thioesterase (TEI) domain at the C-terminus. The amino acid sequence of the predicted DptD protein is shown in SEQ ID NO: 7 (see Figure 3). The protein is 2379 amino acids in length.

[0275] The *dptE* and *dptF* genes are located between *dptA* and the daptomycin NPRS promoter.

[0276] The DNA sequence of the *dptH* thioesterase-encoding gene is shown in SEQ ID NO: 4 (nucleotides 85498-86350 of SEQ ID NO: 1); the promoter region of *dptH* is shown in SEQ ID NO: 5 (nucleotides 85498-85534 of SEQ ID NO: 1); and the open reading frame of *dptH* is shown in SEQ ID NO: 6 (nucleotides 85535-86350 of SEQ ID NO: 1). The amino acid sequence of the predicted DptH protein is shown in SEQ ID NO: 8 (see Figure 4).

- 134 -

[0277] The promoter region of the daptomycin NRPS (nucleotides 36018-36407 of SEQ ID NO: 1) is shown in SEQ ID NO: 2.

[0278] The sequence for the DNA downstream of the 90 kb contig was generated by Genome Therapeutics Corps. from plasmid pV107 by transposon primed sequencing using a system such as GPS-1 Genome priming system (New England Biolabs). Plasmid pV107 comprises an approximately 28 kb EcoRI fragment subcloned using standard techniques from B12:03A05 genome into the vector pNEB193, cut with EcoRI (New England Biolabs). The fragment is called the GTC2 fragment. An adequate number of transposon tagged library clones were sequenced to generate a 6-fold redundant contig, and additional local sequencing off PCR products was used to polish the sequence where needed. The overlap between the 5' end of the contig and the existing 90 kb was removed from the contig, so that the beginning of the pV107 derived sequences (referred to as GTC2) starts with 1. The sequence of the GTC2 fragment is provided in SEQ ID NO: 106.

EXAMPLE 3: Identification of the *dptD* and *dptH* genes as thioesterases

[0279] Amino acid motifs typical of non-ribosomal peptide synthetases and thioesterases were identified by inspection of the *dptD* and *dptH* genes and predicted translation products thereof. The amino acid sequence motif GX SXG, wherein X is any one of the twenty L-amino acids that are inserted translationally into ribosomally produced proteins, is indicative of thioesterases (See Mootz et al., *J. Bacteriol.* 179:6843-6850, 1997, incorporated herein by reference in its entirety). SEQ

- 135 -

ID NOs 7-8 were inspected for the GX SXG thioesterase motif. In SEQ ID NO:7, the amino acid sequence match to the thioesterase motif GWSFG was found at coordinates 2200-2204, encoded by nucleotides 84654-84668 of SEQ ID NO:1. In SEQ ID NO:8, the amino acid sequence match to the thioesterase motif GTSLG was found at coordinates 97-101, encoded by nucleotides 85823-85838 of SEQ ID NO:1.

[0280] The DptD protein of SEQ ID NO:7 was aligned to the CDA III protein of *Streptomyces coelicolor*. The alignment was performed using the Clustal W (v1.4) program in slow pairwise alignment mode. An open gap penalty of 10.0, an extend gap penalty of 0.1, and a blosum similarity matrix to the CDA III protein was used. The CDA III protein is a non-ribosomal peptide synthetase with a carboxy-terminal thioesterase domain (see GENBANK accession number AL035707, version AL035707.1 GI:4490978, hereby incorporated by reference in its entirety). The CDA III amino acid sequence used for the alignment was generated using the MacVector program by creating a contig from two GENBANK cosmid sequences, AL035707 and AL035640, and then translating the open reading frame in the contig annotated in GENBANK. The sequence comparison (Figure 3) revealed an alignment score of 7705 and 1223 conserved identities, indicating significant similarity between the two compared sequences. The GX SXG thioesterase motifs of the DptD protein and the CDA III protein were aligned in this analysis.

[0281] The GX SXG thioesterase motif of the DptH protein of SEQ ID NO: 8 was aligned to the GX SXG thioesterase motif of the CDA III protein of *Streptomyces coelicolor* (CAA71338 protein, see above). The alignment was performed the Clustal W (v1.4) program in slow pairwise alignment mode. An open gap penalty of 10.0, an

- 136 -

extend gap penalty of 0.1, and a blosum similarity matrix to the *Streptomyces* thioesterase protein of GENPEPT record CAA71338 (version CAA71338.1 GI:2647975, hereby incorporated by reference in its entirety) was used. The alignment (Figure 4) revealed an alignment score of 955 and 145 conserved identities indicating significant similarity between the two compared sequences.

[0282] These analyses show that *dptD* and *dptH* encode thioesterase proteins, specifically, the proteins of SEQ ID NOS: 7-8.

EXAMPLE 4: Identification of a Daptomycin NRPS

Identification of dptD as a daptomycin NRPS subunit

[0283] The predicted translation products of the *dptD* DNA sequences described above (Examples 2 and 3) were inspected visually for the occurrence of various protein motifs described in the NRPS literature. A *dptD* condensation ("M") motif, indicative of a condensation domain, was identified at nucleotides 78486-78509 of SEQ ID NO: 1 (all of the nucleotide positions discussed in Examples 4-6 refer to SEQ ID NO: 1). See, e.g., Kleinkauf et al., *Eur. J. Biochem.*, 236, pp. 335-351 (1996) for the various motifs in the NRPS; and Pospiech et al., *Microbiology*, 142, pp. 741-746 (1996). An ATP-binding ("C") motif was identified at nucleotides 79896-79928, an ATP-binding ("E") motif was identified at nucleotides 80451-80486, an ATPase ("F") motif was identified at nucleotides 80556-80579, and an ATP-binding ("G") motif was identified at nucleotides 80652-80675. These motifs collectively are indicative of an adenylation domain. A thiolation ("J") motif, indicative of a thiolation (PCP) domain, was identified at

- 137 -

nucleotides 81048-81062. The above motifs, and the domains that they signify, belong to module 1 of *dptD*; in terms of daptomycin synthetase, this is module 12.

[0284] Another *dptD* condensation ("M") motif, indicative of a condensation domain, was identified at nucleotides 81621-81644. Another ATP-binding ("C") motif was identified at nucleotides 83114-83147, an ATP-binding ("E") motif was identified at nucleotides 83667-83702, an ATPase ("F") motif was identified at nucleotides 83772-83795, and an ATP-binding ("G") motif was identified at nucleotides 83868-83891. The above motifs collectively are indicative of another adenylation domain. Also a thiolation ("J") motif, an indicator of a thiolation (PCP) domain, was identified at nucleotides 84255-84269. The above motifs, and the domains that they signify, belong to module 2 of *dptD*; in terms of daptomycin synthetase, this is module 13.

[0285] The DptD amino acid sequences corresponding to the above-described predicted motifs and domains were identified (all of the amino acid positions for DptD refer to the amino acid positions in SEQ ID NO: 7). The motifs, and the domains that they signify, belonging to module 1 of DptD (corresponding to module 12 of daptomycin synthetase) are as follows: A DptD condensation ("M") motif was identified at coordinates 144-151; an ATP-binding ("C") motif was identified at coordinates 614-624; an ATP-binding ("E") motif was identified at coordinates 799-810; an ATPase ("F") motif was identified at coordinates 834-841; an ATP-binding ("G") motif was identified at coordinates 866-873; and a thiolation ("J") motif was identified at coordinates 998-1002.

- 138 -

[0286] The DptD motifs, and the domains that they signify, belonging to module 2 of DptD (corresponding to module 13 of daptomycin synthetase) are as follows: A DptD condensation ("M") motif was identified at coordinates 1189-1196; an ATP-binding ("C") motif was identified at coordinates 1687-1697; an ATP-binding ("E") motif was identified at coordinates 1871-1882; an ATPase ("F") motif was identified at coordinates 1906-1913; an ATP-binding ("G") motif was identified at coordinates 1938-1945; and a thiolation ("J") motif was identified at coordinates 2067-2071. The ATP-binding motifs are representative of adenylation domains.

Identification of dptA and dptBC as daptomycin NRPS subunits

[0287] Certain M, C, E, F, G and J motifs were identified in a similar fashion in *dptA* and *dptBC*. The sequence and type of each motif, the genes and modules in which each motif is found, as well as the amino acid and nucleotide coordinates of each motif, are shown below in Table 1:

Table 1

Gene	Module	Motif Type	Sequence	Amino Acid Coordinates	Nucleotide Coordinates
<i>dptA</i>	1	M	HHIALDGY	138-145	38966-38989
<i>dptA</i>	1	C	QTSGSTGRPKG	603-613	40361-40393
<i>dptA</i>	1	E	GELYLAGEGLAR	784-795	40904-40939
<i>dptA</i>	1	F	RMVRTGDL	819-826	41009-41032
<i>dptA</i>	1	G	RIELGEVQ	851-858	41105-41128
<i>dptA</i>	1	J	LGGHS	981-985	41495-41509
<i>dptA</i>	2	M	HHTAGDGA	1167-1174	42053-42076
<i>dptA</i>	2	C	YTSGSTGRPKG	1657-1667	43523-43555
<i>dptA</i>	2	E	GELHVAGEGLAR	1843-1854	44081-44116
<i>dptA</i>	2	F	RMVRTGDL	1878-1885	44186-44209
<i>dptA</i>	2	G	RIELGEVE	1910-1917	44282-44305
<i>dptA</i>	2	J	LGGDS	2041-2045	44675-44689
<i>dptA</i>	3	M	HHVILDGW	2751-2758	46805-46828
<i>dptA</i>	3	C	YTSGSTGLPKG	3238-3248	48266-48298
<i>dptA</i>	3	E	GELYVAGDGLAR	3420-3431	48812-48847
<i>dptA</i>	3	F	RMVRTGDL	3455-3462	48917-48940
<i>dptA</i>	3	G	RIELGEVE	3487-3494	49013-49036

- 139 -

Gene	Module	Motif Type	Sequence	Amino Acid Coordinates	Nucleotide Coordinates
dptA	3	J	LGGHS	3616-3620	49400-49414
dptA	4	M	HHIAGDGW	3806-3813	49970-49993
dptA	4	C	YTSGSTGRPKG	4292-4302	51428-51460
dptA	4	E	GEMYVAGAGLAR	4490-4501	52022-52057
dptA	4	F	RLYRTGDL	4525-4532	52127-52150
dptA	4	G	RIELGEIE	4557-4564	52223-52246
dptA	4	J	LGGHS	4688-4692	52616-52630
dptA	5	M	HHIAGDGW	4873-4880	53171-53194
dptA	5	C	HTSGSTGRPKG	5363-5373	54641-54673
dptA	5	E	GEIHIAGSGLAR	5553-5564	55211-55246
dptA	5	F	RMYRTGDL	5587-5594	55313-55336
dptA	5	G	RIELGDVE	5619-5626	55409-55432
dptA	5	J	LGGDS	5749-5753	55799-55813
dptBC	1	M	HHVILDGW	142-149	56467-56490
dptBC	1	C	HTSGSTGRPKG	611-621	57874-57906
dptBC	1	E	GELYLAGTQLAR	803-814	58450-58485
dptBC	1	F	RMYRTGDL	838-845	58555-58578
dptBC	1	G	RIEPAEIE	870-877	58651-58674
dptBC	1	J	AGGHS	998-1002	59035-59049
dptBC	2	M	HHIAGDGW	1184-1191	59593-59616
dptBC	2	C	YTSGSTGRPKG	1691-1701	61114-61146
dptBC	2	E	GELYVAGVGLAR	1873-1884	61660-61695
dptBC	2	F	RMYRTGDL	1908-1915	61765-61788
dptBC	2	G	RVELGEVE	1940-1947	61861-61884
dptBC	2	J	LGGHS	2069-2073	62248-62262
dptBC	3	M	HHVAFDAM	2259-2266	62818-62841
dptBC	3	C	YTSGSTGRPKG	2740-2750	64261-64293
dptBC	3	E	GELYVAGVGLAR	2923-2934	64810-64845
dptBC	3	F	RMYRTGDL	2958-2965	64915-64938
dptBC	3	G	RVELGEVE	2990-2997	65011-65034
dptBC	3	J	LGGDS	3118-3122	65395-65409
dptBC	4	M	HHVVLDGW	3805-3812	67456-67479
dptBC	4	C	YTSGSTGRPKG	4282-4292	68887-68919
dptBC	4	E	GELYVAGVGLAR	4464-4475	69433-69468
dptBC	4	F	RMYRTGDL	4499-4506	69538-69561
dptBC	4	G	RVELGEVE	4531-4538	69634-69657
dptBC	4	J	LGGHS	4662-4666	70027-70041
dptBC	5	M	HHIAGDGW	4852-4859	70597-70620
dptBC	5	C	YTSGSTGQPKG	5340-5350	72061-72093
dptBC	5	E	GELYIAGDGLAR	5526-5537	72619-72654
dptBC	5	F	RMYRTGDL	5561-5568	72724-72747
dptBC	5	G	RVELGEVE	5593-5600	72820-72843
dptBC	5	J	LGGHS	5722-5726	73206-73221
dptBC	6	M	HHIAGDGW	5913-5920	73780-73803
dptBC	6	C	YTSGSTGRPKG	6394-6404	75223-75255
dptBC	6	E	GELYLAGAGLAR	6584-6595	75793-75828
dptBC	6	F	RMYRTGDL	6619-6626	75898-75921
dptBC	6	G	RVELGEVE	6651-6658	75994-76017
dptBC	6	J	LGGDS	6781-6785	76384-76398

- 140 -

[0288] The amino acid coordinates in Table 1 refer to the amino acid sequence of each protein (DptA: SEQ ID NO: 9; DptBC: SEQ ID NO: 11). The nucleotide position refers to the nucleotide position in SEQ ID NO: 1.

5 EXAMPLE 5: Amino acid pocket code annotation

[0289] The amino acid pocket code refers to a set of amino acid residues in the adenylation (A) domain that are believed to be involved in recognition and or binding of the cognate amino acid. The amino acid pocket code
10 for the thirteen daptomycin synthetase modules are shown below (Table 2).

[0290] The amino acid pocket code for the daptomycin synthetase modules was identified by Blast analysis or visual inspection of alignments created using MacVector
15 7.0 of the putative Dpt translation product aligned with NRPS A domains (amino acid binding pockets) as described in Stachelhaus et al. (1999), The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases, *Chem. Biol.*, 6:493-505. See also Challis et al., (2000), Predictive, structure-based model of amino
20 acid recognition by nonribosomal peptide synthetase adenylation domains, *Chem. Biol.* 7:211-224.

Table 2.				
Protein	Module (Amino acid)	Pocket Code	Amino Acid Coordinates	Nucleotide Position
DptA	1 (Trp)	DVSSIGAV	649, 650, 653, 690, 711, 713, 734, 742	40499-40780
DptA	2 (Asn)	DLTKLGDV	1702, 1703, 1706, 1741, 1764, 1766, 1790, 1798	43658-43949

DptA	3 (Asp)	DLTKLGAV	3284, 3285, 3288, 3318, 3341, 3343, 3367, 3375	48404-48679
DptA	4 (Thr)	DFWSVGMV	4338, 4339, 4342, 4381, 4410, 4412, 4438, 4446	51566-51892
DptA	5 (Gly)	DILQLGVI	5409, 5410, 5413, 5452, 5479, 5481, 5503, 5511	54779-55087
DptBC	1 (Orn)	DTWDMGYV	662, 663, 665, 704, 730, 732, 755, 763	58027-58332
DptBC	2 (Asp)	DLTKLGAV	1737, 1738, 1741, 1771, 1794, 1796, 1820, 1828	61252-61527
DptBC	3 (Ala)	DVVSAAFV	2786, 2787, 2790, 2824, 2847, 2849, 2873, 2881	64399-64686
DptBC	4 (Asp)	DLTKLGAV	4328, 4329, 4332, 4362, 4385, 4387, 4411, 4419	69025-69300
DptBC	5 (Gly)	DILQVGMI	5386, 5387, 5390, 5429, 5452, 5454, 5476, 5484	72200-72495
DptBC	6 (Ser)	DVWHISLV	6440, 6441, 6444, 6483, 6508, 6510, 6533, 6541	75361-75666
DptD	1 (3-MG)	DLGKTGVI	659, 660, 663, 697, 720, 722, 746, 754	80031-80318
DptD	2 (Kyn)	DAWTTTGV	1733, 1734, 1737, 1775, 1796, 1798, 1820, 1828	83253-83540

[0291] The amino acid coordinates in Table 2 refer to the amino acid sequence of each protein (DptA: SEQ ID NO: 9; DptBC: SEQ ID NO: 11; DptD: SEQ ID NO: 7). The nucleotide position refers to the nucleotide position in SEQ ID NO: 1.

[0292] Similarities between essentially the entire adenylation domains for aspartate and asparagine in the daptomycin gene cluster and for the adenylation domains for aspartate, asparagine and threonine in the CDA III

- 142 -

NRPS of *Streptomyces coelicolor* are shown in Figure 10. Amino acids were aligned and the dendrogram was constructed using the MacVector. The nomenclature is as follows: the name of the gene--the module number in the gene--the amino acid activated (one letter code). The alignment shows that the adenylation domains for aspartate and asparagine in the daptomycin gene cluster are more similar to each other than they are to a domain from an unrelated amino acid such as threonine. Further, the alignment shows that the adenylation domains for aspartate and asparagine in the daptomycin gene cluster are more similar to each other than they are similar to the modules for aspartate and asparagine in CDA.

EXAMPLE 6: Identification of Epimerase Domains in Daptomycin NRPS

[0293] The amino acid sequences of DptA, DptBC and DptD were inspected for sequences that are characteristic of epimerase domains. Epimerase domains are responsible for converting an L-amino acid to a D-amino acid and are typically encoded by approximately 1.4-1.6 kb of DNA.

[0294] It was expected that there would be a total of two epimerase domains in the daptomycin gene cluster, because it was known that daptomycin contained two D-amino acids, D-Ala and D-Ser. One epimerase domain was identified in each of module 8 (D-Ala) and module 11 (D-Ser). Module 8 and 11 are approximately 1.4 kb larger than modules that did not contain an epimerase domain (approximately 4.6 kb each for modules 8 and 11 compared to 3.2 kb each for modules not containing an epimerase domain). Further, modules 8 and 11 contain motifs that are indicative of an epimerase domain, including the

motifs K, L, M, N, O, P and Q (see Kleinkauf and Von Dohren, 236:355-351 (1996)). See Table 3.

[0295] Surprisingly, an epimerase domain was also identified in module 2. Module 2 is 1.6 kb larger than expected. Further, module 2 contains a number of motifs that are characteristic of an epimerase domain, including motifs K, L, M, N, O, P and Q. See Table 3. This unexpected finding suggests that the asparagine in daptomycin is in the D configuration.

10 Table 3

Gene	Module	Motif Type	Sequence	Amino Acid Coordinates	Nucleotide Coordinates
<i>dptA</i>	2	K	RWPVVEWL	2100-2107	44852-44875
<i>dptA</i>	2	L	VRERHDAW	2146-2153	44990-45013
<i>dptA</i>	2	M	HHLVVDGVS WRIVLG	2237-2251	45263-45307
<i>dptA</i>	2	N	VVDVEGHGRN	2374-2383	45674-45703
<i>dptA</i>	2	O	TVGWFTSIYPVRL	2395-2407	45737-45775
<i>dptA</i>	2	P	PDQGLGY	2439-2445	45869-45889
<i>dptA</i>	2	Q	FGFNLYG	2467-2473	45953-45973
<i>dptBC</i>	3	K	RWPVVEWL	3183-3190	65590-65613
<i>dptBC</i>	3	L	VRDRHEAW	3229-3236	65728-65751
<i>dptBC</i>	3	M	HHLVVDGVS WRVVLG	33315-33329	65986-66030
<i>dptBC</i>	3	N	VVDVEGHGRN	3452-3461	66397-66426
<i>dptBC</i>	3	O	TVGWFTSVYPVRV	3473-3485	66460-66498
<i>dptBC</i>	3	P	PDQGLGY	3517-3523	66592-66612
<i>dptBC</i>	3	Q	FGFNLYG	3545-3551	66676-66696
<i>dptBC</i>	6	K	RWPVVEWL	6846-6853	76579-76602
<i>dptBC</i>	6	L	VRDRHEAW	6892-6899	76717-76740
<i>dptBC</i>	6	M	HHLVVDGVS WRVVLG	6978-6992	76975-77019
<i>dptBC</i>	6	N	VVDVEGHGRN	7115-7124	77383-77415
<i>dptBC</i>	6	O	TVGWFTSVYPVRV	7136-7148	77449-77487
<i>dptBC</i>	6	P	PDQGLGY	7180-7186	77581-77601
<i>dptBC</i>	6	Q	FGFNLYG	7208-7214	77665-77685

[0296] The amino acid coordinates in Table 3 refer to the amino acid sequence of each protein (DptA: SEQ ID NO: 9; DptBC: SEQ ID NO: 11; DptD: SEQ ID NO: 7). The nucleotide position refers to the nucleotide position in
5 SEQ ID NO: 1.

[0297] To confirm that the asparagine in daptomycin was in the D configuration, high pressure liquid chromatography (HPLC) was performed. A hexa-peptide containing the amino acids ornithine, glycine, threonine,
10 aspartic acid, asparagine, and deacylated tryptophan (Trp-Asn-Asp-Orn-Gly-Thr) was isolated from daptomycin by degradation. The peptide above was analyzed by HPLC under conditions that would separate the peptide containing either the D-Asn or L-Asn. The HPLC showed only a single
15 large peak for the isolated peptide above. See Figure 11, left panel. The peptide isolated from daptomycin was mixed with a peptide of the same sequence that had been synthesized in the laboratory and which contained D-Asn. The peptide mixture was analyzed by HPLC under the same
20 conditions as before and shown to contain only a single peak. See Figure 11, middle panel. In addition, the peptide isolated from daptomycin was mixed with a synthetic peptide of the same sequence that contained L-Asn. HPLC analysis displayed two peaks. See Figure 11,
25 right panel. These experiments confirm that naturally-occurring daptomycin contains D-Asn, not L-Asn.

[0298] From the experiments presented in Examples 2-7, the organization of the daptomycin NRPS was determined. Figure 12 shows the organization of *dptA*, *dptBC*, and
30 *dptD*. *dptA* contains five modules (modules 1-5), *dptB* contains 6 modules (modules 6-11), and *dptD* contains two modules (modules 12-13) and a thioesterase domain. Table 4 summarizes the correspondence between the 13 modules,

- 145 -

their domains, the *dpt* genes, and their cognate amino acids. "C" represents a catalytic domain, "A" represents an adenylation domain, "T" represents a thiolation domain, "E" represents an epimerase domain, and "Te" represents a thioesterase domain.

Table 4.

Module	Cognate Amino Acid	Domains	Gene
01	L-Trp	CAT	<i>dptA</i>
02	D-Asn	CATE	<i>dptA</i>
03	L-Asp	CAT	<i>dptA</i>
04	L-Thr	CAT	<i>dptA</i>
05	Gly	CAT	<i>dptA</i>
06	L-Orn	CAT	<i>dptBC</i>
07	L-Asp	CAT	<i>dptBC</i>
08	D-Ala	CATE	<i>dptBC</i>
09	L-Asp	CAT	<i>dptBC</i>
10	Gly	CAT	<i>dptBC</i>
11	D-Ser	CATE	<i>dptBC</i>
12	L-MG	CAT	<i>dptD</i>
13	Kyn	CAT-Te	<i>dptD</i>

EXAMPLE 7: Transformation of *Streptomyces lividans* With The Daptomycin Gene Cluster From *Streptomyces roseosporus*

[0299] *E. coli* cells containing the BAC DNA from clone B12:03A05 (see Example 2) were grown in 5 mL of Luria Broth (LB; Difco) with agitation (250 rpm) overnight at 37°C. The BAC DNA was isolated by a standard alkaline lysis procedure (see Sambrook et al., *supra*, "Small scale preparation of plasmid DNA").

- 146 -

[0300] *S. lividans* TK64 spores were used to inoculate 25 mL of YEME + sucrose media and the culture was incubated for 40 hours at 30°C. The cultures were then harvested and the mycelium was pelleted away from the supernatant and washed several times with P-buffer (Practical *Streptomyces* Genetics; Tobias Kieser, Mervyn J. Bibb, Mark J. Buttner, Keith F. Chater and David Hopwood (John Innes Foundation, Norwich, 2000) ("Practical *Streptomyces* Genetics"). Fresh protoplasts were prepared according to the method described in Practical *Streptomyces* Genetics, supra (p. 56) and aliquoted into 0.5 mL portions (approximately 10^8 - 10^9 protoplasts) and pelleted by centrifugation at 3000 rpm for 7 minutes. Most of the supernatant was removed, leaving the pellet and approximately 50µL of the supernatant. The pellet was resuspended in the remaining supernatant, to which was added 5µL of BAC DNA from clone B12:03A05 (50 ng/µL in TE). This suspension was gently mixed before and after adding 350µL of a 25% PEG-1000 in P-buffer solution (Practical *Streptomyces* Genetics, supra).

[0301] The protoplast suspension mixture was spread, in equal amounts, onto three dried R5T plates (dried to lose approximately 15% of their original weight; see Practical *Streptomyces* Genetics, supra). Inoculated plates were incubated overnight at 30°C. After 16-18 hours of growth, the plates were overlaid with 3 mL of an apramycin solution (1 mg/mL) in 20% glycerol to provide a final concentration of approximately 100µg/mL on each plate, and the plates incubated at 30°C. After three days, the plates were determined, by examination, to contain colonies which were growing in the presence of the apramycin selection. Two colonies were picked and

- 147 -

streaked onto two F10A agar plates (2.5% agar, 0.3% calcium carbonate, 0.5% distillers solubles, 2.5% soluble starch, 0.5% yeast extract, 0.2% dextrose and 0.5% bactopectone; suspended in 1 L deionized and autoclaved water) containing 100 μ L/mL of apramycin and allowed to incubate at 30°C until the colonies sporulated. Spores were harvested according to the methods described in Practical *Streptomyces* Genetics, supra and stored as 20% glycerol suspensions at -20°C.

- 10 **[0302]** The spores derived from the transformation of *S. lividans* with BAC DNA containing the daptomycin gene cluster (from clone B12:03A05, CBUK136742) were grown in an appropriate medium and analyzed by high pressure liquid chromatography (HPLC) and LC-MS to determine if they produced a wild-type lipopeptide profile (see Example 9).

EXAMPLE 8: Fermentation of *Streptomyces lividans* TK64 clone containing the daptomycin gene cluster

- 20 **[0303]** Spores of the *Streptomyces lividans* TK64 clone containing the daptomycin gene cluster (from clone B12:03A05) were harvested by suspending a 10 day old slant culture of medium A (2% irradiated oats (Quaker), 0.7% tryptone (Difco), 0.2% soya peptone (Sigma), 0.5% sodium chloride (BDH), 0.1% trace salts solution, 1.8% agar no. 2 (Lab M), 0.01 % apramycin (Sigma)) in 5 mL 10% aqueous glycerol (BDH)). 1 mL of this suspension, in a 1.5 mL cryovial, comprises the starting material, which was retrieved from storage at -135 °C. A pre-culture was produced by aseptically placing 0.3 mL of the starting

- 148 -

material onto a slope of medium A1 and incubating for 9 days at 28 °C.

[0304] A seed culture was generated by aseptically treating the pre-culture with 4 mL of a 0.1% Tween 80 (Sigma) solution and gently macerating the slope surface to generate a suspension of vegetative mycelium and spores. A two mL aliquot of this suspension was transferred into a 250 mL baffled flask containing 40 mL of nutrient solution S (1% D-glucose (BDH), 1.5% glycerol (BDH), 1.5% soya peptone (Sigma), 0.3% sodium chloride (BDH), 0.5% malt extract (Oxoid), 0.5% yeast extract (Lab M), 0.1% Junlon PW100 (Honeywell and Stein Ltd), 0.1% Tween 80 (Sigma), 4.6% MOPS (Sigma) adjusted to pH 7.0 and autoclaved)) and shaken at 240 rpm for 44 hours at 30 °C.

[0305] Production cultures were generated by aseptically transferring 5% of the seed culture to baffled 250 mL flasks containing 50 mL medium P (1% glucose (BDH), 2% soluble starch (Sigma), 0.5% yeast extract (Difco), 0.5% casein (Sigma), 4.6% MOPS (Sigma) adjusted to pH 7 and autoclaved)) and shaken at 240 rpm for up to 7 days at 30 °C.

25 EXAMPLE 9: Purification and Analysis of the A21978C Lipopeptides from Fermentations of the Streptomyces lividans TK64 Clone Containing the Daptomycin Gene Cluster

[0306] Production cultures described in Example 8 were sampled for analysis by aseptically removing 2 mL of the whole culture and centrifuging for 10 minutes prior to analysis. Volumes up to 50 microlitres of the supernatant were analyzed to monitor for production of the native lipopeptides (A21978C) as produced by

- 149 -

Streptomyces roseosporus. This analysis was performed at ambient temperature using a Waters Alliance 2690 HPLC system and a 996 PDA detector with a 4.6 x 50 mm Symmetry C8 3.5µm column and a Phenomenex Security Guard C8 cartridge. The gradient initially holds at 90% water and 10% acetonitrile for 2.5 minutes, followed by a linear gradient over 6 minutes to 100% acetonitrile. The flow rate is 1.5 mL per minute and the gradient is buffered with 0.01% trifluoroacetic acid. By day 2 of the fermentation, production of three of the native lipopeptides, C1, C2 and C3, with UV/visible spectra identical to that of daptomycin, was evident, as shown by HPLC peaks with retention times of 5.62, 5.77 and 5.90 minutes (λ_{max} 223.8, 261.5 and 364.5 nm) under the analytical conditions stated, as shown in Figure 5A. The lipopeptides then remained evident in the fermentation at each sample point during the 7-day period. Total yields of lipopeptides C1, C2 and C3 ranged from 10-20 mg per liter of fermentation material.

[0307] Liquid chromatography-mass spectrometry (LC-MS) analysis was performed on a Finnigan SSQ710c LC-MS system using electrospray ionization in positive ion mode, with a scan range of 200-2000 daltons and 2 second scans. Chromatographic separation was achieved on a Waters Symmetry C8 column (2.1x 50mm, 3.5µm particle size) eluted with a linear water-acetonitrile gradient containing 0.01% formic acid, increasing from 10% to 100% acetonitrile over a period of six minutes after a initial delay of 0.5 minutes, then remaining at 100% acetonitrile for a further 3.5 minutes before re-equilibration. The flow rate was 0.35 mL/minute and the method was run at ambient temperature.

- 150 -

[0308] The identification of the three native lipopeptides was confirmed, as indicated by molecular ions ($[M+H]^+$) at m/z of 1634.7, 1648.7 and 1662.7, which is in agreement with the masses reported for the major
5 A21978C lipopeptide metabolites C1, C2 and C3, respectively, produced by *Streptomyces roseosporus* (Debono et al., *J. Antibiotics*, 40, pp. 761-777 (1987)).

[0309] Similar experiments were performed using the BAC clones 01G06, B12:06A12, B12:12F06 and B12:18H04.
10 None of the *S. lividans* cells containing any one of these BAC clones was able to produce daptomycin.

EXAMPLE 10: Fed-batch fermentation of *Streptomyces lividans* TK64 Clone Containing the Daptomycin Gene Cluster for the production of Daptomycin

15 [0310] Cells of the *Streptomyces lividans* TK64 clone containing the daptomycin gene cluster (from clone B12:03A05) were regenerated by suspending a 10 day old slope culture of medium A (see Practical *Streptomyces* Genetics; 2% irradiate oats (Quaker), 0.7% tryptone
20 (Difco), 0.2% soya peptone (Sigma), 0.5% sodium chloride (BDH), 0.1% trace salts solution, 1.8% agar no. 2 (Lab M), 0.01% apramycin (Sigma) in 5 mL 10% aqueous glycerol (BDH)). A 1.5 mL cryovial containing 1 mL of starting material was retrieved from storage at -135 °C and thawed
25 rapidly. A pre-culture was produced by aseptically placing 0.3 mL of the starting material onto a slope of medium A and incubating for 9 days at 28 °C. Material for inoculation of the seed culture was generated by aseptically treating the preculture with 4 mL of a 0.1%
30 Tween 80 (Sigma) solution and gently macerating the slope surface to generate a suspension of vegetative mycelium and spores.

- 151 -

[0311] A seed culture was produced by aseptically placing 1 mL of the inoculation material into a 2L baffled Erlenmeyer flask containing 250 mL of nutrient solution S (see Practical *Streptomyces* Genetics, supra) shaken at 240 rpm for 2 days at 30 °C.

[0312] A production culture was generated by aseptically transferring the seed culture to a 20L fermenter containing 14 liters of nutrient solution P (see Practical *Streptomyces* Genetics, supra). The production fermenter was stirred at 350 rpm, aerated at 0.5 vvm, and temperature controlled at 30 °C. After 20 hours incubation a 50% (w/v) glucose solution was fed to the culture at 5 g/hr throughout the fermentation.

[0313] After 40 hours incubation, a 50:50 (w/w) blend of decanoic acid:methyl oleate (Sigma and Acros Organics, respectively) was fed to the fermenter at 0.5 g/hr for the remainder of fermentation. The culture was harvested after 112 hours, and the biomass removed from the culture supernatant by batch processing through a bowl centrifuge.

[0314] The biomass was discarded and the clarified fermentation broth was retained for extraction. The broth (approximately 10L) was loaded onto a 60mm (diameter) by 300mm (length) column of HP20 resin, which had been pre-equilibrated with water, at a rate of 100 mL/min. The column was washed with 2L of water and then with 1.5L of 80% methanol (in water) at a similar flow rate. Finally, the bound material was eluted with 2L methanol and then taken to an aqueous concentrate under vacuum. The concentrate was diluted to 1L with purified water and partitioned with ethyl acetate (700 mL) three times. The ethyl acetate fraction was analyzed and discarded, and the aqueous layer was lyophilized to a powder.

- 152 -

[0315] Daptomycin was isolated by high performance liquid chromatography (HPLC) using a radially compressed cartridge column consisting of two 40x100mm Waters Nova-Pak C18 6 μ m units and a 40x10mm Guard-Pak with identical packing. Lyophilized material (150 to 200mg) was dissolved in water and chromatographed on the columns using a gradient in which the initial conditions were 90% water and 10% acetonitrile, followed by a linear increase over 10 minutes to 20% water and 80% acetonitrile, and then immediately ramping up to 100% acetonitrile over a further minute. UV absorption at 223nm was monitored for elution of daptomycin. The daptomycin peak eluted at about 9 minutes and was collected and combined over many repeated runs. The sample was then evaporated under vacuum and then dried *in vacuo* to yield 30 mg of purified compound. Only a proportion of the total material was processed.

[0316] The purified compound was first analyzed by reversed phase HPLC at ambient temperature on a 4.6 x 50 mm Waters Symmetry C8 3.5 μ m particle size column with a Phenomenex Security Guard C8 cartridge using a Waters Alliance 2690 HPLC system and a 996 PDA detector. The column was eluted with a water-acetonitrile gradient, initially holding at 90% water for 2.5 minutes and then rising linearly over 6 minutes to 100% acetonitrile, at a flow rate of 1.5 mL/minute. The gradient was buffered with 0.01% trifluoroacetic acid. This chromatographic analysis confirmed that the retention time (5.52 mins) and the UV absorption spectrum ($\lambda_{\max}$ 223.8, 261.5, 366.9nm) of the purified compound matched those of daptomycin. LC-MS(ESI) confirmed the molecular ion MH^+ as 1620.6 (Figure 5B) and the ^1H NMR (D6-DMSO) gave a good

- 153 -

visual match with that recorded for daptomycin (Figure 5C).

[0317] The identification of the material as daptomycin was further confirmed by ¹³CNMR experiments.

5 [0318] Feed-batch fermentation may also be accomplished at a larger scale, for example at 60,000 liters.

EXAMPLE 11: The use of daptomycin genes for yield enhancement

10 Chapter 1 Duplication of a positive regulatory gene
[0319] A neutral genomic site in the chromosome of *Streptomyces roseosporus* is identified by transposon mutagenesis with TN5097, or a related transposon, followed by fermentation analysis. The neutral site is
15 excised from the chromosome using a restriction endonuclease that cuts outside of the neutral site and transposon, and cloned in *Escherichia coli*, selecting for the expression of the antibiotic resistance marker in the transposon (hygromycin resistance in the case of TN5097).
20 An example of this approach was used to identify a neutral site in *Streptomyces fradiae*, the tylosin producer. See Baltz et al., *Antonie van Leeuwenhoek*, 71, pp. 179-187 (1997), incorporated herein by reference in its entirety. An example of identifying a neutral site
25 in *S. roseosporus* is described in McHenney et al., *J. Bacteriol.*, 180, pp. 143-151 (1998), incorporated herein by reference in its entirety.
[0320] The regulatory gene from the daptomycin gene cluster (SEQ ID NO: 109) is cloned into a plasmid within
30 the neutral site. A suitable plasmid would be one containing an antibiotic resistance gene for the

- 154 -

selection of primary recombinants containing single crossovers, a counter-selectable marker such as the wild type *rpsL* gene, a ribosomal protein gene that confers sensitivity to streptomycin (Hosted and Baltz, *J. Bacteriol.*, 179, pp. 180-186 (1997)) for selection of recombinants containing double crossovers that insert the cloned regulatory gene, and upstream and downstream sequences, into the chromosomal neutral site, and eliminate the plasmid sequences, and a thermal sensitive replicon that would facilitate the curing of the plasmid. The double crossover is done in a host strain that is normally resistant to streptomycin because it contains a mutation in the *rpsL* gene. Since the wild type (streptomycin-sensitive) allele of *rpsL* is dominant over streptomycin resistance, recombinants expressing streptomycin resistance must have eliminated the *rpsL* gene on the plasmid by a double crossover in the two arms of the neutral site, thus inserting the cloned daptomycin regulatory gene into the chromosome. Recombinants are fermented to verify that they produce an increased yield compared to the parental strain lacking the cloned daptomycin regulatory gene.

Duplication of ABC transporter genes

[0321] One or more of the ABC transporter genes from the daptomycin gene cluster, including upstream and downstream sequences, are cloned into the neutral site vector described above and inserted by double crossover into the *S. roseosporus* chromosome as described in Example 11A. Recombinants are fermented to verify that they produce increased levels of daptomycin compared to the parental strain lacking the cloned ABC transporter genes.

- 155 -

Duplication of novA,B,C homologs

[0322] The segment of DNA containing the *novA*, *B*, *C* homology from the daptomycin gene cluster, including the upstream and downstream sequences, is cloned into the neutral site vector and inserted by double crossover into the *S. roseosporus* chromosome as described in Example 11A. Recombinants are fermented to verify that they produce increased levels of daptomycin compared to the parental strain lacking the cloned *novA*, *B*, *C* genes.

10 *Duplication of daptomycin biosynthetic genes*

[0323] The daptomycin biosynthetic genes, *dptA*, *dptBC*, *dptD*, *dptE*, *dptF*, *dptG* and *dptH*, including the fatty acyl-CoA ligase, the three subunits of the NRPS, the integral thioesterase of *dptD* and the free thioesterase of *dptH*, are cloned into a BAC vector that contains the phiC31 attachment and integration functions (*att/int*) and *oriT* from plasmid RK2 (Baltz, *Trends in Microbiol.*, 6: 76-83 (1998), incorporated herein by reference in its entirety) for conjugation from *E. coli* to *S. roseosporus*.

20 The BAC containing the daptomycin genes is introduced into *S. roseosporus* by conjugation from *E. coli* S17.1, or a strain containing a self-replicating plasmid RK2 (*Id.*). Alternatively, the BAC vector inserts into the chromosome by homologous recombination into the daptomycin gene

25 cluster. Recombinants are fermented to verify that they produce increased levels of daptomycin compared to the parental strain lacking the cloned daptomycin genes.

Duplication of daptomycin thioesterase genes

[0324] The daptomycin gene cluster (SEQ ID NO: 1) contains at least two genes (*dptD* and *dptH*) having open reading frames (SEQ ID NO: 3 and SEQ ID NO: 6, respectively) or domains thereof that encode amino acid

- 156 -

sequences which include conserved sequence motifs characteristic of proteins having thioesterase activity. See SEQ ID NO:7 and SEQ ID NO:8 for DptD and DptH amino acid sequences, respectively. Either one (or both) of these thioesterase genes or the thioesterase domains thereof can be duplicated by following the procedure of Example 11A, above.

[0325] A segment of DNA containing the *dptD* ORF sequences (e.g., SEQ ID NO: 1; SEQ ID NO: 3) optionally linked in an operative fashion to an expression control sequence (such as the natural one in SEQ ID NO: 1 or 2) and optionally including the upstream and downstream sequences, is cloned into a neutral site vector and inserted by double crossover into the *S. roseosporus* chromosome as described in Example 11A. Recombinants are fermented to verify that they produce increased levels of daptomycin compared to the parental strain lacking the cloned *dptD* gene.

[0326] Similarly, a segment of DNA containing the *dptH* ORF sequences (e.g., SEQ ID NO: 4, SEQ ID NO: 6) optionally linked in an operative fashion to an expression control sequence (such as the natural one in SEQ ID NOS: 1, 4 or 5) and optionally including the upstream and downstream sequences, is cloned into a neutral site vector and inserted by double crossover into the *S. roseosporus* chromosome as described in Example 11A. Recombinants are fermented to verify that they produce increased levels of daptomycin compared to the parental strain lacking the cloned *dptH* gene.

[0327] Other suitable hosts (i.e., those having NRPS or PKS multienzyme complexes) may be transformed with segments of DNA encoding proteins from the daptomycin gene cluster having thioesterase activity for improved

- 157 -

peptide production. Alternatively, polypeptides encoded by such segments of DNA may be introduced into *S. roseosporus* or said other suitable hosts by protein transfer techniques well-known to those of skill in the art.

Duplication of daptomycin resistance genes

[0328] The daptomycin resistance gene(s) are identified by cloning and expression in an appropriate streptomycete host that is naturally susceptible to daptomycin. The cloned daptomycin resistance gene(s) are inserted into the neutral site vector within the neutral site, and inserted into the *S. roseosporus* chromosome by double crossover as described in Example 11A. Recombinants are fermented to verify that they produce increased levels of daptomycin compared to the parental strain lacking the cloned daptomycin resistance genes.

Duplication of daptomycin biosynthetic genes and accessory genes

[0329] The BAC clone B12:03A05 was introduced into wild-type *Streptomyces roseosporus* A21978.6 (ATCC Deposit No. 31568, CBUK 136737) and *Streptomyces roseosporus* A21978.65 (NRRL Deposit No. 15998, CBUK 136879) by conjugation as described in Practical *Streptomyces* Genetics, supra, to create exconjugants, CBUK 136927 and CBUK 138016, respectively. The parent and the exconjugant strains were fermented as described in Example 8 in triplicate in medium P to which 0.1% each of filter sterilized glutamic acid, Na salt (BDH) ornithine (Sigma) and aspartic acid (Sigma) were added.

[0330] Cultures were sampled and analyzed by HPLC during a 10 day experiment. The HPLC was performed as described in Example 9 except that only 5 μ L of the

- 158 -

supernatant was injected. By 40-56 hours after commencing fermentation, production of three of the *S. roseosporus* native lipopeptides (A21978C1, A21978C2 and A21978C3) was clearly evident. These lipopeptides were present in the fermentation at every subsequent sample point, with diversification to other A21978C factors evident at the end of the time course. Averaged over the three fermentation replicates, the maximum yields of A21978C lipopeptides produced by the exconjugates CBUK 136927 (284 mg/L) and CBUK138016 (1488 mg/L) were approximately twice that produced by the parent strains CBUK 136737 (143 mg/L) and CBUK 136879 (726 mg/L), respectively.

EXAMPLE 12: The Use of Daptomycin Biosynthetic Genes To Produce Novel Products

A. Modification of the peptide structure by site-directed mutagenesis of an amino acid specificity code: conversion of position 2 D-Asn to D-Asp.

[0331] The amino acid specificity codes for the thirteen amino acids in daptomycin are shown in Table 1 (see Example 6, above). See also Stachelhaus et al., *Chem. Biol.*, 6, pp. 493-505 (1999), incorporated herein by reference in its entirety, for a discussion of identifying and altering adenylation domain amino acid specificity codes in NRPSs. The code for all three L-aspartic acid residues in positions 3, 7, and 9 of daptomycin are identical: DLTKLGAV (where the letters indicate standard amino acid abbreviations). The code for D-Asn in position 2 is DLTKLGDV, and it differs by a single amino acid (a D instead of A in position 7). The D-Asn specificity code is changed to that specifying D-Asp by

- 159 -

making a site specific change in the adenylation domain of module 2 in PS I.

[0332] The mutant version of module 2 is inserted into the *S. roseosporus* chromosome by gene replacement (see
5 Example 11). A counter selectable marker (e.g., the wild type *rpsL* gene) is inserted into the adenylation domain of module 2 by gene replacement. The mutant module 2 adenylation domain containing the coding sequence for D-Asp, and containing flanking DNA (about 1 to 5 kb on
10 each side of the specificity code) on an appropriate thermal sensitive plasmid is introduced into the *S. roseosporus* strain disrupted for daptomycin biosynthesis. Recombinants containing single crossovers are selected at the non-permissive temperature by selection for an
15 antibiotic resistance marker on the plasmid (e.g., hygromycin, apramycin or thiostrepton resistance). If the host strain is streptomycin resistant by a mutation in the chromosomal *rpsL* gene, then the second crossover completing the gene replacement can be selected for
20 streptomycin resistance. The recombinant is screened for antibiotic production. The novel derivative of daptomycin is separated and analyzed to confirm the structure according to methods described, e.g., in United States Patents RE 32,333, RE 32,455, 4,874,843,
25 4,482,487, 4,537,717, and 5,912,226.

B. Molecular exchange of an amino acid coding module for one of different amino acid specificity.

[0333] Daptomycin has four acidic amino acids: three L-asp residues at positions 3, 7, and 9, and a
30 3-methyl-Glu (3-MG) at position 12 (see Table 1, Example 6). Novel derivatives of daptomycin are generated by exchanging the adenylation domain that specifies 3-MG for one that specifies L-asp. The adenylation domain of the

- 160 -

3-MG module is inserted into segments of the L-asp module flanking the L-asp adenylation domain which has been removed by molecular genetic procedures. The hybrid 3-MG module containing the flanking DNA from an L-asp module is inserted into an appropriately constructed gene replacement vector, and the hybrid module is exchanged for an L-asp module by homologous double crossover as in Example 11A. This same procedure is repeated for the other two L-asp modules. The recombinants produce three novel derivatives of daptomycin containing 3-MG substituted for L-asp in positions 3, 7, or 9, and maintain the overall four negative charges in the molecule.

C. Exchange of a non-ribosomal peptide synthetase (NRPS) subunit for one that catalyzes the incorporation of different amino acid(s).

[0334] The gene that encodes the third subunit of the daptomycin NRPS (see Table 1, Example 6) contains two modules that encode the specificity for incorporation of amino acids 12 (3-MG) and 13 (L-kyn). The gene that encodes the third subunit for the biosynthesis of the cyclic lipopeptide CDA (Kempter et al., *Angew. Chem. Int. Ed. Engl.*, 36, pp. 498-501 (1997); Chong et al., *Microbiology*, 144, pp. 193-199 (1998); each of which is incorporated by reference herein in its entirety) also encodes the last two amino acids, in this case amino acids 10 (3-MG) and 11(L-trp). A derivative of daptomycin containing L-trp instead of L-kyn in position 13 is generated by disrupting gene *dptD*, and by replacing it with the gene that encodes PSIII for CDA. Expression of the PSIII gene from a strong promoter (e.g., the *ermEp** promoter; Baltz, *Trends Microbiol.*, 6, pp. 76-83 (1998), incorporated herein by reference in its

- 161 -

entirety), and inserted into a neutral site in the *S. roseosporus* genome as described in Example 11A, allows CDA PSIII to complement the *dptD* mutation and results in the production of the altered daptomycin with L-trp replacing L-kyn. The recombinant is fermented and the product(s) of the recombinant are analyzed by LC-MS as described in Example 9.

[0335] Similar manipulations can be performed for *trans*-complementation for other subunits, i.e. to generate a disruption or deletion in a subunit of the daptomycin biosynthetic gene cluster, and then complement in *trans* by one or more natural or modified subunits from an NRPS (the latter can include *trans*-complementation by modified versions of daptomycin biosynthetic gene cluster subunits, such as can be generated using methods described throughout example 12, especially examples 12A, 12B or 12H, 12J). *Trans*-complementation between the NRPS subunits then leads to production of a novel nonribosomal peptide which can be analysed for as described in previous examples.

[0336] To perform a *trans*-complementation experiment using portions of the daptomycin biosynthetic gene cluster and the calcium dependent antibiotic (CDA) biosynthetic gene cluster, the set of daptomycin biosynthetic genes, or the set of daptomycin biosynthetic genes and accessory genes, such as those contained on the BAC clone B12:03A05 are introduced by transformation or conjugation into other natural or engineered strains or species of actinomycetes. The recipients may be known producers of secondary metabolites or uncharacterized strains, or may be generated by recombinant techniques to carry biosynthetic pathways other than that for biosynthesis of daptomycin. The transformants or

- 162 -

ex-conjugants are fermented in a variety of media and whole broth or extracts thereof are screened for either novel daptomycin-like compounds or biological activity against daptomycin-resistant tester organisms.

5 **[0337]** In some instances, complementation is facilitated by inactivation of some of the genes in the daptomycin biosynthetic pathway. Sequences encoding a subunit of the NRPS in the BAC B12:03A05 can be deleted or replaced by a marker gene to form a modified B12:03A05
10 before introduction into a heterologous host that already expresses one or more native or introduced NRPS gene clusters. Enzyme subunits encoded by the modified B12:03A05 clone and by endogenous host NRPS genes may then associate in the cytoplasm to form a heteromeric
15 multienzyme complex that leads to production of a novel peptide. In other cases, a deletion may be created after B12:03A05 or a portion comprising the daptomycin NRPS is introduced into a heterologous host that already expresses one or more native or introduced NRPS gene
20 clusters. For example, a *dptD*-disrupted or -deleted version of B12:03A05 can be created in a strain of *S. lividans* into which B12:03A05 has been introduced. *S. lividans* carries a native copy of the gene cluster for CDA. The resulting strain is fermented and analyzed to
25 show that complementation between the CDA PS III and the modified B12:03A05 leads to production of a derivative of daptomycin containing L-trp instead of L-kyn in position 13. In one embodiment of this example, it was shown that a novel lipopeptide could be produced by *trans*-
30 complementation using the *S. lividans* TK64/B12:03A05 strain described in Example 7.

[0338] To produce the novel lipopeptide, homologous recombination across flanking DNA sequences was used to

- 163 -

exchange the bulk of the coding region of *dptD* in *S. lividans* TK64/B12:03A05 for a heterologous marker gene. To perform the homologous recombination, two fragments comprising the regions directly upstream ("5' fragment") and downstream ("3' fragment") of *dptD* were amplified from chromosomal *S. roseosporus* DNA using the following primer sets with 5'-terminal extensions in which unique restriction sites have been introduced (underlined):

5' fragment (1122 bp):
10 5' GCG AAG CTT CTG GTG GCG CAT CAC CTG G 3' (SEQ ID NO: 156)
5' GCT CTA GAT GGA AGT ATG TCC TCC ATC GC 3' (SEQ ID NO: 157)

3' fragment (1535 bp):
15 5' CGG ATC CCG CCG GCA CCT GAC CC 3' (SEQ ID NO: 158)
5' CCG AAT TCC GCC TCC GAG TAC ATC GAG G 3' (SEQ ID NO: 159)

20 **[0339]** The amplified fragments were cloned in succession into the corresponding unique sites in the multiple cloning site of pNEB193 (New England Biolabs). The resulting construct, pSD002, was confirmed by
25 restriction digest analysis for orientation, and by sequencing for the absence of errors in the portions generated by PCR. A *SpeI* fragment containing the marker gene, *ermE* (erythromycin resistance gene; see Hopwood, *supra*) was inserted into pSD002 at an *XbaI* site and
30 verified by restriction digest analysis. The resulting plasmid, pSD005, thus includes a cassette composed of *ermE* flanked by DNA stretches homologous to DNA sequences upstream and downstream of *dptD*. Once inserted into the daptomycin biosynthetic gene cluster pathway by
35 homologous recombination, this cassette would essentially replace all of *dptD*, except for the first 31 bp and the last 12 bp, with *ermE*. The region comprising the

- 164 -

replacement cassette was then subcloned into a vector (a cloning site-modified version of pRHB538; Hosted et al., *J. Bacteriol.* 179: 180-186, 1997) carrying a temperature-sensitive replication origin and *rpsL* (a gene conferring sensitivity to streptomycin that can be used in a TK64 background) to create pSD030, the final plasmid in the series for introduction into *S. lividans*.

[0340] The plasmid, pSD030, was introduced into *S. lividans* by protoplast transformation essentially as described above in Example 7. The transformation mix of protoplasts and cells was gently spread over R2Ye plates and incubated at 30°C for approximately 16 hours. Each plate was then flooded with 1 mL of water containing 1.25 mg of erythromycin, resulting in a final concentration of 50 µg/ml once the liquid was absorbed into the media. Erythromycin-resistant colonies arising on the transformation plate after 7 days were inoculated into 25 mL of TSB (Hopwood, *supra*) plus erythromycin and incubated at 30°C for 48 hours. The mycelium was harvested, and 1/10th of the mycelial mass was macerated and transferred to a new aliquot of 25 mL TSB plus erythromycin. The resultant solution was then incubated at 40°C to select against the temperature-sensitive replicon of pSD030. After 48 hours, the mycelium was harvested by centrifugation, macerated and resuspended in a final volume of 2 mL TSB. 100 µL of this suspension was spread on SPMR plates (Babcock et al., *J. Bacteriol.* 170: 2802-2808, 1988) containing 50 µg/mL erythromycin and 30 µg/mL of streptomycin. Colonies that survived were screened and shown to have the correct genotype by PCR to identify strains such as CBUK 137860, in which *ermE* had successfully replaced *dptD*.

- 165 -

[0341] Starting material of CBUK 137860 was prepared essentially as described in Example 10 and used to produce a seed culture. The production culture was also generated essentially as described in Example 10, but the
5 aeration was at 0.7 vvm. The pH of the fermenter was computer controlled at 6.50 with a 14% (v/v) ammonium hydroxide solution. A 50% (w/v) glucose solution was fed to the culture at 0.36g/L/hr throughout the fermentation.

[0342] The biomass from the 20 L fermentation was
10 discarded and the clarified liquor was applied to an open glass column, packed with Mitsubishi HP20 resin (60 x 300 mm) and conditioned with methanol and water. Prior to elution, the column was washed with 2 L of water followed by 2 L of methanol/water (1:4). The column was then
15 eluted with 2 L of methanol/water (4:1) followed by 1 L methanol, and collected as two separate fractions.

[0343] Liquid chromatography-mass spectroscopy (LC-MS) electrospray ionization (ESI) analysis indicated that both fractions contained the A21978C/CDA hybrid
20 molecules, and the less complex methanol/water (4:1) fraction was processed further. This was evaporated under vacuum to an aqueous residue and then made up to 500 mL with water. It was then back extracted with 3 x 500 mL of ethyl acetate in a 2 L separating funnel, to
25 give an aqueous and organic fraction. LC-MS (ESI) indicated that the hybrid molecules were absent from the organic phase and it was discarded. The aqueous fraction was lyophilized overnight.

[0344] The hybrid molecules were purified by
30 preparative high performance liquid chromatography (HPLC) using a Waters Prep LC system and a Waters 40 x 200mm Nova-Pak C18 60Å 6µm radially-compressed double cartridge with 40 x 10mm guard. The freeze-dried material was

- 166 -

dissolved in water and purified using a gradient method. This method held at 90% water and 10% acetonitrile for 2 minutes and was followed by a linear gradient over 13 minutes to 25% water and 75% acetonitrile. The flow was 55 mL/min and the whole gradient was buffered with 0.04% trifluoroacetic acid. Fractions were collected and analyzed by LC-MS on a Finnigan SSQ710c LC-MS system using electrospray ionisation (ESI) in positive ion mode, with a scan range of 200-2000 daltons and 2 second scans. Chromatographic separation for this LC-MS analysis was achieved on a Waters Symmetry C8 column (4.6x 50mm, 3.5µm particle size) eluted with a linear water-acetonitrile gradient containing 0.01% formic acid, increasing from 10% to 100% acetonitrile over a period of six minutes after a initial delay of 0.5 minutes, then remaining at 100% acetonitrile for a further 3.5 minutes before re-equilibration. The flow rate was 1.5 mL/minute and the method was run at ambient temperature.

[0345] The analysis identified the A21978C/CDA hybrids as the expected analogues of the native A21978C lipopeptides A21978C1, with a branched native C₁₁ acyl chain, and A21978C2, with a branched C₁₂ acyl chain, in which the native kynurenine residue is replaced with a tryptophan residue. Both fractions required further purification prior to NMR studies. The C₁₁ hybrid (A) was further purified using an isocratic method with 60% water and 40% acetonitrile buffered with 0.04% trifluoroacetic acid. 1.8mg of material was isolated. The C₁₂ hybrid (B) final purification step used an isocratic method with 58% water and 42% acetonitrile buffered with 0.04% trifluoroacetic acid. Approximately 1.5 mg of material was isolated. A ¹H NMR spectrum is shown in Figure 13. The UV maxima and ESI-MS molecular ion information

- 167 -

(doubly-charged ions observed in negative ion mode) for A and B are presented below:

	A	B
ESI -MS (m/z)	814 (M-2H) ²⁻	821 (M-2H) ²⁻
UV-vis $\lambda_{\max}$ /nm	221, 280,	221, 280

[0346] Similar experiments where modified versions of the daptomycin NRPS (e.g. *dptA* deletion, *dptA* plus *dptD* deletion, etc.) are introduced into other secondary metabolite producing strains such as *S. lividans*, *S. fradiae*, *S. viridochromogenes* and others may also generate derivative compounds based on the daptomycin backbone. Given that NRPS subunits can be expressed separately and exchanged, one may *trans*-complement or exchange all subunits, alone or in combination, with one or more natural or modified NRPS subunits (including modified versions of daptomycin synthetase as described above) in *S. roseosporus* or other expressing hosts. As an example, the following modified *S. roseosporus* strain can be created: *dptA* and *dptD* deleted from the native locus, *dptBC* expression intact. Then, this strain can be complemented by an ectopically integrated *dptA* that is modified by site-directed mutagenesis to incorporate asp instead of asn at position 2 (Example 12A) and an ectopically integrated *dptD* modified so that its kyn-accepting module is exchanged for the trp-accepting module of CDA PSIII or by an ectopically integrated CDAIII. Another way to create the same strain is to bring a *dptA*, *dptD* deleted BAC B12:03A05 into a *S. lividans* TK64 derivative that already carries a modified *dptA* subunit that incorporates aspartate instead of asparagine. Such strains may be fermented to recover a

- 168 -

daptomycin derivative with asp in position 2, and trp in position 13.

D. Insertion of one or more modules to cause the expansion of the Daptomycin ring or lengthening of the tail.

5 **[0347]** A simple NRPS elongation module may be defined as comprising domains "C-A-T" (condensation-, adenylation- and thiolation-domains). To link modules, and to identify a permissive site within the daptomycin NRPS in which to insert additional internal modules, the domain and inter-domain regions are examined for sequences indicative of flexible "linker" sequences. See, e.g., Mootz et al., *Proc. Natl. Acad. Sci. U.S.A.*, 97, pp. 5848-5853 (2000), which is incorporated herein by reference in its entirety. Sequences encoding an additional module are inserted in the linker sequence between an upstream T-domain and a downstream C-domain using well-known genetic recombination techniques, e.g., see Example 11A, above.

20 **[0348]** Isolation of the module DNA is obtained from the chromosomal DNA extracted from the producer organism. Various isolation techniques can be used such as, cutting the chromosomal DNA with restriction enzymes and isolating a fragment coding for the module(s) of interest after it is identified by means of Southern blot or isolation of the module(s) of interest by e.g. genetic amplification (PCR) using suitable primers. Sequencing and characterization of the amplified fragments as well as cloning can be performed according to conventional techniques. New modules can be inserted between the modules specifying L-Thr and Gly in *dptA*; between the modules specifying L-Orn and L-Asp or L-Asp and D-Ala in *dptBC*; between L-Asp and Gly or Gly and D-Ser in *dptBC*;

25

30

- 169 -

and between modules specifying 3-MG and L-Kyn in *dptD* to expand the ring of daptomycin. New modules can be inserted in the *dptA* gene between the modules specifying L-Trp and D-Asn, D-Asn and L-Asp, or L-Asp and L-Tyr to
5 lengthen the tail of daptomycin. The module(s) insertions can be carried out using the methods for double crossovers described in Example 11A.

E. Insertion of an additional carboxyl terminus module adjacent to and upstream from the thioesterase module.

10 **[0349]** Carboxy-terminal thioesterase domains ("Te-domains") of a variety of NRPSs and PKSs can cleave (i.e., catalyze chain termination) non-natural peptide and polyketide substrates. See Mootz et al., *supra*; see also de Ferra et al., *J. Biol. Chem.*, 272, 25304-25309
15 (1997); each of which is hereby incorporated by reference in its entirety. Te-domains can act as hydrolases, releasing a linear product, or as cyclases, releasing cyclic peptides. Evidence suggests that a Te-domain which functions as a cyclase in its natural configuration
20 within a NRPS or PKS may, nonetheless, function as a hydrolase when engineered into new modular configurations. (An isolated C-terminal Te-domain has been shown to catalyze cyclization on various substrates as long as key "recognition amino acids" are at the C-
25 and N-termini of the substrate; see Trauger et al., *Nature*, 407, pp. 215-218, 2000.

[0350] It has also been shown that some C-terminal Te-domains function best, when moved, by retaining their association with a portion of the protein domain
30 occurring directly upstream in the natural NRPS or PKS modular configuration. See Guenzi et al., *J. Biol. Chem.*, 273, pp. 14403-14410 (1998), incorporated herein by reference in its entirety. It is possible that

- 170 -

retaining the boundary between the Te-domain and a portion of the domain directly upstream (N-terminal) may also contribute to retaining cyclase function of the Te-domain within a new modular configuration.

5 **[0351]** Accordingly, to insert an additional module upstream from a Te-domain and have it be operatively linked thereto, one can identify linker sequences between the C-A-T modules and the C-terminal Te-domain, as described above, and insert sequences encoding the additional module therein, using standard genetic manipulations. Optionally, one can engineer a new, hybrid C-terminal Te-domain in which the C-terminal portion of the penultimate thiolation (T-) domain remains linked (or is otherwise grafted) to the Te-domain ("T-
10 /Te- domain"). See Guenzi et al., 1998, *supra*. Sequences encoding the additional module are then inserted within the identified linker region upstream from a hybrid T-/Te domain using well-known genetic recombination techniques, as described in Example 11A,
15 above.
20

F. Internal deletion of one or more modules to cause the contraction of the Daptomycin ring or shortening of the tail.

[0352] To obtain a deletion of an internal module(s)
25 on the chromosome by double crossing-over and selection on antibiotic plates it is necessary to prepare a plasmid containing a fragment of chromosomal DNA situated upstream from the module to be deleted fused by ligation to a fragment downstream of the module(s) to be deleted.
30 The plasmid also carries a wild type *rpsL* gene to confer streptomycin sensitivity on recombinants in a streptomycin-resistant genetic background (see Example 11A), an antibiotic resistance gene (e.g., apramycin

- 171 -

resistance, thiostrepton resistance or hygromycin resistance) for selection of single crossovers, and a temperature sensitive replicon that can be cured at elevated temperature. A single crossover inserting the plasmid by homologous recombination into the region of DNA upstream of the module(s) to be exchanged can be selected for antibiotic resistance at elevated temperature. The second crossover that deletes the module(s) can then be selected on media containing streptomycin (thus eliminating all plasmid sequences). Recombinants containing deletions of the appropriate module(s) can be verified by Southern blot hybridization of *S. roseosporus* DNA cleaved with appropriate restriction endonucleases. This approach can be taken to delete the L-Asp module or the Gly module from *dptBC*, for example. It can also be used to delete the modules in the *dptA* gene specifying L-Asn, L-Asp or both L-Asn and L-Asp together.

G. Translocation of the terminal thioesterase module to cause the contraction of the Daptomycin ring.

[0353] Sequences encoding the thioesterase (Te) region which resides at the carboxyl terminus of the last module in the daptomycin NRPS (DptD) may be translocated upstream to the end of an internal module encoding region. This translocation will result in the release of a defined shortened product that will yield a truncated linear or cyclic peptide. The translocation of the Te can be accomplished by double crossovers much the same way as described above in Examples 12A and 12F.

Molecular exchange between Daptomycin NRPS and other NRPS or PKS genes

a. *Daptomycin thioesterase onto different NRPS or PKS*

- 172 -

[0354] Using well-known molecular and genetic methods such as those described above, sequences encoding a C-terminal Te-domain of the daptomycin NRPS of the invention (e.g., DptD) may be moved (either alone or in combination with one or more upstream modules or portions thereof) into association with sequences encoding other NRPS or PKS modular genes from a variety of other hosts to produce hybrid modular synthetases that are capable of producing new peptide and/or hybrid peptide/polyketide products having useful properties. See, e.g., Stachelhaus et al., *Science*, 269, pp. 69-72 (1995) and Cane and Khosla, *Chem. Biol.*, 6, pp. 319-325 (1999); each of which is incorporated herein by reference in its entirety. Similarly, daptomycin sequences encoding a free thioesterase of the invention (e.g., DptH) may be moved into association other NRPS or PKS encoding modular genes to produce hybrid modular synthetases.

b. Module and domain exchanges between daptomycin and other NRPS and/or PKS

[0355] Various sequences derived from the daptomycin biosynthetic gene cluster of the invention -- including but not limited to domains and modular structures -- may be used to construct plasmids and other vectors for use in genetic recombination reactions (gene duplication, conversion, replacement, etc.) between daptomycin sequences and natural or synthetic NRPS and PKS sequences in homologous and heterologous hosts to produce hybrid NRPS and hybrid NRPS/PKS modular synthetases comprising sequences from the daptomycin biosynthetic gene cluster. Such hybrid synthetases will produce novel peptide and polyketide products which are expected to have new and useful properties.

- 173 -

Creation of lipopeptide derivatives of nonribosomally-synthesized peptides that are not normally acylated.

[0356] The fatty acid tail of daptomycin is thought to be attached by the products of the *dptE* and *dptF* genes, working in conjunction with the condensation domain at the start of *dptA*. These genes and gene fragments may be transferred to the beginning of a foreign nonribosomal peptide synthase gene, or to an internal location within the daptomycin gene cluster, either at the start of a gene (e.g. 5' of *dptBC* or *dptD*) or within a gene at the start of a module (e.g. 5' of module 2), to create acylated versions of the foreign nonribosomally synthesized peptide, or to create acylated, truncated derivatives of daptomycin. The foreign gene may be derived from another natural organism, or one generated by recombinant techniques, e.g. various versions of daptomycin that have undergone modifications to expand or contract the ring, to have substituted amino acids in the peptide sequence as described herein.

Modification of amino acid stereoisomers in the peptide structure.

[0357] Stereospecificity in the amino acid backbone produced by an NRPS is determined by the presence of epimerase domains in the donor module and distinctive condensation domains in the acceptor module. An alteration in stereochemistry of the amino acids may be achieved by addition of an epimerase domain to a donor module, and substitution of the appropriate condensation domain to the acceptor module. An alteration can also be made by removal of the epimerase domain from a donor module, and the substitution of the appropriate condensation domain in the acceptor, e.g. the epimerase domain can be excised from module 2 of *dptA*, and the

- 174 -

condensation domain of module 3 of *dptA* can be exchanged for the condensation domain from another module that does not normally accept a D-amino acid. Useful epimerase and condensation domains may be found in the daptomycin cluster as well as in other nonribosomal peptide synthetase genes.

EXAMPLE 13: Use of Free Thioesterase

A. Expression of *dptD* or *dptH* related sequences in homologous or heterologous systems to increase efficiency of product formation by modular NRPSs and PKSs

[0358] The C-terminal Te-domain excised from tyrocidine synthetase has been shown to catalyze cyclization on various peptide substrates as long as key "recognition amino acids" are at the C- and N-termini of the substrate. See Trauger et al., *Nature*, 407, 215-218, 2000. Sequences derived from the C-terminal domain of daptomycin NRPS (e.g., *dptD*) may similarly be isolated and expressed - alone or in the form of suitable fusion proteins - in a homologous or heterologous host (or *in vitro* system) to catalyze cyclization of peptide and polyketide products which naturally (or which have been engineered to) possess key substrate recognition amino acids required for the daptomycin Te-domain to bind and join substrate ends (see below).

[0359] When isolating sequences derived from the C-terminal Te-domain of daptomycin synthetase (NRPS) for independent expression, it may be preferable to include natural C-terminal sequences from the penultimate amino acid module. See, e.g., Guenzi et al., 1998, *supra*.

Various *dptD* and upstream-derived sequence combinations can be tested using techniques well-known in the art to optimize the thioesterase activity of the C-terminal Te-

- 175 -

domain of daptomycin NRPS when expressed independently from upstream polypeptides such as DptA and/or DptBC. Independent expression of the C-terminal Te-domain of daptomycin may be accomplished using standard molecular biology techniques. Independent expression of the C-terminal Te-domain of daptomycin NRPS is accomplished by inserting sequences derived from the thioesterase domain of the *dptD* ORF (SEQ ID NO:3) downstream from natural daptomycin NRPS promoter sequences (SEQ ID NO:2) in an appropriately constructed expression vector. Alternatively, independent expression of the C-terminal Te-domain of daptomycin NRPS is accomplished by inserting the thioesterase domain of the *dptD* ORF (SEQ ID NO:3) downstream from a heterologous promoter, which is constitutively active or from a heterologous promoter which may be turned on or off in a regulated manner. Those of skill in the art will appreciate the factors to be considered in selecting appropriate promoters and vectors for expression or over-expression in a host-dependent manner.

[0360] Sequences derived from the free thioesterase domain of the daptomycin biosynthetic gene cluster of the invention (*dptH*) may be similarly expressed in a homologous or heterologous host to test and develop novel cyclic peptides and the like.

[0361] The key recognition amino acids of daptomycin are identified by systematic mutagenesis of the amino acid residues of daptomycin followed by cyclization assays using each modified daptomycin substrate in a reaction catalyzed by the isolated Te-domain. C- and N-terminal amino acid residues required for daptomycin cyclization are identified and engineered into new substrate backbones into which peptide and polyketide

- 176 -

building block units can be inserted. Substrate engineering can be performed at the nucleic acid sequence level or at the peptide level using techniques well-known to those of skill in the art. The length and composition of preferred substrates may be determined empirically, taking into consideration factors well-known to the skilled worker and including (but not limited to) substrate binding efficiency, catalytic rate, biological activity of resulting cyclic product(s), and ease of purification of the final products.

*B. Mutagenize *dptD* or *dptH* to affect proofreading function*

[0362] The *dptH* gene from the daptomycin gene cluster is related to free thioesterase enzymes which are known to participate in the biosynthesis of some peptide and polyketide secondary metabolites. See e.g., Schneider and Marahiel, *Arch. Microbiol.*, 169, pp. 404-410 (1998), and Butler et al., *Chem. Biol.*, 6, pp. 87-292 (1999), hereby incorporated by reference in their entirety. It has been suggested that editing thioesterases are often required for efficient natural product synthesis. Butler et al. have postulated that the free thioesterase found in the polyketide tylosin gene cluster may be involved in editing and proofreading functions, consistent with the suggested role of the thioesterases in efficient product formation.

[0363] Homologous or heterologous expression of the daptomycin *dptH* (encoding a free thioesterase) or the thioesterase-encoding domain of *dptD* (encoding the C-terminal Te) genes may affect the efficiency of product formation by modular NRPSs and PKSs. The proposed editing and proofreading functions of the daptomycin thioesterase type II enzyme (DptH) (and potentially of

- 177 -

the type I thioesterase enzyme when detached from the C-terminus of the daptomycin gene cluster and separately expressed) may be altered by conventional mutagenesis and other recombinant DNA techniques, e.g., those known to affect adversely the fidelity of DNA replication.

Altered and mutated forms of thioesterase genes may be expressed in appropriate expression systems and screened for those which encode thioesterase enzymes having altered biological properties. Especially desirable would be thioesterase enzymes that have higher than normal rates of amino acid misincorporation. Such mutants would be useful for creating a larger diversity of peptide and peptide/polyketide hybrid products having new and useful biological properties.

EXAMPLE 14: Using Daptomycin Biosynthetic Genes to Identify and Isolate Related Genes

[0364] The nucleic acid and amino acid sequences of the invention can be compared to the corresponding sequences from another lipopeptide pathway in order to identify features that can then be used to identify sequences from an NRPS or a component of an NRPS encoding another lipopeptide.

[0365] The amino acid 3-methyl glutamic acid (3MG) is uncommon, but is found in daptomycin, the calcium dependent antibiotic (CDA) from *S. coelicolor*, and the A54145 compound made by *S. fradiae*. Comparison of the *S. roseosporus* and *S. coelicolor* nucleic acid sequences that encode the 3MG adenylation domain, as well as from analogous sequences from genes that adenylate other amino acids, were used to create the primer pair P140 and P141:

P140 ACSSWSGGSGTSSCCTTCATGAA (SEQ ID NO: 160)

- 178 -

P141 ATGGTGTTCGAGAACTAYCC (SEQ ID NO: 161).

[0366] *S. fradiae* cosmid library clones were screened by PCR using P140 and P141 using standard techniques. The PCR reaction yielded a nucleic acid molecule product of approximately 700 bp, whose sequence proved similar to the region encoding the 3MG adenylation domain in *S. roseosporus* and *S. coelicolor*. Extension of the sequence by primer walking confirmed that the region identified was the 3MG module in A54145.

10 [0367] This method was also used to identify portions of an NRPS pathway that encode condensation domains downstream of a D-amino acid activating module. D-amino acids are unusual amino acids found in non-ribosomally synthesized peptides, and primers for condensation domains associated with them can be used to identify pathways with such amino acids. The nucleic acid sequences of the *S. roseosporus* daptomycin and *S. coelicolor* CDA sequences that encode these D-amino acid condensation domains were compared to each other and to analogous sequences from other condensation domains associated with L-amino acids in order to create the primer pair P144 and P145:

P144 SCSTSCAGGAGGGSHTSSTSTTCC (SEQ ID NO: 162)

P145 CCGAASACSACGTCGTCSCGSCC (SEQ ID NO: 163).

25 [0368] *S. fradiae* cosmid library clones were screened by PCR using P144 and P145 using standard techniques. The PCR reaction yielded a nucleic acid molecule products of approximately 800 basepairs, the sequences of which proved to be similar to the condensation domains following the D-amino acids in *S. roseosporus* and *S. coelicolor*. Sequences corresponding to more than one domain were obtained, indicating that the pathway had more than one D-amino acid.

[0369] These approaches, based on understanding the sequence of the daptomycin pathway, can be used to develop special primer sets for other genetic features of lipopeptide pathway gene clusters, such as regions encoding epimerase domains or the condensation domain of the first adenylation module responsible for condensing the fatty acid to the peptide, as well as genes involved in acylation, such as DptE and F.

Table 5

ORF# - Fragment	Nucleotide Sequence SEQ ID NO:	Amino Acid Sequence SEQ ID NO:
1 - 90 kb*	20	19
2 - 90 kb	22	21
3 - 90 kb	24	23
4 - 90 kb	26	25
5 - 90 kb	28	27
6 - 90 kb	30	29
7 - 90 kb	32	31
8 - 90 kb	34	33
9 - 90 kb	36	35
10 - 90 kb	38	37
11 - 90 kb	40	39
12a - 90 kb	42	41
12b - 90 kb	44	43

* ORF-1 of the 90 kb fragment is a partial sequence of the ORF because the 3' end of the ORF, including the stop codon, terminates in the SP6 fragment. The nucleic acid sequence of the 3' end of the ORF-1 sequence, including the stop codon, corresponds to nucleotides 13020-12876 of SEQ ID NO: 103. Thus, the full open reading frame of ORF-1 of the 90 kb fragment consists of SEQ ID NO: 19 (the complementary strand of nucleotides 1635-1 of SEQ ID NO: 1) followed by the complementary strand of nucleotides 13020-12876 of SEQ ID NO: 103.

13 - 90 kb	46	45
14 - 90 kb	48	47
15 - 90 kb	50	49
16 - 90 kb	52	51
17 - 90 kb	54	53
18 - 90 kb	56	55
19 - 90 kb	58	57
20 - 90 kb	60	59
21 - 90 kb	62	61
22 - 90 kb	64	63
23 - 90 kb	66	65
24 - 90 kb	68	67
25 - 90 kb	70	69
26a - 90 kb	72	71
26b - 90 kb	74	73
27 - 90 kb	76	75
28 - 90 kb	78	77
29 - 90 kb <i>dptE</i>	16	15
30 - 90 kb <i>dptF</i>	18	17
31 - 90 kb <i>dptA</i>	10	9
32 - 90 kb <i>dptB</i>	12	11
33 - 90 kb <i>dptC</i>	14	13
34 - 90 kb <i>dptD</i>	3	7
35 - 90 kb	80	79
36 - 90 kb <i>dptH</i>	6	8
37 - 90 kb	82	81
38 - 90 kb	84	83
41 - 90 kb	105	104

- 181 -

1 - SP6	86	85
2 - SP6	88	87
3 - SP6	90	89
4 - SP6	92	91
5 - SP6	94	93
6 - SP6	96	95
7 - SP6	98	97
8 - SP6	100	99
9 - SP6	102	101
2 - GTC2	107	108
3 - GTC2	109	110
4 - GTC2	111	112
5 - GTC2	113	114
6 - GTC2	115	116
7 - GTC2	117	118
8 - GTC2	119	120
9 - GTC2	121	122
10 - GTC2	123	124
11 - GTC2	125	126
12 - GTC2	127	128
13 - GTC2	129	130
14 - GTC2	131	132
15 - GTC2	133	134
16 - GTC2	135	136

Table 6: BlastX Results for ORFs in 90 kb Fragment

ORF	Start	Stop	Str	BLASTX (accession numbers, entry title, P-value, E-value)	Polypeptide
1	1637	1	-	emb CAB88932.1 (AL353863) putative ABC transporter [Strept... pir S57562 strW protein - Streptomyces glaucescens >gj 212... emb CAB88932.1 (AL353863) putative ABC transporter [Streptomyces coelicolor A3(2)] Length = 593 Score = 732 bits (1870), Expect(2) = 0.0 Identities = 367/462 (79%), Positives = 405/462 (87%)	Type III ABC transporter similar to Streptomyces glaucescens strW gene (resistance to streptomycin); has Walker A, B motifs. Translationally coupled to Orf2.
2	3502	1634	-	emb CAB88931.1 (AL353863) putative ABC transporter transme. pir S57561 strV protein - Streptomyces glaucescens >gj 212 emb CAB88931.1 (AL353863) putative ABC transporter transmembrane subunit [Streptomyces coelicolor A3(2)] Length = 623 Score = 854 bits (2183), Expect = 0.0 Identities = 456/637 (71%), Positives = 510/637 (79%), Gaps = 17/637 (2%)	ABC transporter similar to Streptomyces glaucescens strV gene (resistance to streptomycin); has Walker B motif. Translationally coupled to Orf1.
3	5144	3659	-	gi 3913215 1-CARBOXY-3-CHLORO-3,4-DIHYDROXYCYCLO HE gi 3914351 PUTATIVE 4,5-DIHYDROXYPHTHALATE DEHYDRO gi 3913215 sp Q44258 CBAC_ALCSB 1-CARBOXY-3-CHLORO-3,4-DIHYDROXYCYCLO HEXA-1,5-DIENE DEHYDROGENASE Length = 397 Score = 158 (66.0 bits), Expect = 1.6e-10, P = 1.6e-10 Identities = 59/218 (27%), Positives = 180/218 (82%), Gaps = 24/218 (11%)	Oxidoreductase

4	8364	5410	-	gi 2506961 D-LACTATE DEHYDROGENASE [CYTOCHROME], MI... 251 5.1e-21 gi 3023651 D-LACTATE DEHYDROGENASE [CYTOCHROME] PRE... 212 1.9e-16 gi 2506961 sp P32891 DLD1_YEAST D-LACTATE DEHYDROGENASE [CYTOCHROME], MITOCHONDRIAL PRECURSOR (D-LACTATE FERRICYTOCHROME C OXIDOREDUCTASE) (D-LCR) Length = 587 Score = 251 (102.2 bits), Expect = 5.1e-21, P = 5.1e-21 Identities = 119/502 (23%), Positives = 374/502 (74%), Gaps = 91/502 (18%)	Transmembrane, FAD- dependent dehydrogenase
5	8916	8416	-	gi 10803169 emb CAC13097.1 (AL445503) putative marR-family.. 107 3e-23 gi 15896528 ref NP_349877.1 Transcriptional regulator, Mar.. 56 1e-07 gi 10803169 emb CAC13097.1 (AL445503) putative marR-family regulator [Streptomyces coelicolor] Length = 153 Score = 107 bits (268), Expect = 3e-23 Identities = 66/110 (60%), Positives = 79/110 (71%)	Mar family-related protein Transcriptional regulator Involved in antibiotic susceptibility and resistance
6	9030	10853	+	Gb AAF67494.1 AF170880_1 (AF170880) NovA [Streptomyces spe 1017 0.0 emb CAC13096.1 (AL445503) putative ABC transporter ATP-bin 946 0.0 gb AAF67494.1 AF170880_1 (AF170880) NovA [Streptomyces spheroides] Length = 635 Score = 1017 bits (2602), Expect = 0.0 Identities = 526/609 (86%), Positives = 559/609 (91%), Gaps = 3/609 (0%)	NovA-related protein (novobiocin biosynthetic gene cluster) that is ABC transporter; has Walker A, B motifs
7	10933	11544	+	emb CAB91142.1 (AL355913) putative translation initiation ... 64 3e-09 pir JQ0405 hypothetical 119.5K protein (uvrA region) - Mic... 62 7e-09 emb CAB91142.1 (AL355913) putative translation initiation factor IF-2(fragment)[Streptomyces coelicolor A3(2)] Length = 835 Score = 63.6 bits (152), Expect = 3e-09 Identities = 74/237 (31%), Positives = 84/237 (35%), Gaps = 6/237 (2%)	Hypothetical protein with no significant match identified by BlastX

8	11990	12850	+	gi 7688708 gb AAAF67495.1 AF170880_2 (AF170880) NovB [Strept gi 10803167 emb CAC13095.1 (AL445503) conserved hypothetical 319 2e-86 297 9e-80 gi 7688708 gb AAAF67495.1 AF170880_2 (AF170880) NovB [Streptomyces spheroides] Length = 284 Score = 319 bits (817), Expect = 2e-86 Identities = 156/247 (63%), Positives = 188/247 (75%)	NovB-related protein (novobiocin biosynthetic gene cluster)
9	14038	12878	-	gb AAAF67496.1 AF170880_3 (AF170880) NovC [Streptomyces sphe emb CAB71851.1 (AL138667) putative monooxygenase. [Strepto 520 e-146 261 1e-68 gb AAAF67496.1 AF170880_3 (AF170880) NovC [Streptomyces spheroides] Length = 352 Score = 520 bits (1324), Expect = e-146 Identities = 260/346 (75%), Positives = 283/346 (81%), Gaps = 1/346 (0%)	Nov-C related protein that is oxidoreductase
10	14348	14070	-	pir J39929 hypothetical protein orfM - Bacillus subtilis pir D69817 sulfate starvation-induced protein 6 homolog yg 78 2e-14 78 2e-14 pir J39929 hypothetical protein orfM - Bacillus subtilis (fragment) gb AAA64350.1 (L16808) Gene disrupted by Tn917 insertion after base 3033. Translation product hydrophilic, no homologues in the databases.; putative [Bacillus subtilis] Length = 372 Score = 78.0 bits (189), Expect = 2e-14 Identities = 37/53 (69%), Positives = 41/53 (76%)	Monooxygenase
11	15697	14522	-	gi 1723069 HYPOTHETICAL 69.5 KDA PROTEIN RV1364C gi 8928323 SIGMAB REGULATION PROTEIN PHOSPHATASE 2C 86 0.04 85 0.053 gi 1723069 sp Q11034 YD64_MYCTU HYPOTHETICAL 69.5 KDA PROTEIN RV1364C Length = 653 Score = 86 (37.9 bits), Expect = 0.041, P = 0.04 Identities = 45/153 (29%), Positives = 132/153 (86%), Gaps = 6/153 (3%)	Hypothetical protein

12a	17597	16938	-	gi 728850 GLUCOAMYLASE S1/S2 PRECURSOR (GLUCAN 1,4 GLYCOPROTEIN X PRECURSOR 91 0.0072 gi 728850 sp P08640 AMYH_YEAST GLUCOAMYLASE S1/S2 PRECURSOR (GLUCAN 1,4-ALPHA-GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE) Length = 1367 Score = 113 (48.4 bits), Expect = 1.9e-05, P = 1.9e-05 Identities = 47/186 (25%), Positives = 158/186 (84%), Gaps = 12/186 (6%)	113 1.9e-05 gi 138350	Hypothetical protein
12b	17870	18682	+	gi 8546911 emb CAB94663.1 (AL359216) hypothetical protein ... 34 1.3 gi 8546913 emb CAB94625.1 (AL359215) putative membrane pro... 33 2.9 gi 8546911 emb CAB94663.1 (AL359216) hypothetical protein SC1D2.05 (fragment). [Streptomyces coelicolor A3(2)] Length = 192 Score = 34.3 bits (77), Expect = 1.3 Identities = 28/94 (29%), Positives = 40/94 (41%), Gaps = 5/94 (5%)		Hypothetical Protein
13	19898	18915	-	emb CAB94641.1 (AL359215) putative iron transport lipoprot... 250 2e-65 pir C83282 hypothetical protein PA2913 [imported] - Pseudo... 168 1e-40 emb CAB94641.1 (AL359215) putative iron transport lipoprotein. [Streptomyces coelicolor A3(2)] Length = 345 Score = 250 bits (632), Expect = 2e-65 Identities = 133/322 (41%), Positives = 188/322 (58%), Gaps = 13/322 (4%)		Iron (ABC) transporter Association with orfs 14 and 15
14	20674	19907	-	emb CAB94640.1 (AL359215) putative iron transport protein,... 279 3e-74 emb CAC14366.1 (AL445963) Fe uptake system permease [Strep... 250 2e-65 emb CAB94640.1 (AL359215) putative iron transport protein, ATP-binding component. [Streptomyces coelicolor A3(2)] Length = 258 Score = 279 bits (706), Expect = 3e-74 Identities = 141/251 (56%), Positives = 181/251 (71%)		Iron transporter Association with orfs 13 and 15

15	21782	20676	-	embjCAB94639.1 (AL359215) putative FecCD-family membrane t embjCAC14365.1 (AL445963) Fe uptake system integral membra embjCAB94639.1 (AL359215) putative FecCD-family membrane transport protein.[Streptomyces coelicolor A3(2)] Length = 368 Score = 371 bits (943), Expect = e-102 Identities = 192/365 (52%), Positives = 248/365 (67%)	371 e-102 277 2e-73	Iron transporter Association with orfs 13 and 14
16	23130	21877	-	gij138350 GLYCOPROTEIN X PRECURSOR gij728850 GLUCOAMYLASE S1/S2 PRECURSOR (GLUCAN 1,4... gij138350 sp P28968 VGLX_HSVB GLYCOPROTEIN X PRECURSOR Length = 797 Score = 94 (41.0 bits), Expect = 0.0088, P = 0.0088 Identities = 51/216 (23%), Positives = 181/216 (83%), Gaps = 9/216 (4%)	94 0.0088 83 0.16	Hypothetical protein
17	23987	23127	-	gij14591289 ref NP_143367.1 hypothetical protein [Pyrococc... gij322598 pir S28604 St12p protein - Arabidopsis thaliana gij14591289 ref NP_143367.1 hypothetical protein [Pyrococcus horikoshii] Length = 248 Score = 46.2 bits (108), Expect = 3e-04 Identities = 31/119 (26%), Positives = 62/119 (52%), Gaps = 2/119 (1%)	46 3e-04 42 0.006	Hypothetical protein
18	24966	23953	-	gij543960 CYSTATHIONINE BETA-SYNTHASE (SERINE SULF CYSTEINE SYNTHASE (O-ACETYL SERINE SULFHY... gij543960 sp P32232 CBS_RAT CYSTATHIONINE BETA-SYNTHASE (SERINE SULFHYDRASE) (BETA-THIONASE) (HEMOPROTEIN H-450) Length = 561 Score = 162 (67.5 bits), Expect = 4.3e-11, P = 4.3e-11 Identities = 76/290 (26%), Positives = 243/290 (83%), Gaps = 17/290 (5%)	162 4.3e-11 147 2.4e-09	Hypothetical protein

19	25228	26127	+	gij8928195 MEVALONATE KINASE (MK) gij8928178 MEVALONATE KINASE (MK) gij8928195 sp Q9V187 KIME_PYRAB MEVALONATE KINASE (MK) Length = 335 Score = 99 (43.0 bits), Expect = 0.00096, P = 0.00096 Identities = 25/61 (40%), Positives = 49/61 (80%)	99 0.00096 90 0.011	Hypothetical protein
20	26445	27212	+	gij731172 SKIN SECRETORY PROTEIN XP2 PRECURSOR (AP... gij127749 MYOSIN IC HEAVY CHAIN gij731172 sp P17437 XP2_XENLA SKIN SECRETORY PROTEIN XP2 PRECURSOR (APEG PROTEIN) Length = 439 Score = 87 (38.3 bits), Expect = 0.019, P = 0.019 Identities = 20/54 (37%), Positives = 39/54 (72%)	87 0.019 86 0.025	Hypothetical protein
21	28124	27381	-	emb CAB56736.1 (AL121600) ABC transport protein, ATP-bindi. pir H75293 probable manganese ABC transporter, ATP-binding... emb CAB56736.1 (AL121600) ABC transport protein, ATP-binding subunit [Streptomyces coelicolor A3(2)] Length = 252 Score = 351 bits (892), Expect = 4e-96 Identities = 181/247 (73%), Positives = 193/247 (77%)	351 4e-96 154 1e-36	ABC Transporter (Mn transporter)
22	28139	29098	+	emb CAB56735.1 (AL121600) ABC transporter protein, integra... pir G75293 probable manganese ABC transporter, permease pr... emb CAB56735.1 (AL121600) ABC transporter protein, integral membrane subunit [Streptomyces coelicolor A3(2)] Length = 283 Score = 462 bits (1177), Expect = e-129 Identities = 241/272 (88%), Positives = 252/272 (92%)	462 e-129 208 1e-52	ABC transporter (integral membrane protein) Role in Mn or Fe transport

23	29095	30285	+	gi 6002369 emb CAB56734.1 (AL121600) hypothetical protein ... 484 e-136 gi 13592175 gb AAK31375.1 AC084329_1 (AC084329) ppg3 [Leish... 61 2e-08 gi 6002369 emb CAB56734.1 (AL121600) hypothetical protein SCF76.14c [Streptomyces coelicolor A3(2)] Length = 415 Score = 484 bits (1247), Expect = e-136 Identities = 245/395 (62%), Positives = 287/395 (72%), Gaps = 1/395 (0%)	Hypothetical protein
24	30282	31244	+	gi 6002368 emb CAB56733.1 (AL121600) putative solute-bindin... 439 e-122 gi 15807666 ref NP_296243.1 adhesin B [Deinococcus radiodu... 123 2e-27 gi 6002368 emb CAB56733.1 (AL121600) putative solute-binding lipoprotein [Streptomyces coelicolor A3(2)] Length = 329 Score = 439 bits (1128), Expect = e-122 Identities = 222/315 (70%), Positives = 253/315 (79%)	ABC transporter protein Translationally coupled to orf 23
25	31332	32537	+	emb CAB56732.1 (AL121600) putative secreted protein [Strep... 620 e-176 gb AAA59875.1 (M74027) mucin [Homo sapiens] 130 3e-29 emb CAB56732.1 (AL121600) putative secreted protein [Streptomyces coelicolor A3(2)] Length = 402 Score = 620 bits (1581), Expect = e-176 Identities = 299/402 (74%), Positives = 341/402 (84%), Gaps = 1/402 (0%)	Hypothetical Protein
26a	32816	33427	-	gi 8039818 HYPOTHETICAL 23.1 KDA PROTEIN MLC581.27 159 5.3e-11 gi 2829591 HYPOTHETICAL 23.0 KDA PROTEIN RV2637 143 4e-09 gi 8039818 sp Q49642 YQ37_MYCLE HYPOTHETICAL 23.1 KDA PROTEIN MLC581.27 Length = 214 Score = 159 (66.3 bits), Expect = 5.3e-11, P = 5.3e-11 Identities = 57/197 (28%), Positives = 166/197 (84%), Gaps = 14/197 (7%)	Hypothetical protein

26b	32686	32868	+	gi 15805506 ref NP_294202.1 penicillin-binding protein 1 [...] gi 7248459 gb AAAF43497.1 AF134579_1 (AF134579) arabinogalac... 33 0.72 32 0.95 gi 15805506 ref NP_294202.1 penicillin-binding protein 1 [Deinococcus radiodurans] gi 7473266 pir B75514 penicillin-binding protein 1 - Deinococcus radiodurans (strain R1) gi 6458167 gb AAAF10059.1 AE001907_5 (AE001907) penicillin-binding protein 1 [Deinococcus radiodurans] Length = 873 Score = 32.7 bits (73), Expect = 0.72 Identities = 24/55 (43%), Positives = 28/55 (50%)	Hypothetical Protein
27	34195	35154	+	pir T36741 probable ABC-type transport system ATP-binding ... gb AAD44229.1 AF143772_35 (AF143772) DrrA [Mycobacterium av... 291 6e-78 290 2e-77 pir T36741 probable ABC-type transport system ATP-binding protein - Streptomyces coelicolor emb CAB50934.1 (AL096849) putative ABC-transporter ATP-binding protein [Streptomyces coelicolor A3(2)] Length = 332 Score = 291 bits (738), Expect = 6e-78 Identities = 168/303 (55%), Positives = 204/303 (66%), Gaps = 2/303 (0%)	Type I ABC transporter similar to daunorubicin resistance gene, DrrA, in Streptomyces antibioticus; has Walker A, B motifs.
28	35148	36017	+	pir S32909 hypothetical protein 5 - Streptomyces antibioti... pir T50567 probable ABC-type transport protein, transmembr... 120 2e-26 115 6e-25 pir S32909 hypothetical protein 5 - Streptomyces antibioticus gb AAA26794.1 (L06249) membrane protein [Streptomyces antibioticus] Length = 273 Score = 120 bits (299), Expect = 2e-26 Identities = 72/226 (31%), Positives = 113/226 (49%)	ABC transporter (integral membrane protein) similar to daunorubicin resistance gene, DrrB, in Streptomyces antibioticus; has Walker A, B motifs.

- 190 -

35	85270	85497	+	pir T36310 probable small conserved hypothetical protein S... 111 9e-25 gb AAG29779.1 AF235050_2 (AF235050) CumB [Streptomyces rish... 101 1e-21 pir T36310 probable small conserved hypothetical protein SCE8.11c - Streptomyces coelicolor gb AAD18046.1 (AF124138) Cda-orfX [Streptomyces coelicolor A3(2)] emb CAB38589.1 (AL035654) putative small conserved hypothetical protein [Streptomyces coelicolor A3(2)] Length = 71 Score = 111 bits (276), Expect = 9e-25 Identities = 46/67 (68%), Positives = 56/67 (82%)	Hypothetical Protein
37	86434	87420	+	pir T36307 hypothetical protein SCE8.08c - Streptomyces co... 175 7e-43 gb AAA59875.1 (M74027) mucin [Homo sapiens] 94 3e-18 pir T36307 hypothetical protein SCE8.08c - Streptomyces coelicolor emb CAB38586.1 (AL035654) hypothetical protein [Streptomyces coelicolor A3(2)] Length = 338 Score = 175 bits (439), Expect = 7e-43 Identities = 120/330 (36%), Positives = 164/330 (49%), Gaps = 13/330 (3%)	Hypothetical Protein Translationally coupled to orf 38

- 191 -

38	87417	88154	+	pir E83323 hypothetical protein PA2579 [imported] - Pseudo... pir G75588 probable tryptophan 2,3-dioxygenase - Deinococc... pir G75588 probable tryptophan 2,3-dioxygenase - Deinococcus radiodurans (strain R1) gb AAF12443.1 AE001863_68 (AE001863) tryptophan 2,3-dioxygenase, putative [Deinococcus radiodurans] Length = 287 Score = 87.4 bits (213), Expect = 2e-16 Identities = 73/259 (28%), Positives = 107/259 (41%), Gaps = 37/259 (14%)	102 3e-21 87 2e-16	Hypothetical Protein Translationally coupled to orf37
41	89910	90563	+	gij 7480757 pir T36281 probable hydrolase - Streptomyces co... gij 7480757 pir T36281 probable hydrolase - Streptomyces coelicolor gij 5123678 emb CAB45367.1 (AL079345) putative hydrolase [Streptomyces coelicolor A3(2)] Length = 215 Score = 114 bits (285), Expect = 2e-24 Identities = 72/170 (42%), Positives = 96/170 (56%), Gaps = 3/170 (1%)	114 2e-24	

Str refers to whether the gene is encoded on the DNA molecule (relative to SEQ ID NO: 1) from left to right (+) or from right to left on the complementary strand.

5 The BlastX box contains the two top BlastX scores for each ORF (top two lines) and details regarding the database protein entry and the alignment of the ORF to the database entry.

Table 7: BlastX Results for ORFs in SP6 Fragment

ORF	start	stop	Str	BlastX (accession numbers, entry title, P-value, E-value)	Polypeptide
1	965	1	-	pir T34645 hypothetical protein SC10H5.07 SC10H5.07 - Stre... 352 2e-96 pir T36710 hypothetical protein SCH69.11c - Streptomyces c... 206 2e-52 pir T34645 hypothetical protein SC10H5.07 SC10H5.07 - Streptomyces coelicolor emb CAA20279.1 (AL031232) hypothetical protein SC10H5.07 [Streptomyces coelicolor A3(2)] Length = 469 Score = 352 bits (904), Expect = 2e-96 Identities = 179/305 (58%), Positives = 216/305 (70%)	Hypothetical Protein
2	989	1948	-	pir T35566 probable integral membrane protein - Streptomyc... 206 3e-52 gb AAA53486.1 (U03114) unknown [Streptomyces albus] 139 3e-32 pir T35566 probable integral membrane protein - Streptomyces coelicolor emb CAA20393.1 (AL031317) putative integral membrane protein [Streptomyces coelicolor] Length = 315 Score = 206 bits (523), Expect = 3e-52 Identities = 114/311 (36%), Positives = 180/311 (57%), Gaps = 2/311 (0%)	Hypothetical Protein
3	2099	2392	+		Hypothetical Protein
4	3277	2405	-	emb CAB88937.1 (AL353863) acyl-coA thioesterase [Streptomy... 535 e-151 emb CAB87210.1 (AL163641) acyl CoA thioesterase II [Strept... 293 1e-78 emb CAB88937.1 (AL353863) acyl-coA thioesterase [Streptomyces coelicolor A3(2)] Length = 288 Score = 535 bits (1379), Expect = e-151 Identities = 258/288 (89%), Positives = 273/288 (94%)	Acyl CoA thioesterase; enzyme involved in short chain fatty acid biosynthesis

5	5885	3312	-	emb CAB88936.1 (AL353863) putative helicase [Streptomyces ... 548 e-155 gb AAG45420.1 AF309494_1 (AF309494) vegetative cell wall pr... 121 1e-26 emb CAB88936.1 (AL353863) putative helicase [Streptomyces coelicolor A3(2)] Length = 854 Score = 548 bits (1413), Expect = e-155 Identities = 266/323 (82%), Positives = 291/323 (89%)	DNA helicase
6	5963	6754	+	emb CAB88935.1 (AL353863) putative integral membrane prote... 491 e-138 gb AAK31375.1 AC084329_1 (AC084329) ppg3 [Leishmania major] 106 2e-22 emb CAB88935.1 (AL353863) putative integral membrane protein [Streptomyces coelicolor A3(2)] Length = 264 Score = 491 bits (1265), Expect = e-138 Identities = 235/264 (89%), Positives = 246/264 (93%), Gaps = 1/264 (0%)	Hypothetical Protein
7	6850	8403	+	sp Q9FCB1 DNLJ_STRCO PROBABLE DNA LIGASE (POLYDEOXYRIBONUCL... 461 e-141 ref NP_337667.1 DNA ligase [Mycobacterium tuberculosis CDC... 294 4e-85 sp Q9FCB1 DNLJ_STRCO PROBABLE DNA LIGASE (POLYDEOXYRIBONUCLEOTIDE SYNTHASE [ATP]) emb CAC01484.1 (AL391017) putative DNA ligase [Streptomyces coelicolor A3(2)] Length = 512 Score = 461 bits (1186), Expect(2) = e-141 Identities = 252/341 (73%), Positives = 267/341 (77%)	DNA Ligase

- 194 -

8	9860	8433	-	emb CAB93757.1 (AL357613) putative oxidoreductase. [Strept... pir T34726 probable dehydrogenase - Streptomyces coelicolo... 299 8e-81 130 9e-30 emb CAB93757.1 (AL357613) putative oxidoreductase. [Streptomyces coelicolor A3(2)] Length = 481 Score = 299 bits (766), Expect = 8e-81 Identities = 147/185 (79%), Positives = 165/185 (88%), Gaps = 1/185 (0%)	Oxidoreductase
9	10784	9921	-	emb CAB57411.1 (AL121746) hypothetical protein SCF73.06c [... gb AAK61383.1 (AY035849) basic proline-rich protein [Sus s... 311 3e-84 115 6e-25 emb CAB57411.1 (AL121746) hypothetical protein SCF73.06c [Streptomyces coelicolor A3(2)] Length = 333 Score = 311 bits (798), Expect = 3e-84 Identities = 166/264 (62%), Positives = 182/264 (68%)	Hypothetical Protein

Str refers to whether the gene is encoded on the DNA molecule (relative to SEQ ID NO: 1) from left to right (+) or from right to left on the complementary strand.

5 The BlastX box contains the two top BlastX scores for each ORF (top two lines) and details regarding the database protein entry and the alignment of the ORF to the database entry.

Table 8: BlastX Results for ORFs in GTC Fragment

ORF	start	stop	frame	DNA	Polyptide
2	2941	74	-3	>gil7435848 pir S72176 thermolysin (EC 3.4.24.27) precursor - Bacillus caldolyticus (strain YP-T) gi 995782 gb AAB18652.1 (U25629) neutral proteinase [Bacillus caldolyticus] Length = 546 Score = 180 bits (457), Expect = 7e-44 Identities = 159/550 (28%), Positives = 251/550 (44%), Gaps = 20/550 (3%)	Thermostable neutral protein.
3	3078	4103	3	>gil8977943 emb CAB95810.1 (AL1359949) putative transcriptional regulator [Streptomyces coelicolor A3(2)] Length = 343 Score = 89.0 bits (219), Expect = 7e-17 Identities = 93/335 (27%), Positives = 132/335 (38%), Gaps = 11/335 (3%)	Positive regulatory gene for daptomycin synthesis
4	5246	4131	-2	>gil6434729 emb CAB61176.1 (AL132973) putative DeoR-family transcriptional regulator [Streptomyces coelicolor A3(2)] Length = 368 Score = 179 bits (454), Expect = 4e-44 Identities = 133/361 (36%), Positives = 171/361 (46%), Gaps = 7/361 (1%)	Negative regulatory gene for daptomycin synthesis
5	5536	6888	1	>gil6434730 emb CAB61177.1 (AL132973) probable solute-binding lipoprotein. [Streptomyces coelicolor A3(2)] Length = 443 Score = 226 bits (575), Expect = 7e-58 Identities = 129/358 (36%), Positives = 184/358 (51%), Gaps = 5/358 (1%)	ABC Transporter
6	7017	7814	3	>gil6434731 emb CAB61178.1 (AL132973) putative binding protein dependent transport protein. [Streptomyces coelicolor A3(2)] Length = 328 Score = 243 bits (619), Expect = 3e-63 Identities = 124/239 (51%), Positives = 163/239 (67%), Gaps = 5/239 (2%)	ABC Transporter
7	7943	8743	2	>gil6434732 emb CAB61179.1 (AL132973) putative binding protein dependent transport protein. [Streptomyces coelicolor A3(2)] Length = 287 Score = 265 bits (677), Expect = 5e-70 Identities = 131/252 (51%), Positives = 169/252 (66%)	ABC Transporter

8	8815	9795	1	>gil6434733 emb CAB61180.1 (AL132973) putative 2-hydroxyacid-family dehydrogenase. [Streptomyces coelicolor A3(2)] Length = 343 Score = 190 bits (482), Expect = 3e-47 Identities = 120/330 (36%), Positives = 166/330 (49%), Gaps = 1/330 (0%)	Dehydrogenase?
9	9843	11852	3		Hypothetical protein
10	11860	12738	1		Hypothetical protein
11	13799	12783	-2		Hypothetical protein
12	14051	14674	2		Hypothetical protein
13	14671	15846	1	>gil7480768 pir T35943 probable hydrolytic protein - Streptomyces coelicolor gil4158202 emb CAA22765.1 (AL035206) putative hydrolytic protein [Streptomyces coelicolor A3(2)] Length = 464 Score = 145 bits (366), Expect = 8e-34 Identities = 78/198 (39%), Positives = 105/198 (52%), Gaps = 12/198 (6%)	Hydrolase
14	15954	19265	3	>gil8894813 emb CAB96009.1 (AL360055) hypothetical protein [Streptomyces coelicolor A3(2)] Length = 833 Score = 369 bits (948), Expect = e-101 Identities = 271/831 (32%), Positives = 378/831 (44%), Gaps = 33/831 (3%)	SpoVK-like protein
15	19262	20530	2	>gil7481390 pir T42024 probable serine proteinase - Streptomyces coelicolor gil1151075 gb AAA85224.1 (U33176) serine protease [Streptomyces coelicolor] Length = 390 Score = 68.2 bits (165), Expect = 2e-10 Identities = 63/209 (30%), Positives = 88/209 (41%), Gaps = 21/209 (10%)	Serine protease
16	23947	20585	-2	>gil3413388 emb AL031231.1 SC3C3 Streptomyces coelicolor cosmid 3C3 Length = 31382 Score = 421 bits (914), Expect(6) = 0.0 Identities = 180/291 (61%), Positives = 222/291 (75%)	FtsK/SpoIIIE homologue

EXAMPLE 15: Heterologous Production of Daptomycin in
Streptomyces lividans in the Absence of Actinorhodin

[0370] Both genetic and medium effects were modified to improve the expression of A21978C lipopeptides in an
5 heterologous host. Various strains containing the dpt gene cluster BAC, along with control strains without the gene cluster, were grown in shake-flask fermentation and clarified broths analyzed for the presence of the A21978C lipopeptide series by HPLC. The dpt cluster on the BAC
10 clone B1203A05 was introduced into *S.lividans* by protoplast generation using standard techniques (Keiser, T., et al., Practical Streptomycete genetics. John Innes Foundation, Norwich, 2000). Strains examined included both *S.lividans* TK23 and TK64 strains containing the dpt
15 gene cluster and a genetically altered version of *S.lividans* TK23 with a partially deleted actinorhodin pathway. Other comparable and suitable act knockout strains are known to those in the art. TK64 differs from TK23 in possessing an rpsL (*str-6*) mutation conferring
20 resistance to streptomycin, which has also been implicated in enhancement in the production of actinorhodin (Shima et al., J. Bacteriol, 178 (24), 7276-7284 (1996)). The actinorhodin family are colored polyketides produced in copious quantities by *S.lividans*
25 under many fermentation conditions and which interfere with the detection and purification of other secondary metabolites from the fermentation.

[0371] To eliminate actinorhodin production from *S. lividans*, a cassette was constructed to delete part of
30 the pathway. An 8kb fragment containing the actinorhodin polyketide synthase pathway (Malpartida and Hopwood, Mol.

Gen. Genet., 205, 66-73 (1986)) was cloned into a pUC19; 1.4 kb of DNA was removed from the center of this fragment, thus deleting the 3' end of actIorfI and almost all of actIorfII. This fragment was then replaced by the
 5 resistance marker *ermE* (Bibb et al., Gene, 38 (1-3), 15-26 (1985)). This deletion cassette was then transferred to the temperature sensitive plasmid pGM160 and introduced into *S. lividans* TK23. These recombinant strains were then fermented for 40 hr before plating on to selective
 10 media, from this screening several colonies were isolated with the appropriate phenotype. The genotype of these strains was then confirmed by Southern blots. *S. lividans* strains of both TK64 and TK23 containing the BAC vector alone were also examined as control strains. See Table 9
 15 for strain notation.

Table 9

CBUK strain number	Lineage	Presence of act	rpsL status	Transforming DNA
136736	TK64	+	<i>str-6</i>	BAC vector only
136742	TK64	+	<i>str-6</i>	B12:03A05 dpt gene cluster
137028	TK23	+	+	BAC vector only
137027	TK23	+	+	B12:03A05 dpt gene cluster
137026	TK23 (521)	-	+	BAC vector only
137024	TK23 (521)	-	+	B12:03A05 dpt gene cluster

+ in rpsL column indicates wild type status

[0372] Although a number of different media were
 20 initially explored, two different media were examined in more detail for their ability to support production of the A21978C lipopeptides in *S. lividans*. Both of these media also support good production of the A21978C lipopeptides in *S. roseosporus*. Medium A was a complex
 25 medium consisting of 1% glucose (BDH), 2% soluble starch (Sigma), 0.5% yeast extract (Difco), 0.5% casein (Sigma). 4.6% MOPS (Sigma), adjusted to pH 7 and autoclaved. Medium B was a defined medium consisting of 2% glycerol, 0.25% sucrose, 1.2% proline 1.5% MOPS, 0.056% K₂HPO₄,

- 199 -

0.05% NaCl, together with trace elements and vitamins, adjusted to pH 7 and filter sterilized.

[0373] Fermentations were initiated by inoculation of an enriched oatmeal slope containing 100mg/L apramycin with approximately 0.25 ml material from a cryovial stored at -135°C. After 7-10 days incubation at 28 °C, a mixed mycelial and spore suspension was generated by the addition of 4 ml 0.1% Tween 80 and 2ml inoculated into 40ml of seed medium containing 25mg/L apramycin in a baffled flask to initiate the seed stage. Seed flasks were shaken at 240 rpm and 30 °C for 24-28 hours before a 5% transfer to production flasks containing 50 ml of medium A or B. Replicate flasks were sampled from day 2 until day 6 of the production fermentation period by aseptically removing approximately 1 ml broth, centrifuging for 10 min. at 10,000 rpm and analyzing the supernatant by HPLC. Analysis was performed at ambient temperature using a Waters Alliance 2690 HPLC system and a 996 PDA detector with a 4.6 x 50 mm Symmetry C8 3.5 µM column and a Phenomenex Security Guard C8 cartridge. The gradient initially holds at 90% water and 10% acetonitrile for 2.5 min., followed by a linear gradient over 6 minutes to 100% acetonitrile. The flow rate was 1.5 ml min⁻¹ and the gradient was buffered with 0.01% trifluoroacetic acid. Up to 50 microliters of the supernatant was injected to monitor for production of the native A21987C lipopeptides.

[0374] Confirmation of expected molecular weights was obtained by LC-MS analysis using a Finnigan SSQ710c system using electrospray ionization in positive ion mode, with a scan range of 200-2000 Daltons and 2 second scans. The LC method was run on a Waters Symmetry C8 column (2.1x 50mm 3.5µm particle size). The method held

- 200 -

at the initial conditions of 90% water, 10% acetonitrile and 0.01% formic acid for 0.5 minutes, followed by a linear gradient to 100% acetonitrile and 0.01% formic acid over 6 min. The method then held for 3.5 min..

5 before re-equilibration. The method was run at ambient temperature.

[0375] The heterologous expression of the A21978C lipopeptide series in *S.lividans* TK64 (136742) in medium A was analyzed by HPLC. Production of three of the
10 A21978C lipopeptides with characteristic UV/visible spectra was evident, with retention times of 5.61, 5.77 and 5.89 minutes (λ_{max} 223.8, 261.5 and 364.5 nm) under the analytical conditions stated above. On LC-MS analysis, these three A21978C lipopeptides yielded
15 molecular ions (M-H)⁺ at m/z of 1634.7, 1648.7 and 1662.7, which is in agreement with the masses reported for the major A21978C lipopeptide metabolites C₁, C₂ and C₃ respectively produced by *Streptomyces roseosporus* (Debono et al., J. Antibiotics, XL (6), 761-777 (1987)). A
20 similar product profile was also obtained for heterologous expression of the dpt gene cluster in *S.lividans* TK23 (137027) under the same conditions. Similarly high production levels of actinorhodin were observed in this strain despite the absence of the rpsL
25 mutation that is reported to enhance actinorhodin production. A21978C lipopeptides were not detected in fermentations of the TK64 control strain (136736) or the TK23 control strain (137028) with the BAC vector only integrated.

30 [0376] The amount of A21978C lipopeptides produced in crude broth by these strains could not be accurately quantitated due to co-chromatography with host peaks, including members of the CDA complex; however, a total

- 201 -

maximum yield of the three main lipopeptides was estimated at approximately 20 mg/L. The A21978C lipopeptides were produced early on in the fermentation along with numerous other host metabolites.

5 **[0377]** The profile of production of the A21978C lipopeptides was also observed from fermentations of the *S.lividans* TK23 (137024) act knockout strain in medium A. Absence of an intact act pathway in this strain allowed application of the defined medium B, in which normally
10 high levels of act are supported. Variations of the defined medium were evaluated and a 2 to 4 g/L level of K_2HPO_4 was found advantageous for both production of the A21978C lipopeptides and suppression of some of the host metabolites. HPLC analysis revealed a much cleaner HPLC
15 profile obtained from crude broths of 137024 grown in a higher phosphate supplemented medium at 50 hrs, as compared to production of the A21978C lipopeptides in medium B without the phosphate supplementation. As the fermentation progressed and phosphate derepression
20 occurred, the level and diversity of host metabolites increased, although never to the level previously observed in medium A. Although, early in the fermentation, the production of many host metabolites was suppressed, the production of the CDA series of
25 lipopeptides was not. CDA can exist in both non-phosphorylated and phosphorylated forms. Under the chromatography conditions used, the non-phosphorylated forms of CDA co-chromatographed in the same region as the A21978C lipopeptides and complicated detection and
30 quantification. Incorporation of phosphate into the fermentation medium biased production, at least initially, to the phosphorylated forms of CDA, which were well resolved from the three A21978C lipopeptides by

- 202 -

HPLC. This effect on CDA production was also clearly evident from fermentation of the control strain of the *S.lividans* TK23 act knockout strain with an integrated BAC plasmid not containing the dpt gene cluster (137026)
5 in high phosphate supplemented medium B.

Claims:

1. An isolated nucleic acid molecule comprising a nucleic acid sequence encoding a daptomycin non-ribosomal peptide synthetase (NRPS) or subunit thereof from *Streptomyces roseosporus*, wherein the nucleic acid molecule encodes DptBC and wherein said nucleic acid molecule is not pRHB159.

2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule encodes a polypeptide comprising amino acid sequence SEQ ID NO: 11.

3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is *dptBC* or wherein the nucleic acid molecule comprises nucleic acid sequence SEQ ID NO: 12.

4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule comprises a nucleic acid sequence from an *S. roseosporus* nucleic acid sequence from BAC clone B12:03A05.

5. An isolated nucleic acid molecule that encodes a daptomycin NRPS or subunit thereof, wherein the isolated nucleic acid molecule selectively hybridizes to a reference nucleic acid molecule that encodes a daptomycin NRPS or subunit thereof, wherein the reference nucleic acid molecule comprises a nucleic acid sequence selected from the group consisting of:

- a) the nucleic acid sequence of *dptBC*;
- b) a nucleic acid sequence encoding the amino acid sequence of DptBC;
- c) a nucleic acid sequence encoding the amino acid sequence SEQ ID NO: 11; and

- 204 -

(d) the nucleic acid sequence SEQ ID NO: 12;

wherein said nucleic acid molecule is not
15 pRHB159.

6. The isolated nucleic acid molecule according to claim 5, wherein the nucleic acid molecule hybridizes under conditions selected from the group consisting of low stringency conditions, moderate stringency conditions
5 and high stringency conditions.

7. An isolated nucleic acid molecule that encodes a daptomycin NRPS or subunit thereof, wherein the isolated nucleic acid molecule comprises a nucleic acid sequence that has at least 50% sequence identity to a
5 nucleic acid sequence selected from the group consisting of:

(a) the nucleic acid sequence of dptBC;

(b) a nucleic acid sequence encoding the amino acid sequence of DptBC;

10 (c) a nucleic acid sequence encoding the amino acid sequence SEQ ID NO: 11; and

(d) the nucleic acid sequence SEQ ID NO:
12;

wherein said nucleic acid molecule is not
15 pRHB159.

8. The nucleic acid molecule according to claim 7, wherein the homologous molecule exhibits at least 60% sequence identity to the nucleic acid sequence of any one of (a) to (d).

- 205 -

9. The nucleic acid molecule according to claim 8, wherein the sequence identity is at least 70%.

10. The nucleic acid molecule according to claim 9, wherein the sequence identity is at least 80%.

11. The nucleic acid molecule according to claim 10, wherein the sequence identity is at least 90%.

12. The nucleic acid molecule according to claim 11, wherein the sequence identity is at least 95%.

13. An isolated nucleic acid molecule that encodes a daptomycin NRPS or subunit thereof, wherein the isolated nucleic acid molecule is an allelic variant of a nucleic acid molecule that comprises a nucleic acid
5 sequence selected from the group consisting of:

(a) the nucleic acid sequence of dptBC;

(b) a nucleic acid sequence encoding the amino acid sequence of DptBC;

(c) a nucleic acid sequence encoding the
10 amino acid sequence SEQ ID NO: 11; and

(d) the nucleic acid sequence SEQ ID NO:
12.

14. An isolated nucleic acid molecule comprising a part of a nucleic acid sequence, wherein
15 said part is at least 14 nucleotides, selected from the group consisting of:

- 206 -

(a) a nucleic acid sequence encoding the amino acid sequence of a polypeptide comprising amino acids 5003 to 5007 of SEQ ID NO: 11; and

20 (b) a nucleic acid sequence comprising nucleotides 68350 to 68364 of SEQ ID NO: 12;

wherein said nucleic acid molecule is not pRHB159.

15. The isolated nucleic acid molecule according to claim 14, wherein said nucleic acid molecule encodes at least one domain from a daptomycin NRPS.

16. The isolated nucleic acid molecule according to claim 14, wherein said nucleic acid molecule encodes at least one module from a daptomycin NRPS.

17. The nucleic acid molecule according to claim 14, wherein the part comprises at least 17 nucleotides of the nucleic acid sequence.

18. The nucleic acid molecule according to claim 17, wherein the part comprises at least 20 nucleotides of the nucleic acid sequence.

19. The nucleic acid molecule according to claim 18, wherein the part comprises at least 25 nucleotides of the nucleic acid sequence.

20. The nucleic acid molecule according to claim 19, wherein the part comprises at least 50 nucleotides of the nucleic acid sequence.

21. The nucleic acid molecule according to any one of claims 14-20 that is an oligonucleotide from 14 to 60 nucleotides in length.

22. A vector comprising the nucleic acid molecule according to any one of claims 1-21.

23. The vector according to claim 22, wherein the vector comprises expression control sequences controlling the transcription of the nucleic acid molecule.

24. The vector according to claim 23 wherein the expression control sequences control the expression of the nucleic acid molecule in a prokaryotic cell.

25. A host cell comprising the nucleic acid molecule according to any one of claims 1-21.

26. A host cell comprising the vector according to any one of claims 22-24.

27. A method for producing DptBC, or a domain, module or part thereof, comprising the step of culturing the host cell according to claims 25 or 26 under conditions in which the polypeptide is produced,
5 optionally comprising the step of isolating the polypeptide.

28. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of

(a) the amino acid sequence encoded by the nucleic acid sequence of *dptBC*;

5 (b) the amino acid sequence of DptBC;

(c) the amino acid sequence of SEQ ID NO:

- 208 -

11; and

(d) the amino acid sequence encoded by the nucleic acid sequence SEQ ID NO: 12.

29. An isolated polypeptide that is encoded by the nucleic acid molecule according to any one of claims 1-13.

30. An isolated polypeptide that is encoded by the nucleic acid molecule according to any one of claims 14-21.

31. An antibody that selectively binds to the polypeptide according to any one of claims 28-30.

32. The antibody according to claim 31 that is an intact immunoglobulin; an antigen-binding portion thereof that is Fab, Fab', F(ab')₂, Fv, dAb or a CDR fragment; a single-chain antibody; a chimeric antibody; a
5 diabody; or a polypeptide comprising at least a portion of the immunoglobulin sufficient to confer specific antigen binding to the polypeptide.

33. The antibody according to claim 32, wherein the antibody is a neutralizing antibody.

34. The antibody according to claim 32, wherein the antibody is an activating antibody.

35. The antibody according to claim 32, wherein the antibody is a monoclonal antibody or a polyclonal antibody.

36. A method for preparing an antibody that selectively binds to the polypeptide according to any one of claims 28-30, comprising the steps of

(a) immunizing a non-human animal with the

5. polypeptide; and

(b) isolating the antibody

37. A method for determining if a sample contains a nucleic acid molecule encoding a daptomycin NRPS or a daptomycin NRPS subunit, comprising the steps of

5 (a) providing a nucleic acid molecule according to any one of claims 1-21;

(b) contacting the nucleic acid molecule with the sample under selective hybridization conditions; and

10 (c) determining if the nucleic acid molecule selectively hybridized to a nucleic acid molecule in the sample.

38. A method to produce daptomycin comprising the steps of

(a) introducing a nucleic acid molecule comprising SEQ ID NO: 1 in a host cell; and

5 (b) culturing the host cell under conditions in which daptomycin is produced.

39. The method according to claim 38, wherein the nucleic acid molecule is derived from *Streptomyces*.

40. The method according to claim 38 wherein the nucleic acid molecule is derived from *S. roseosporus*.

41. The method according to claim 40, wherein the nucleic acid molecule comprises the entire daptomycin biosynthetic gene cluster.

- 210 -

42. The method according to claim 38, wherein the nucleic acid molecule encodes a peptide comprising DptA, DptBC or DptD.

43. The method according to claim 42, wherein the nucleic acid molecule encodes a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO: 9, SEQ ID NO: 11 and SEQ ID NO: 7.

44. The method according to claim 38 wherein the nucleic acid molecule comprises a nucleic acid sequence selected from the group consisting of *dptA*, *dptBC* and *dptD*, or wherein the nucleic acid molecule
5 comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO: 10, SEQ ID NO: 12 and SEQ ID NO: 3.

45. The method according to claim 38, wherein
10 the nucleic acid molecule comprises a nucleic acid sequence from an *S. roseosporus* nucleic acid sequence from BAC clone B12:03A05.

46. The method according to claim 38, wherein the nucleic acid molecule selectively hybridizes to a reference nucleic acid molecule that encodes a daptomycin gene cluster or a portion thereof, wherein the reference
5 nucleic acid molecule comprises a nucleic acid sequence selected from the group consisting of:

(a) a nucleic acid sequence selected from the group consisting of *dptA*, *dptBC* and *dptD*;

- 211 -

(b) a nucleic acid sequence encoding the
10 amino acid sequence of a polypeptide selected from the
group consisting of DptA, DptBC and DptD;

(c) a nucleic acid sequence encoding the
amino acid sequence of a polypeptide selected from the
group consisting of SEQ ID NO: 9, SEQ ID NO: 11 and SEQ
15 ID NO: 7;

(d) a nucleic acid sequence selected from
the group consisting of SEQ ID NO: 10, SEQ ID NO: 12 and
SEQ ID NO: 3;

(e) a nucleic acid sequence from an *S.*
20 *roseosporus* nucleic acid sequence from BAC clone
B12:03A05.

47. The method according to claim 38, further
comprising the step of isolating the daptomycin.

48. A method to increase the production of
daptomycin by a cell comprising the steps of

(a) providing a host cell that expresses
daptomycin;

5 (b) introducing a nucleic acid molecule into
a neutral site of a chromosome of said host cell, wherein
the introduction of the nucleic acid molecule results in
increased production of daptomycin by a cell compared to
the cell without the nucleic acid molecule; and

10 (c) culturing the host cell under conditions
in which daptomycin is produced;

- 212 -

wherein the nucleic acid molecule comprises a nucleic acid sequence selected from the group consisting of

15 (i) the nucleic acid sequence of dptBC;

(ii) a nucleic acid sequence encoding the amino acid sequence of DptBC;

20 (iii) a nucleic acid sequence encoding the amino acid sequence SEQ ID NO: 11; and

(iv) the nucleic acid sequence SEQ ID NO: 12.

49. The method according to claim 48, wherein the host cell is *S. roseosporus* or *S. lividans* comprising the daptomycin biosynthetic gene cluster.

50. A nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 24, 54, 74, 94, 98 and 100, or encoding a polypeptide having an amino acid sequence selected from
5 the group consisting of SEQ ID NOS: 93, 97 and 99;

wherein said nucleic acid molecule is not pRHB160, pRHB166, pRHB613, pRHB614, or pRHB680.

51. A polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 93, 97, and 99; or encoded by a nucleic acid molecule selected from the group consisting of SEQ ID
5 NOS: 24, 54, 74, 94 98 and 100.

52. A nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting

of SEQ ID NOS: 105, 107, 109, 111, 113, 115, 117, 119,
121, 123, 125, 127, 129, 131, 133 and 135; or encoding a
5 polypeptide having an amino acid sequence selected from
the group consisting of 104, 108, 110, 112, 114, 116,
118, 120, 122, 124, 126, 128, 130, 132, 134 and 136;
wherein said nucleic acid molecule is not
pRHB174.

10 53. A polypeptide comprising an amino acid
sequence selected from the group consisting of SEQ ID
NOS: 104, 108, 110, 112, 114, 116, 118, 120, 122, 124,
126, 128, 130, 132, 134 and 136; or encoded by a nucleic
acid molecule selected from the group consisting of SEQ
15 ID NOS: 102, 105, 107, 109, 111, 113, 115, 117, 119, 121,
123, 125, 127, 129, 131, 133 and 135.

54. An antibody that binds to the polypeptide
according to either of claims 51 or 53.

55. A method for producing a lipopeptide or
peptide core thereof, comprising the steps of

(a) providing a cell comprising an NRPS gene
cluster other than the daptomycin biosynthetic gene
5 cluster in a cell;

(b) introducing into the cell at least one
gene that encodes all or a part of the NRPS from the
daptomycin biosynthetic gene cluster; and

(c) culturing the cell under conditions in
10 which the lipopeptide or peptide core thereof is
produced.

56. A method for producing a lipopeptide or
peptide core thereof, comprising the steps of

(a) providing a cell comprising a daptomycin
biosynthetic gene cluster in a cell;

5 (b) introducing into the cell at least one other gene that encodes all or a part of an NRPS from an NRPS gene cluster other than the daptomycin biosynthetic gene cluster; and

(c) culturing the cell under conditions in
10 which the lipopeptide or peptide core thereof is produced.

57. The method according to either of claims 55 or 56, wherein said daptomycin biosynthetic gene cluster comprises a naturally-occurring nucleotide sequence of the daptomycin biosynthetic gene cluster.

58. The method according to either of claims 55 or 56, wherein said daptomycin biosynthetic gene cluster comprises a modified nucleotide sequence compared to the naturally-occurring nucleotide sequence of the
5 daptomycin biosynthetic gene cluster.

59. The method according to claim 58, wherein said modified nucleotide sequence comprises a modification or replacement of one or more modules of the daptomycin biosynthetic gene cluster that alters one or
5 more amino acids incorporated by the NRPS encoded by the daptomycin biosynthetic gene cluster compared to the amino acids encoded by the NRPS of the naturally-occurring daptomycin biosynthetic gene cluster.

60. The method according to either of claims 55 or 56, wherein said NRPS gene cluster other than the daptomycin biosynthetic gene cluster comprises a naturally-occurring nucleotide sequence of the NRPS gene
5 cluster.

61. The method according to either of claims 55 or 56, wherein said NRPS gene cluster other than the

daptomycin biosynthetic gene cluster comprises a modified
nucleotide sequence compared to the naturally-occurring
5 nucleotide sequence of the NRPS gene cluster.

62. The method according to claim 61, wherein
said modified nucleotide sequence comprises a
modification or replacement of one or more modules of the
NRPS gene cluster that alters one or more amino acids
5 incorporated by the NRPS gene cluster compared to the
amino acids encoded by the NRPS of the naturally-
occurring NRPS gene cluster.

63. The method according to either of claims
55 or 56, wherein said aptomycin biosynthetic gene
cluster comprises a naturally-occurring nucleotide
sequence of the aptomycin biosynthetic gene cluster and
5 said NRPS gene cluster other than the aptomycin
biosynthetic gene cluster comprise a naturally-occurring
nucleotide sequence of the NRPS gene cluster.

64. The method according to either of claims
55 or 56, wherein said aptomycin biosynthetic gene
cluster comprises a modified nucleotide sequence compared
to the naturally-occurring nucleotide sequence of the
5 aptomycin biosynthetic gene cluster and said NRPS gene
cluster other than the aptomycin biosynthetic gene
cluster comprises a modified nucleotide sequence compared
to the naturally-occurring nucleotide sequence of the
NRPS gene cluster.

65. The method according to claim 64, wherein
said modified nucleotide sequence of said aptomycin
biosynthetic gene cluster and of said NRPS gene cluster
other than the aptomycin biosynthetic gene cluster
5 comprises a modification or replacement of one or more

- 216 -

modules of the daptomycin biosynthetic gene cluster or the NRPS gene cluster that alters one or more amino acids incorporated by the NRPS gene cluster compared to the amino acids encoded by the NRPS of the naturally-
10 occurring gene cluster.

66. The method according to either of claims 55 or 56, further comprising the step of introducing into the cell at least one more gene that encodes all or a part of an NRPS from an NRPS gene cluster other than the
5 daptomycin biosynthetic gene cluster.

67. The method according to either of claims 55 or 56, further comprising the step of introducing into the cell at least one more gene that encodes all or a part of an NRPS from the daptomycin biosynthetic gene
5 cluster.

68. The method according to either of claims 55 or 56, further comprising the step of isolating the lipopeptide or peptide core thereof.

69. The cell produced by the method according to either of claims 55 or 56.

70. The lipopeptide or peptide core thereof produced by the method according to either of claims 55 or 56.

71. A method for producing a lipopeptide or peptide core thereof, comprising the steps of

(a) providing a cell comprising a daptomycin biosynthetic gene cluster in a cell;

5 (b) modifying or exchanging one or more modules of the daptomycin biosynthetic gene cluster to

- 217 -

alter at least one amino acid directed for incorporation by the gene cluster;

10 (c) introducing into the cell at least one other gene that encodes all or a part of an NRPS from the daptomycin biosynthetic gene cluster or another NRPS gene cluster; and

(d) culturing the cell under conditions in which the lipopeptide or peptide core thereof is
15 produced.

72. A method for producing a lipopeptide or peptide core thereof, comprising the steps of

(a) providing a cell comprising an NRPS gene cluster in a cell;

5 (b) introducing a daptomycin biosynthetic gene cluster in a cell;

(c) modifying or exchanging one or more modules of the daptomycin biosynthetic gene cluster to alter at least one amino acid directed for incorporation
10 by the gene cluster; and

(d) culturing the host cell under conditions in which the lipopeptide or peptide core thereof is produced.

73. A method for producing a lipopeptide or peptide core thereof, comprising the steps of

(a) providing a cell comprising an NRPS gene cluster in a cell;

5 (b) providing a daptomycin biosynthetic gene cluster in a cell;

(c) modifying or exchanging one or more modules of the daptomycin biosynthetic gene cluster to alter at least one amino acid directed for incorporation
10 by the gene cluster;

- 218 -

(d) introducing the modified daptomycin
biosynthetic gene cluster into the cell; and

(e) culturing the host cell under conditions
in which the lipopeptide or peptide core thereof is
15 produced.

Figure 1

Manipulation of Dpt genes

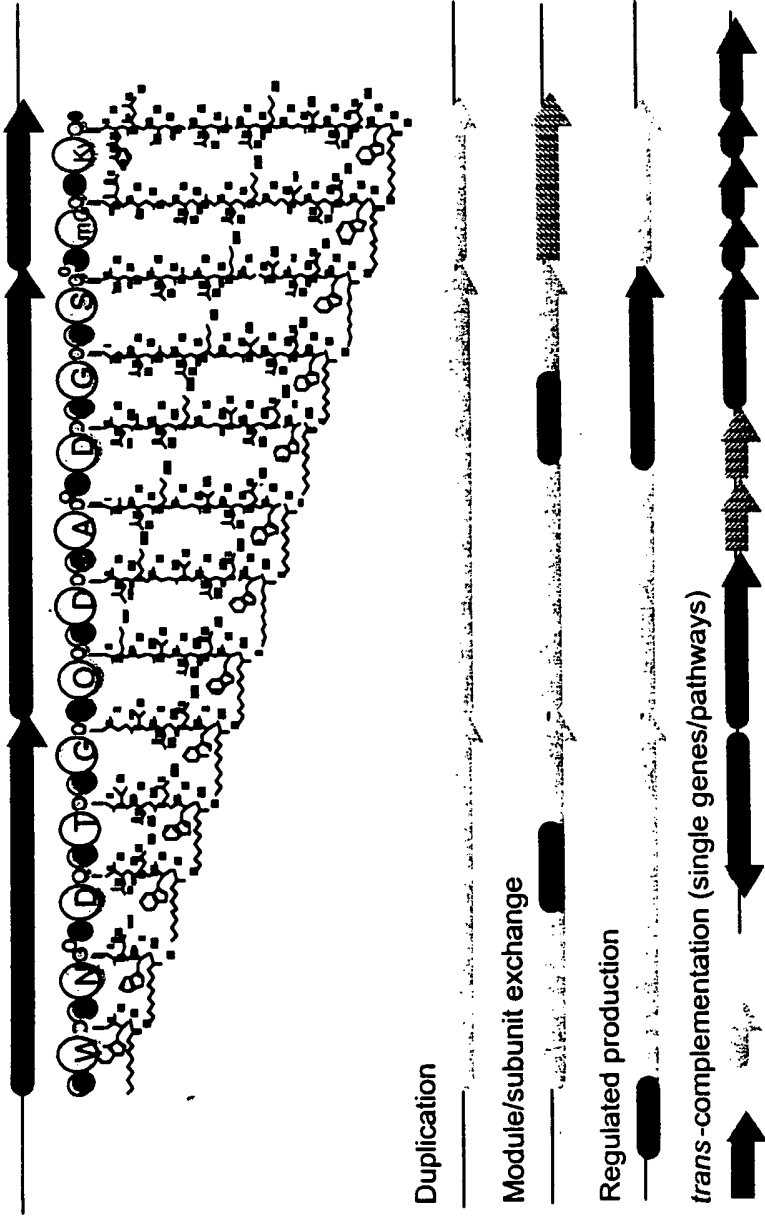


Figure 2A

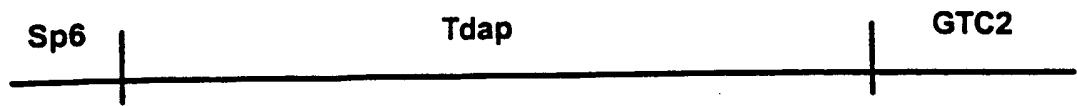


Figure 2B

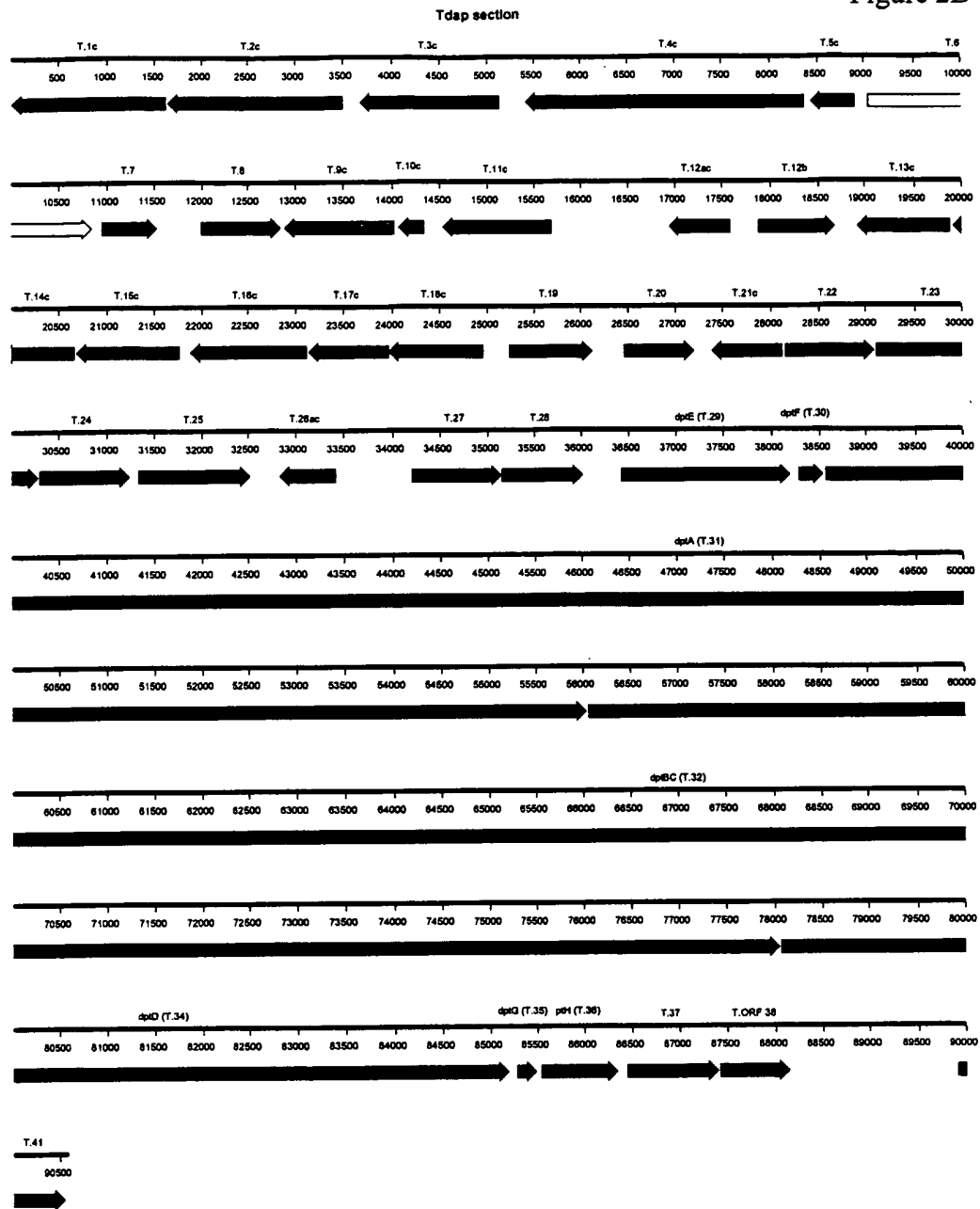


Figure 2C

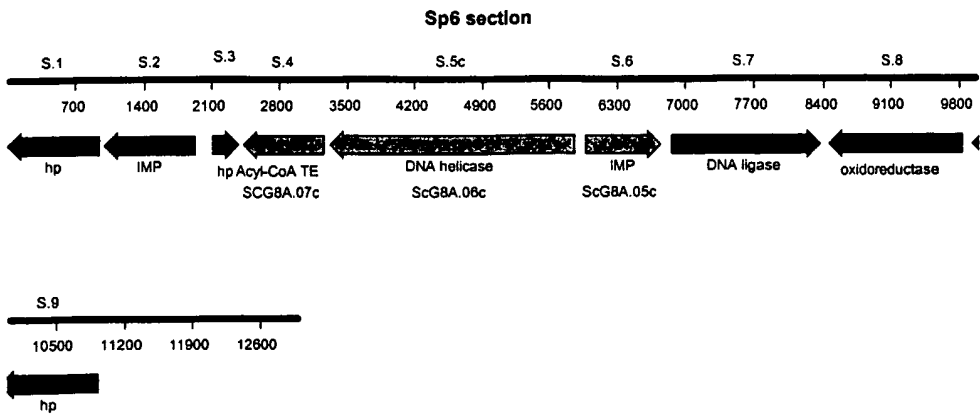
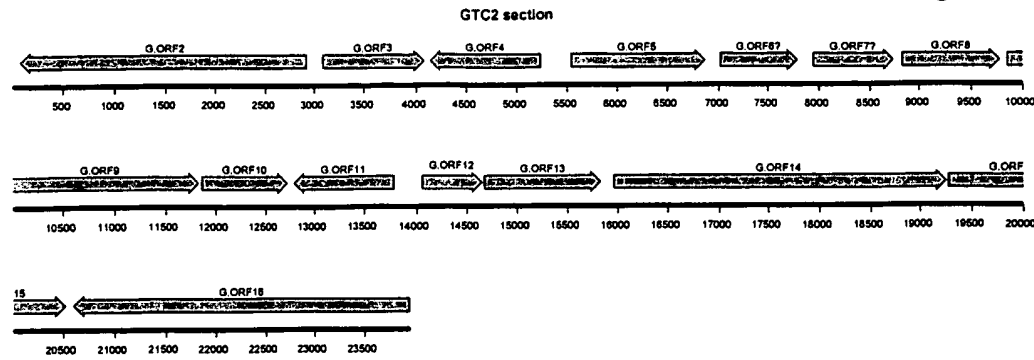


Figure 2D



DpEd.cd11 Multiple Alignments
Tuesday, December 19, 2000 9:25 PM

6/19

FIGURE 3

ClustalW (v1.4) multiple sequence alignment

2 Sequences Aligned Alignment Score = 7705
Gaps Inserted = 22 Conserved Identities = 1223

Pairwise Alignment Mode: Slow
Pairwise Alignment Parameters:
Open Gap Penalty = 10.0 Extend Gap Penalty = 0.1
Similarity Matrix: blosum

Multiple Alignment Parameters:
Open Gap Penalty = 10.0 Extend Gap Penalty = 0.1
Delay Divergent = 404 Gap Distance = 8
Similarity Matrix: blosum

Processing time: 18.0 seconds

```

DpEd      1 MTQAMKEDILPLTFLGECILPHSVYDEOSVYVTVQVWVLEGGFVDFPEALAAAAAALLRHANLAAAFRYELQRFVQII PREVAVPMQHTDVALEGAQKAKTERLLAQDQRRFRLT 120
T36180    1 MUKSLEDILPLTFLGECILPHSVYDEOSVYVTVQVWVLEGGFVDFPEALAAAAAALLRHANLAAAFRYELQRFVQII PREVAVPMQHTDVALEGAQKAKTERLLAQDQRRFRLT 120
      .. .....
DpEd      121 APFLAFLVLTGKHDRHAPALTFFKILKDDKSPVLLAEITLYRTG- DETALPMVPRYDYLMTSRDRDEAGHAKALAGVDEATLVAFGADRAEAPPLATEERLEPDLAATLAAR 239
T36180    121 APFLAFLVLTGKHDRHAPALTFFKILKDDKSPVLLAEITLYRTG- DETALPMVPRYDYLMTSRDRDEAGHAKALAGVDEATLVAFGADRAEAPPLATEERLEPDLAATLAAR 240
      .. .....
DpEd      240 ARSGVTVTLTVQAAALVGLKLTGDDVTVGVTVSGRPFELAGVDMKGLFINTVLAELLPRSLAFTVLRQEQIQLLDKQYERLAVIQLAGRTLDTVMVFTNTVAASS- 358
T36180    241 ARSGVTVTVTVQCCVGLVSLKLTGDDVTVGVTVSGRPFELAGVDMKGLFINTVLAELLPRSLAFTVLRQEQIQLLDKQYERLAVIQLAGRTLDTVMVFTNTVAASS- 360
      .. .....
DpEd      359 -AGADGPAAEPRVADPHVDAHMYFLGLVLPGLPPLAAGHPSALPAERVTTIDSLVRLAEHMDPOLAVGRADILGEEQKMLTGLNDTHRDVPLTVPMQTEQAARTKGRPA 477
T36180    361 PFGAAPDAELPTVLGVRKSDOHYFLGLLALPRTELAFELGYLPVVDPAVSAVIAAFRAALRTVLDAPOTRGAVALLDPVAGTVLEDSGSDGVRPAEDUTLFEQVARTPGKTA 480
      .. .....
DpEd      478 VHARDGELSYAEHARANLARKLAAGVGPETVYTLPLSARMVVALAVHKTGAAYVVDPEYPAERLAPLGLDGPALVLEDS---RSAAMPAGPAPVLTLDGALDTGVRLPE 594
T36180    481 LIAFDGRLTYAELEDAANRLARLVELGVGPENWVAVAGRTELAVGELAVLACAGAYVVDPEYTPORLUMHQDADPALVLTSDVDRIGEEQKMLTGLNDTHRDVPLTVPMQTEQAARTKGRPA 600
      .. .....
DpEd      595 HQLTGTGLAPLD-QPAYVTVTSGSTGRPHGVVILHSVTVGLRT- IEEYPAAGQVYHSFVSFLTVGALYAPLVSGGLALGSPDOKILGGE---DSPTVQATPSRLAVLDEL 709
T36180    601 ALTDADRAAPLPGHAPAYVTVTSGSTGRPHGVVILHSVTVGLRT- IEEYPAAGQVYHSFVSFLTVGALYAPLVSGGLALGSPDOKILGGE---DSPTVQATPSRLAVLDEL 720
      .. .....
DpEd      710 PDEISPTGALTLOGEELSETLDMRANHGVTVVHYGPTETTDICADRLAPOTLLPGVPTIGRPLMTRLYVLDGELVYVTVGVAGELVYAGAGLARGYLRPGLTABRFPVACPG 829
T36180    721 PPEVAPSTLTLOGEELAGELARHREAGGVTVVHYGPTETATGCLDMLAORTVEGPTVIGTGRGVRVTVLDSALAIPVAGLGGVTVLAGVGLARGYLRPGLTABRFPVACPG 840
      .. .....
DpEd      830 APGRVTVTGLVHRTDGLTLEFVGRVDDQVYKRGFTELGSEVATVAATPGVAAATVAVREDRPGDRLVAYVTPADVDPTGELPSAVTAHAAALPAYHVSAYVVLHEVPLTNGKI 949
T36180    841 APGRVTVTGLVHRTDGLTLEFVGRVDDQVYKRGFTELGSEVATVAATPGVAAATVAVREDRPGDRLVAYVTPADVDPTGELPSAVTAHAAALPAYHVSAYVVLHEVPLTNGKI 957
      .. .....
DpEd      950 NHAALPAPEAVSGAGFPAAGTANREVLGLFAEVLGERVGTADOFFELGESHLLATRVSRVSLVGLVGLVRLFDAPTGRILLGERSGAJPAAPLTAERTGRDPLSYAQQRLM 1069
T36180    958 DUALPAPTPTGTAGRAPPTPAEELGLYAEVLSLPGVTVDDOFFELGESHLLATRVSRVSLVGLVGLVRLFDAPTGRILLGERSGAJPAAPLTAERTGRDPLSYAQQRLM 1075
      .. .....
DpEd      1070 FLHEEGKATYHPLAARLTGFLVLTALDALTDAVANKESLTLIARDGTOTANGHILPGDPRATITLAEVPLHDELACRLAEAAHHPFILTAEIPVATVTRTERDHTLLVYTH 1189
T36180    1076 FLHEEGSPPTYHPLVLALEGLPLPDALRAALLDVGWHSUTFTTEDRG-ARGVHHPAGG---VRPVFTAESTEADYADLARAARHATLGAELIPVRAALLALSTREHVLILVH 1192
      .. .....
DpEd      1190 HIASDMSKEPFLADLSAAYANRAHAEPLPLSVQYADAMQNDVLTGDEDTSDHQQALHMRGLAGLPGELPTPRPRPQVGRGGRGLETPAALHRTDITLARTSTTVF 1309
T36180    1193 HIASDMSKFLAQLDTAATACAGAPAPQPLPVQYADALMGGELGDDTPOTLAGRLAYHQQGLAGLPERLGLPTPRPRPQVGRGGRGLETPAALHRTDITLARTSTTVF 1312
      .. .....
DpEd      1310 MYQALACELSLGAGTDIPITPTAGRTDEATELGITFVNTVLATDVGDPFAELARVATLDAVANDVTFEALVEVLANPESLHAPLQILLAPONTEDRSISDRPOTLL 1429
T36180    1313 MYQALACELSLGAGTDIPITPTAGRTDEATELGITFVNTVLATDVGDPFAELARVATLDAVANDVTFEALVEVLANPESLHAPLQILLAPONTEDRSISDRPOTLL 1432
      .. .....
DpEd      1430 PDLQVTEOP---LDAGTAPFLAFTERPFERGEPSTGIVSHALYDESTVQIADCTVQFLDAAVHAPOTRYDAVGLLPDHTLHLLTRESQTVGLSPATLPELFYARVAHHPH 1547
T36180    1433 PDLQVTEOP---LDAGTAPFLAFTERPFERGEPSTGIVSHALYDESTVQIADCTVQFLDAAVHAPOTRYDAVGLLPDHTLHLLTRESQTVGLSPATLPELFYARVAHHPH 1552
      .. .....
DpEd      1548 IAVEVAGRAPSTTYDALRRANRLARLITDGRVPRORVAILPREADLYTAMGLJAGAVCVVDPAYTPORLUMHQDADPALVLTSDVDRIGEEQKMLTGLNDTHRDVPLTVPMQTEQAARTKGRPA 1665
T36180    1553 IAVEVAGRAPSTTYDALRRANRLARLITDGRVPRORVAILPREADLYTAMGLJAGAVCVVDPAYTPORLUMHQDADPALVLTSDVDRIGEEQKMLTGLNDTHRDVPLTVPMQTEQAARTKGRPA 1669
      .. .....
DpEd      1666 AAPDGNPTOTGLPANPAYVTVTSGSTGRPHGVVILHSVTVGLRT- IEEYPAAGQVYHSFVSFLTVGALYAPLVSGGLALGSPDOKILGGE---DSPTVQATPSRLAVLDEL 1785
T36180    1670 ADLTADRLAPLPGHAPAYVTVTSGSTGRPHGVVILHSVTVGLRT- IEEYPAAGQVYHSFVSFLTVGALYAPLVSGGLALGSPDOKILGGE---DSPTVQATPSRLAVLDEL 1789
      .. .....
DpEd      1786 AGAAPPLTVTVTGVGCGPLVDRGQGERILNGTPTFTVGTATVACERTGGAUPPIGAPMPOQVTVLDRALPVYAGVGEITYVAGAGLARGYLRPGLTABRFPVACPG 1905
T36180    1790 AGAAPPLTVTVTGVGCGPLVDRGQGERILNGTPTFTVGTATVACERTGGAUPPIGAPMPOQVTVLDRALPVYAGVGEITYVAGAGLARGYLRPGLTABRFPVACPG 1906
      .. .....
DpEd      1906 RMYRTGLARRASDGLLFEGRADTVKINGRVLEAETDALASHPGVEDAVVTVVDDGLDGLRVAYVYTOP-OTPSAAALRLASLPRHNVGGLVLTLDALPLTANGRVORTALP 2024
T36180    1907 RMYRTGLARRASDGLLFEGRADTVKINGRVLEAETDALASHPGVEDAVVTVVDDGLDGLRVAYVYTOP-OTPSAAALRLASLPRHNVGGLVLTLDALPLTANGRVORTALP 2026
      .. .....
DpEd      2025 GPTOTAPAGRAPOTPOEVRLCALADVLRGTVGVDEGTLOGLHSLATLAAVRAALGVLSVTVLSEAPTALLASACTADAAAYDPFETVPLARTGSPPLPCVHAGLSD 2144
T36180    2027 APSRTVE-GSPRTPREALCALADVLRGTVGVDEGTLOGLHSLATLAAVRAALGVLSVTVLSEAPTALLASACTADAAAYDPFETVPLARTGSPPLPCVHAGLSD 2144
      .. .....
DpEd      2145 YAGLSHLDADVPYGLQARLTAAGLPGSVSEDAEDYAGIDRLCTGCPVRLGASGPTVAHATLQOQHTVELLAVLDAYTPGARPAEDVDSQIVADYLAQGSFAPVRL 2264
T36180    2145 YAGLSHLDADVPYGLQARLTAAGLPGSVSEDAEDYAGIDRLCTGCPVRLGASGPTVAHATLQOQHTVELLAVLDAYTPGARPAEDVDSQIVADYLAQGSFAPVRL 2263
      .. .....
DpEd      2265 EG-DAMLPFLFVTRTGGPARDFDAGRILANDVTLNARLTRAFTPGVTVGDMVFAAARPG-GEQAARVGLHHPHVTGGLDHLITCAHEDTDP-AALTRIGPVLAARLCACTN 2380
T36180    2264 TCGRPPTAATRAFLARAGDPGRGLDEALAAVLYTHNARLHRTKGTTCGVNVLVAERAGDGLARAGAESRPHVGRTERVGVADHGLGVQSDAALAVTGRALAGLDPATCH 2381
      .. .....

```

DptD.CdaII Multiple Alignments
Tuesday, December 19, 2000 9:35 AM

DptD	2381	2380
TJ6180	2384 AAGAAVPEPTEVAMNPSPEPAPSPESLSTEVAN	2418

FIGURE 4

DptH/Cda prot aligna :iple Alignments
 Tuesday, December 19, 3:37 AM

ClustalW (v1.4) multiple sequence alignment

2 Sequences Aligned Alignment Score = 955
 Gaps Inserted = 1 Conserved Identities = 145

Pairwise Alignment Mode: Slow
 Pairwise Alignment Parameters:
 Open Gap Penalty = 10.0 Extend Gap Penalty = 0.1
 Similarity Matrix: blosum

Multiple Alignment Parameters:
 Open Gap Penalty = 10.0 Extend Gap Penalty = 0.1
 Delay Divergent = 40% Gap Distance = 8
 Similarity Matrix: blosum

Processing time: 0.4 seconds

```

      10      20      30      40      50      60
DptH protein MRATSRMIQVNGARIACSDSG-----CGDPVLMIAGTGSTGRVWDAYQVPDLHAAGFRT
T36181      MPVLTVNGIRINYDDAPPAGAQNAPAVLLVMGSGGSGRAWHLHQVPALVAAGFRV
      .. *** **      *      *..*.*.*.* *      *** * *****

      70      80      90      100     110     120
DptH protein ITFTNRGVPPSDECERGFLLADLAADTAALIEQVAGGPCRNVGTSLGACVAGQVALARP
T36181      ISFDNRGLAPSEECPCGFGIDDLVADTAALVEELRLGPCRNVAGISMGHIAQELALSRPD
      *. * ** . **.* ** . * * *****.*.. ***** * *..*.....*.*.

      130     140     150     160     170     180
DptH protein LVTQAVFMATRGRTDAMRAAATRAAAALYDSGVLPAYAAAVRALQNLSPHTLRDRHQV
T36181      LVDRLVLMATRARPDALREALCRAEMELYDQGI RLPAAYEAVVQAMONLSFPTLNDVQA
      ** . * **** * **.* * **      *** * . ** * * . * . ***** **

      190     200     210     220     230     240
DptH protein EDWLPLEFYAERDGPVRAQLELGLLPDRLADYRDITVPCLVIAFEDDVVTPPYLGREVA
T36181      RDWLDVLELTRSGAGYRAQLGVRVDGDRREAYRGIRAAATRVVAFQDDLIAPPHLGREVA
      *** . * . * * * * * * . * * * * . *..*.....* * *****

      250     260     270
DptH protein DAIPGARFETVPRCGHYGYLEDASAVNKILRDFFRTSN
T36181      DAIPGAEYELVPDCGHYGYLESPPDAVNKSLVEFLRRN
      ***** . * * ***** * * * * *
  
```

Figure 5A

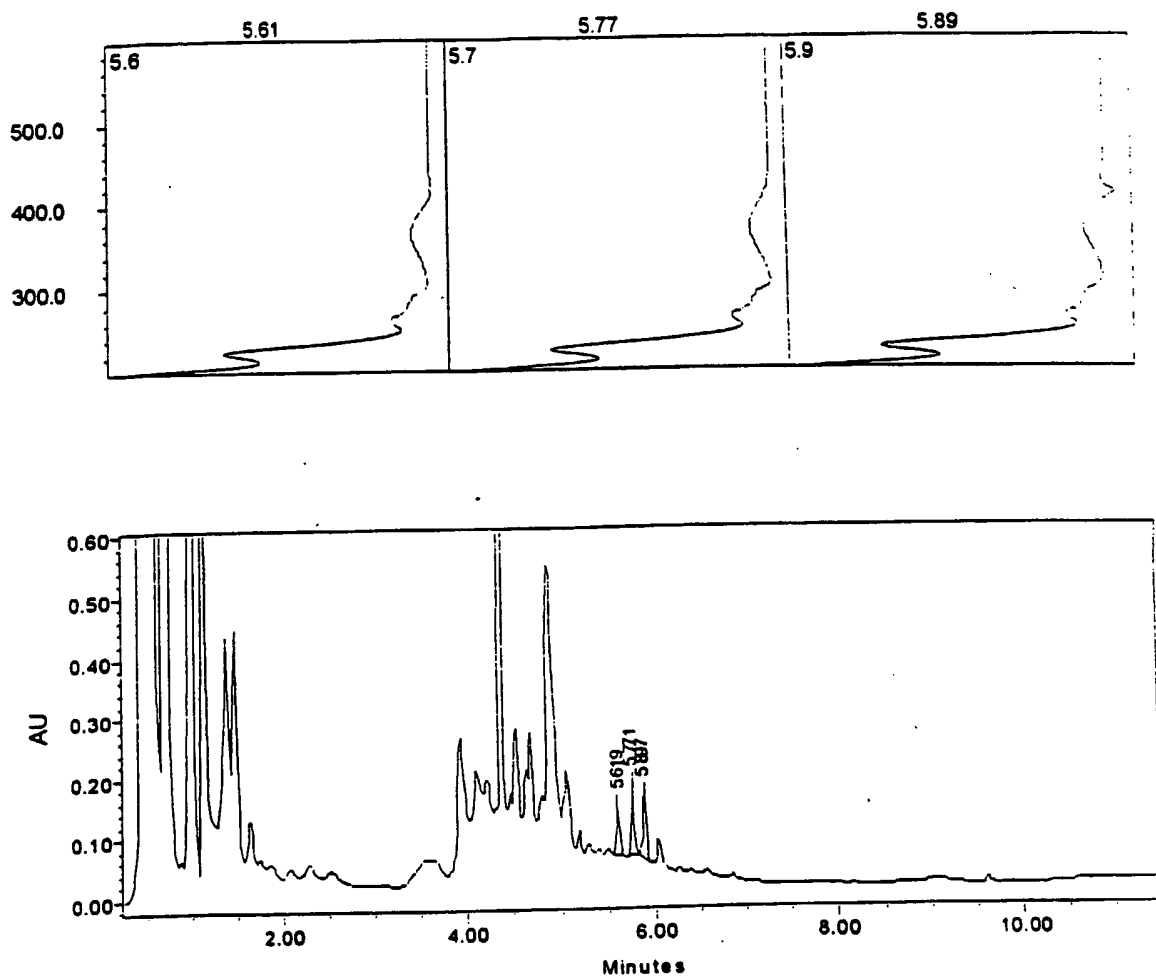


Figure 5B

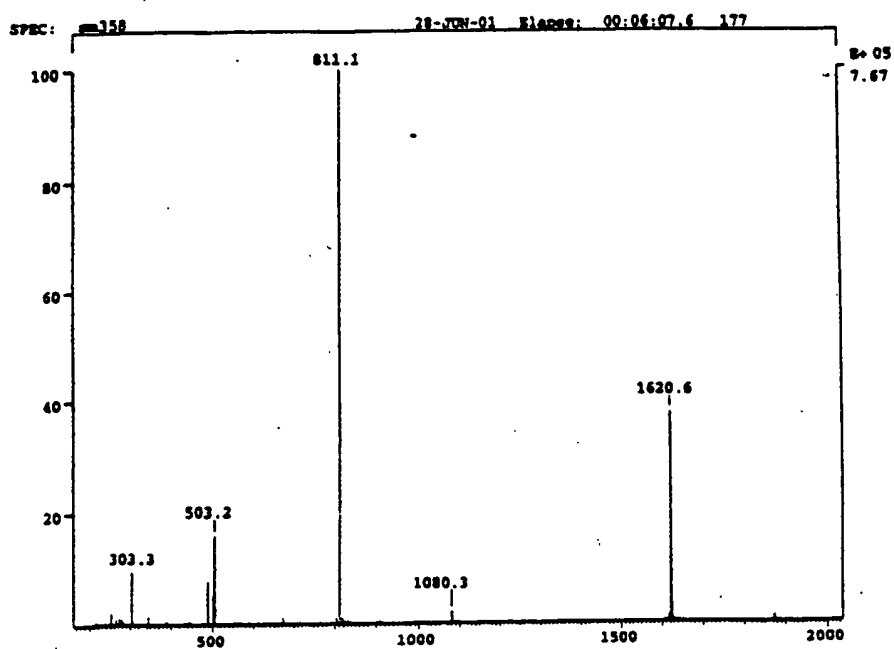


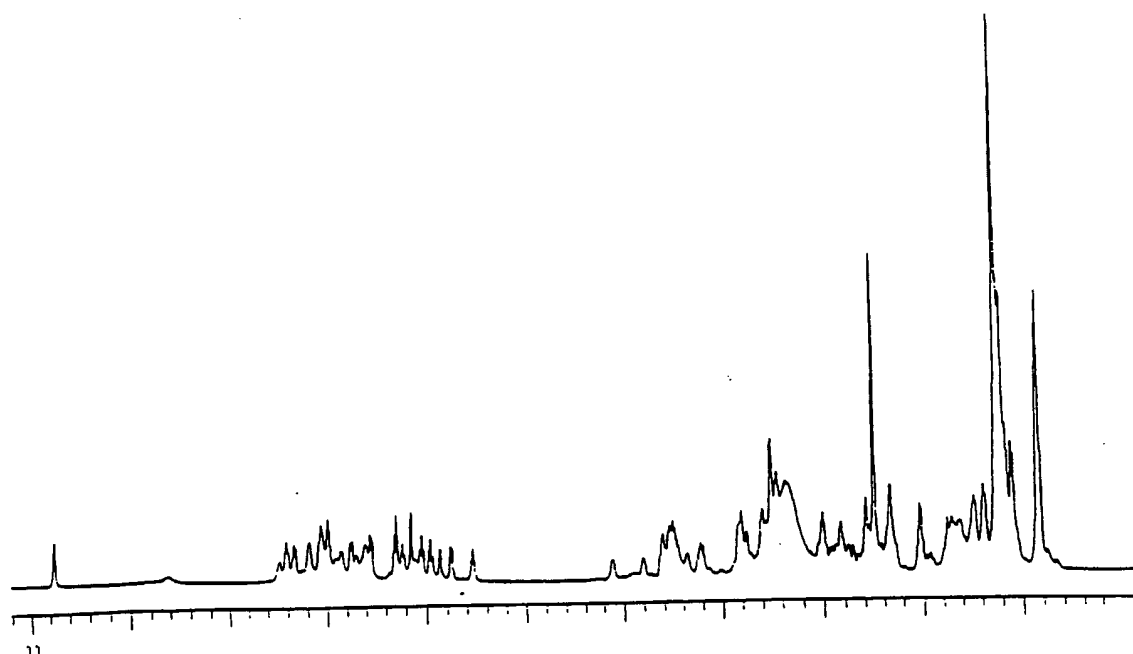
Figure 5C

Figure 6

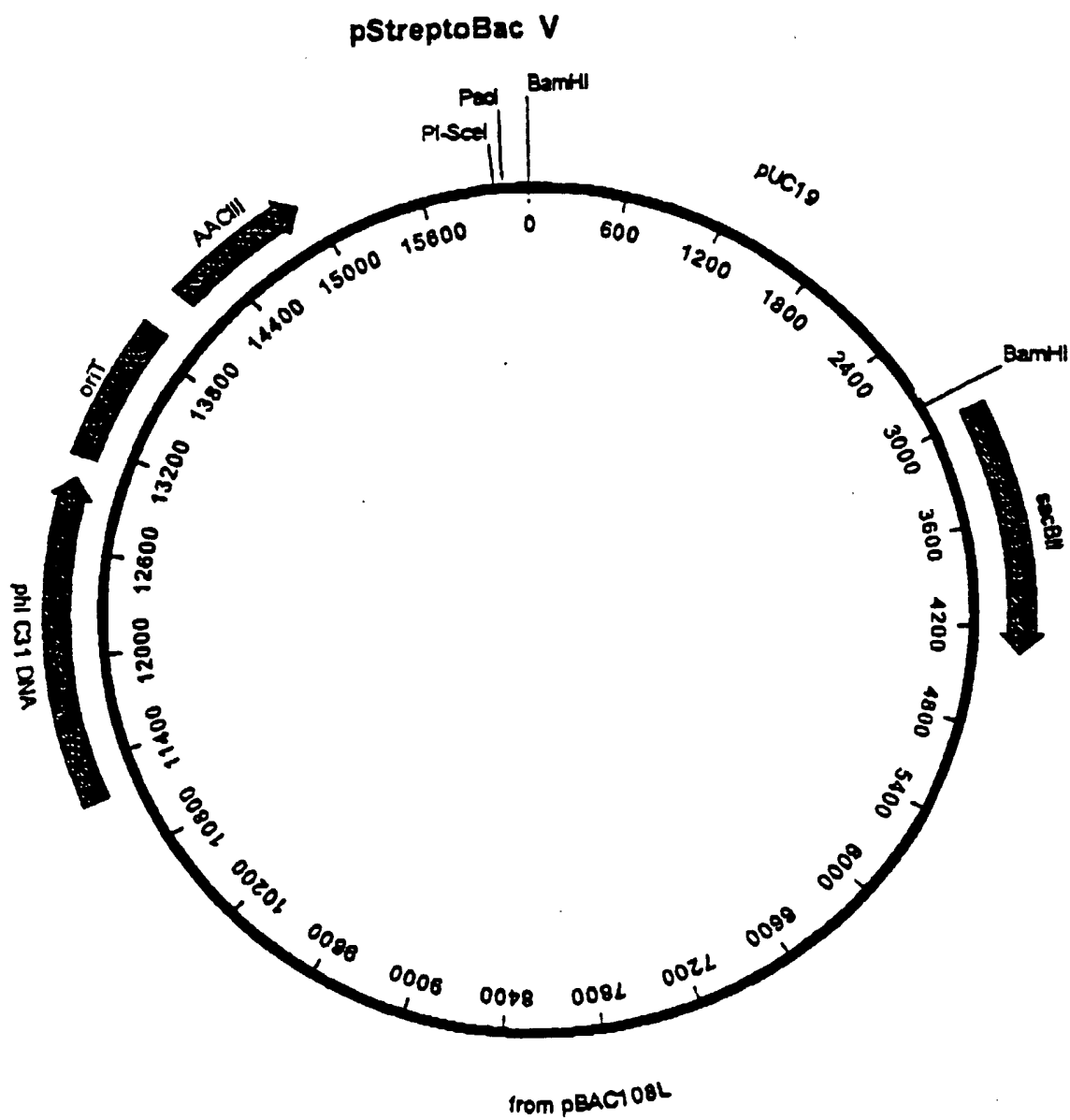


Figure 7

Recovery of *dpt*-related clones

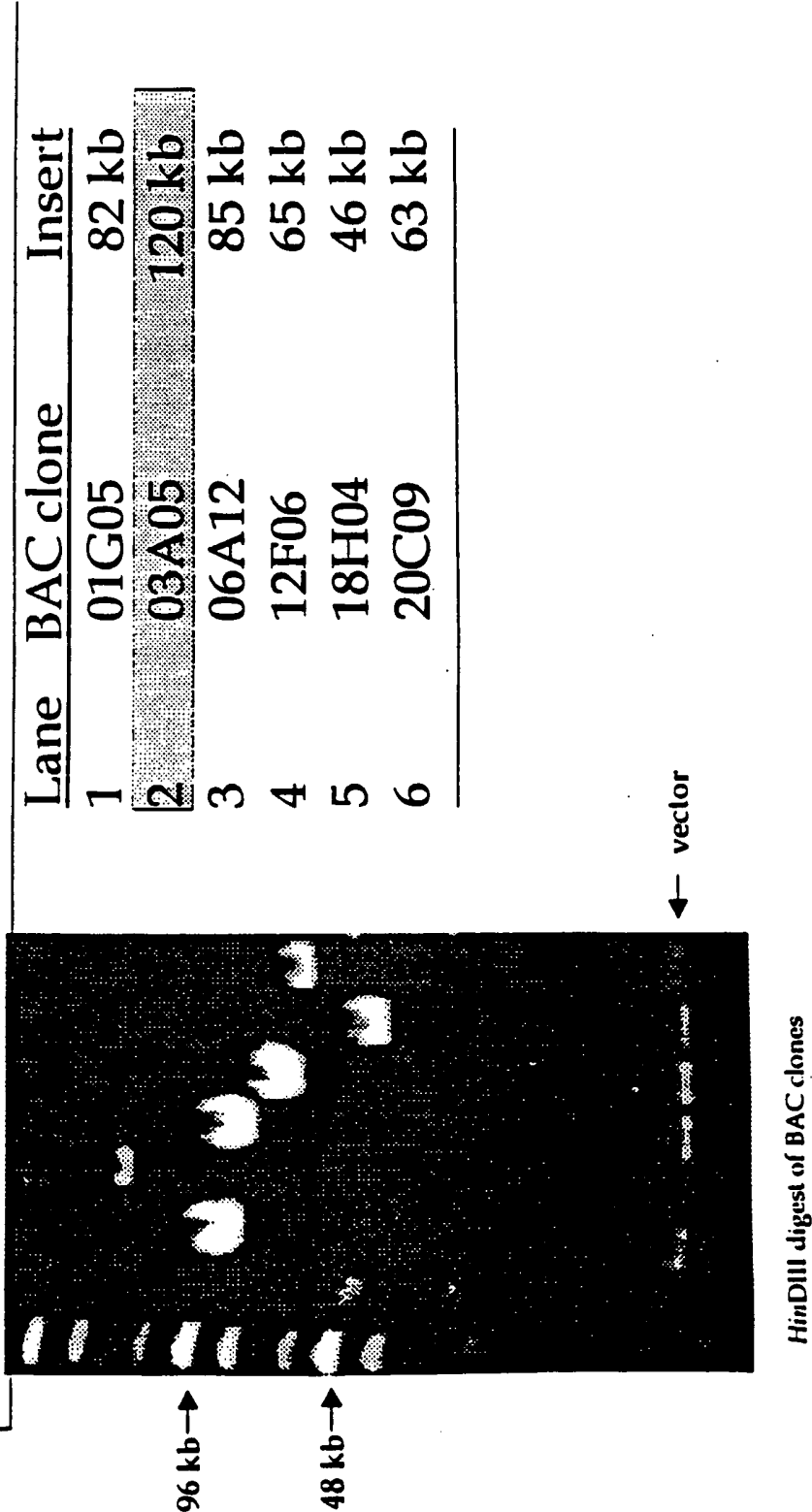


Figure 8

BACs cover 180-200 kb in *dpt* region

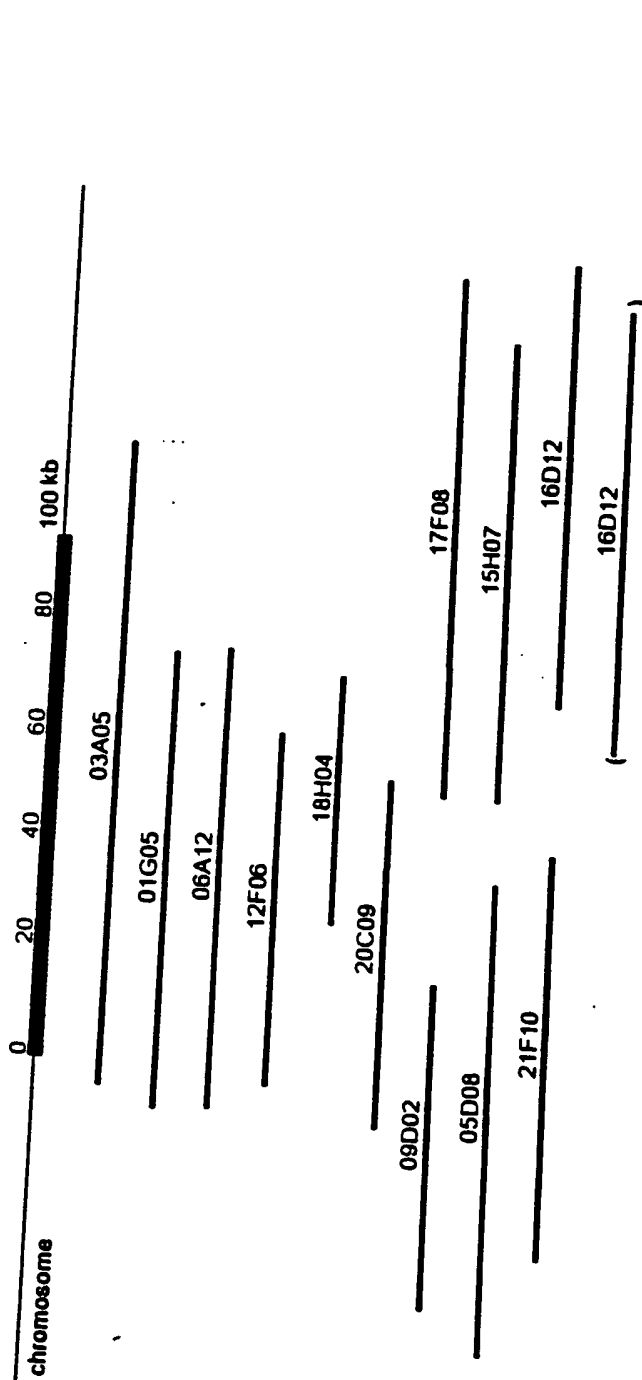
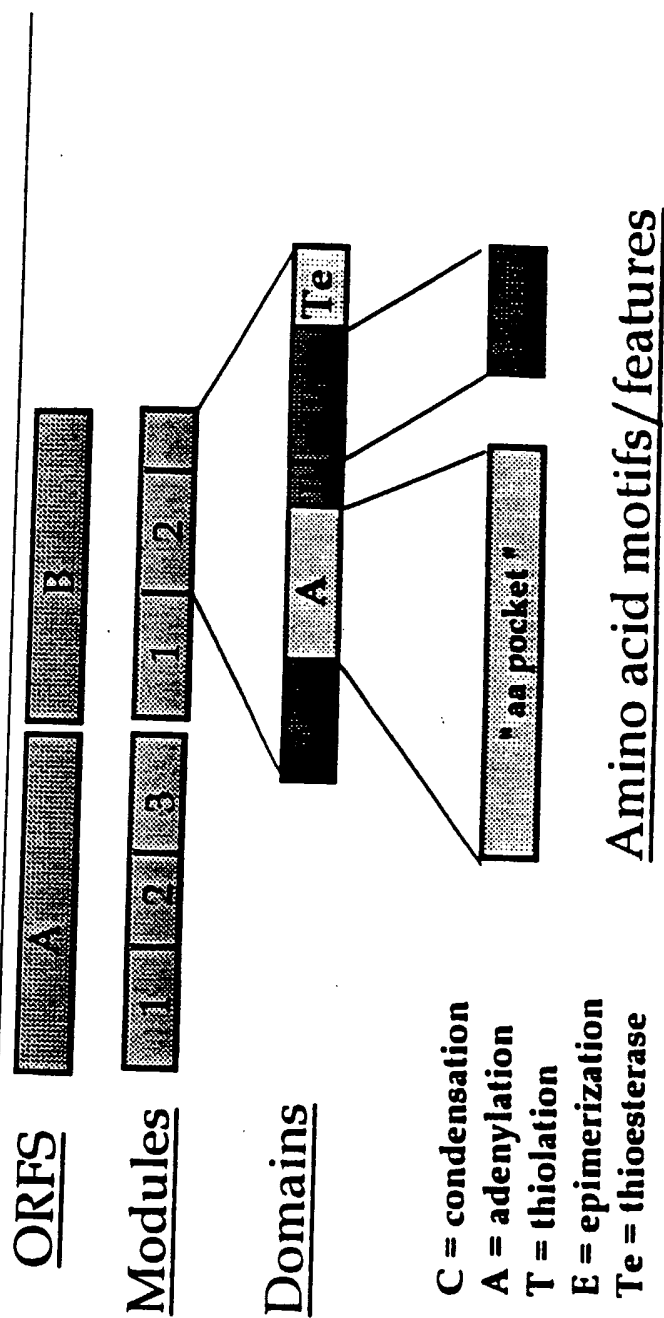


Figure 9

NRPS gene structure



A domain similarities (asn, asp)

Figure 10

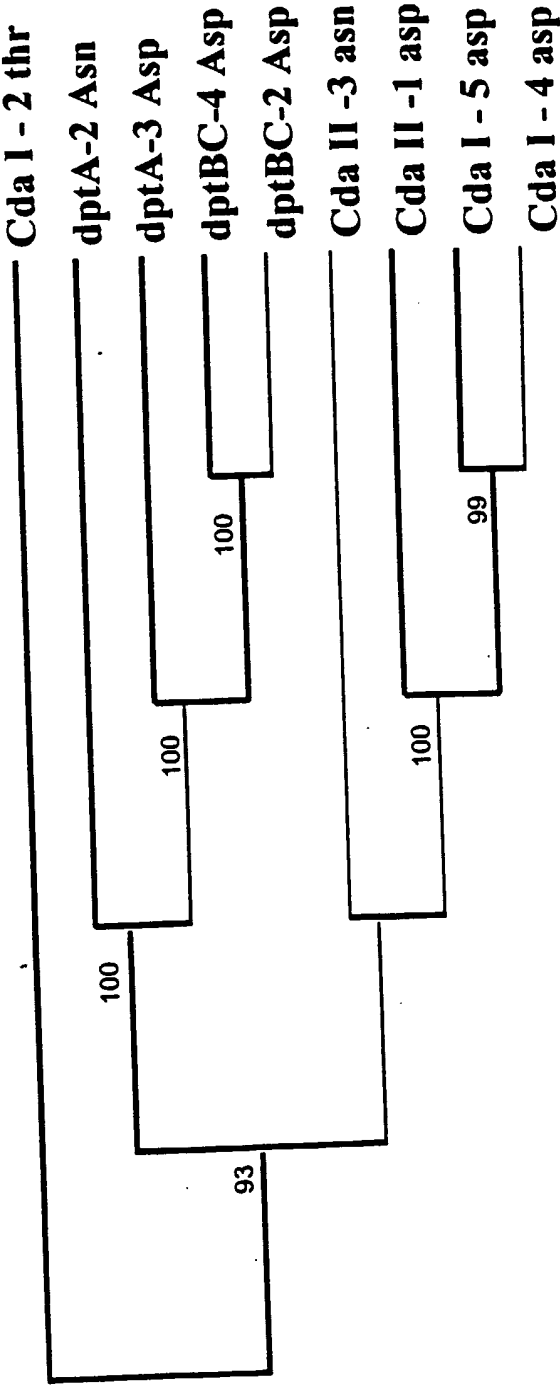


Figure 11

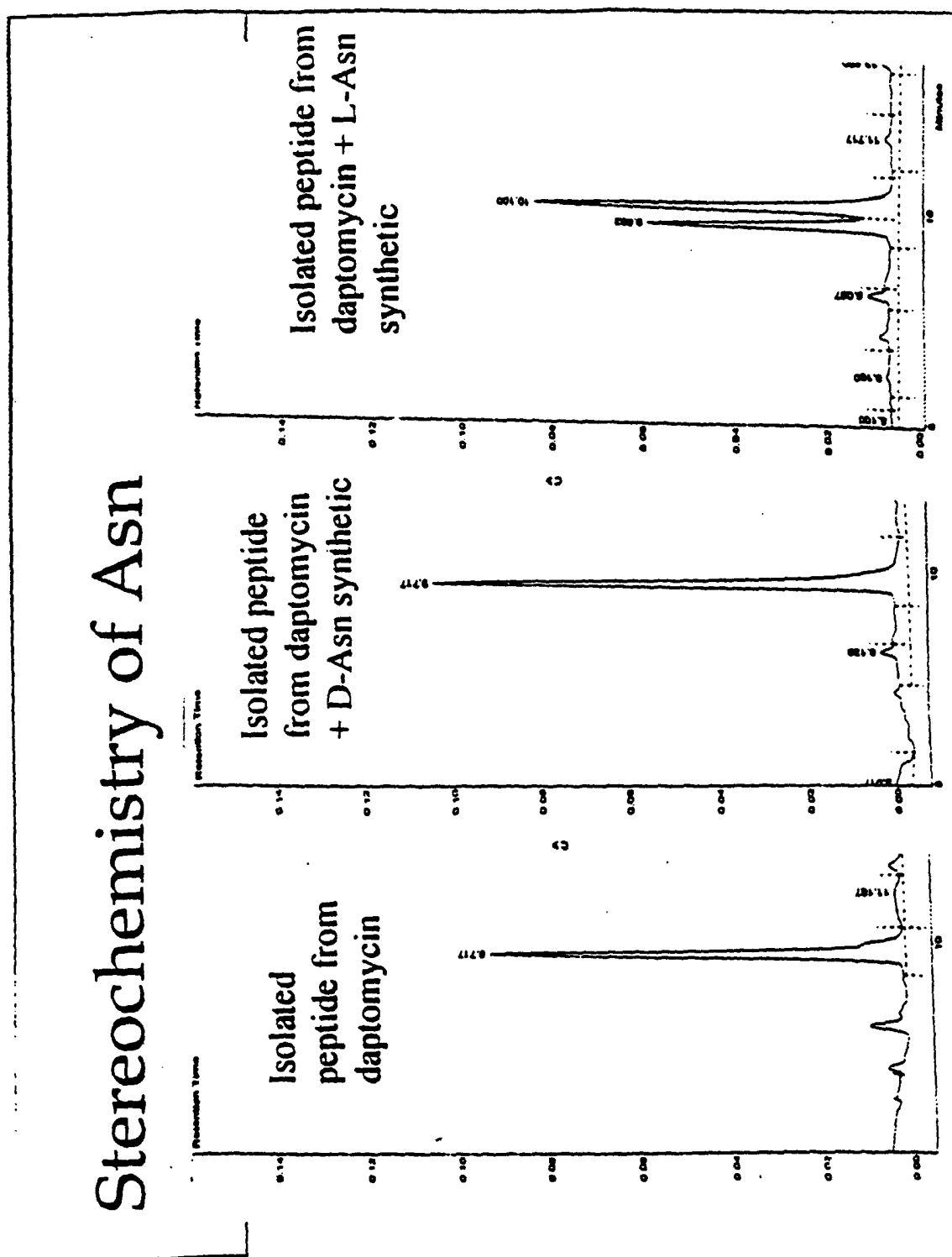


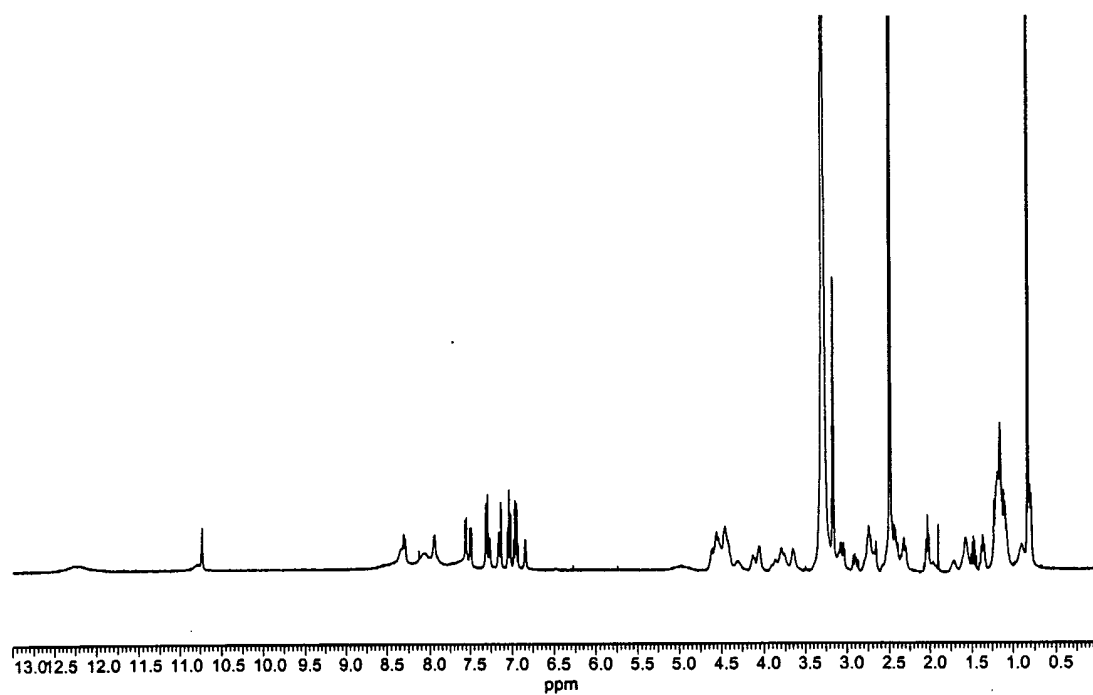
Figure 12

Organization of 46.6 kb *dpt* region



<u>ORF</u>	<u>modules</u>	<u>domains</u>
<i>dptA</i>	5	CAT•CATE•CAT•CAT•CAT
<i>dptBC</i>	6	CAT•CAT•CATE•CAT•CAT•CATE
<i>dptD</i>	2	CAT•CATTe

3 ORFs translationally coupled
Epimerisation domains associated with asn, ala and ser modules
Thioesterase (Te) domain at end of *dptD*

Figure 13

SEQUENCE LISTING

SEQ ID NO: 1

```
1 GCCACCACCG TACGGCCCTC CAGCACCCGG GCCAGGGAAC GCTCCAGATG ACGGGCGGCC
61 CGCGGGTCCA GCAGCGACGT CGCCTCGTCC AGCACCAGCG TGTGCGGATC GGCCAGCACC
121 AGCCGGGCCA GCGCGATCTG CTGCGCCTGG GCCGGGGTCA GCGTGAACCC GCCCGAACCG
181 ACCTCGGTGT CCAGCCCCTT CTCCAGCGCC TTCGCCCAGC CGTCCGCGTC GACCGCGGCC
241 AGCGACGCCC ACAGCTCGGC GTCTTCCTGG CCTTCCCTGG CCAGGCGCAG ATTGTCCCGG
301 AGCGAACCGA CGAAGACATG GTGCTCCTGG TTGACCAGGG CCACATGCTC ACGGACCCGC
361 TCCGCCGTCA TCCGCGACAA CTCCGCCCCG CCGAGCGTCA CCTACCGGT GCGCGGTGCG
421 TAGATCCCCG CCAGCAGCCG GCCCAGCGTC GACTTGCCCG CGCCGACCG GCCGACCAGG
481 GCGAGCCGGG TGCCCGGAGC CACGTCGAGC GACACCTTGT GCAGGACGTC GACACCTTCC
541 CGGTACCCGA AGCGGACCTC GTCCGCCCGT ACGTCCCGGC CTCCGGGCC GACCTCGGCG
601 TCGCCCGCGT CCGGCTCGAT GTCCCGGACG CCGACCAGCC GGGCCAGCGA CACCTGGGCC
661 ACCTGGAGCT CGTCGTACCA GCGCAGGATC AGACCGATCG GGTGACCAT CATCTGGGCC
721 AGCAACGCCC CCGTCGTGAG CTGCCCAGAC GTCAGCCACC CCTCCAGCAC GAACCAGCCG
781 CCGAGCAGCA GGACCGCGCC GAGGATCGTC ACGTACGTGG CGTTGATGAC GGGGAAGAGC
841 ACCGAGCGGA GGAAGAGTGT GTACCGTTCC CACGCTGTCC ATTGAGAAAT CCGCCGGTCC
901 GACAGCGCCA CCCGGCGGCC GCCGAGGCGG TGCGCCTCCA CGGTCCGCC CGCGTCCACG
961 GTCTCCGCGA GCATCGCGGC GACGGCGGCG TAACCGGCGG CCTCCGAGCG GTACGCGGAG
1021 GGGGCCCGGC GGAAGTACCA GCGGCAGCCC ACGATCAGCA CCGGCAGCGG GATCAGCAGC
1081 GCCAGCGCCA GCGGGGGAGC GGTACCGTCC AGCGCGCCGA GCAGCAGCCG GGCCACACAG
1141 ACGCCGATCG CCAGCTGCGG CACGGCCTCG GCGATCGCGT TCGCCAGCGG GTCGATGTCC
1201 GTGGTGATCC GGGACAGCAG ATCGCCCGTC CCGGCCCGCT CCAGCACACC GGGCGGCAGC
1261 CCGAGCGACC GGACGAGGAA GTCCTCGCGC AGATCCGCGA GCATCTCCTC GCCCAGCATC
1321 GCGCCGCGCA GCCGCATGGA GCGGGTGAAC AGGACCTGGA CGACCAGCGC CACCGCGAAG
1381 ATCGCGGCCG TACGCTCCAG ATGCAGGTCG GTGACCCCGG CCGAGAGGTC CTCGACCAGA
1441 CCGCCAGCA GATACGGTCC GGTGATCGAG GCGACCACCG CCACCGCGTT GACCGCGATC
1501 AGGACGGTGA ACGCCCTGCG GTGCCGACGC AGCAGACTCC GTACGTAAC CCGCACGGTC
1561 GTCGGTGTGC CCACGGGCAG TGTCGTGCGC GACTCCGGGG CCGCGGGGTC GTACGCCGGG
1621 GGTGCGACGC CGATCATGCC CTCTCCTCGA TTTCTCTGAT GCTCTTCATG GCGGGGACGT
1681 CGCCGCTCTT CATGACGGAG ACGTCGTAC CAGACGCCGT CACCGCGTCC GCCGCGGCCG
1741 CCTCGTCGTC GGTCTCGCGG GTGACGACCG CCCGGTAGCG CGGTTCGTTG CGCAGCAGGT
1801 CGTGGTGGGT TCCCACGGCG ACGACCGTGC CCTCGTGGAC GAGGACCACC CGGTCCGGCG
1861 CGTCGAGCAG CAGCGGCGAC GAGGCGAACG CCACCGTCGT ACGACCTGG CGCAGCTTCG
1921 CGATGCCGGC GGCAGCCCGT GCCTCGGTGT GCGAGTCGAC CGCGAGGTC GGTCTGTCCA
1981 GCACCAGCGC CTCCGGGTCG GTGACCAAGG ACCGGGCCAG CGCCAGACGC TGGCGTGGC
2041 CGCCGGACAG GGACCGCCG CGCTCGGTGA TCCGGGTCCG CATCGGGTCC CCGTCGTTGT
2101 CGACGACCGC CTGGGCCAGA GCGCTCAGCA CATCGGCGCA CTGGGCCGCC TCCAGCGCCG
2161 TGTCCGGGGT GACCAGGCCC GAGGACGGGA CGTCCAGCAG CTCCTGGAGC GTGCCGGACA
2221 GCAGCACCGG GTCCTTGTC TGGACCAGGA CCGCCGCTCG TGCGGCGTCC AGCGGGATCT
2281 CGTCCAGGGC GACCCCGCCG AGCAGCACCG ACGGGGTCGC CGCGGCGGCC TTGTCTCTCT
2341 CCTCGCCGGT CTCCGCGTGC CCGCCGAGCC GTTCGGCCAG CCGGCCGCC TCCTCCGGGT
2401 CACCGCAGAC GACGGCCGTG AACTGCCCGC GCGGAGCCAT CAGCCCGGTC GCCGGGTCTG
2461 ACAGATCACC GGTGGGCGTC ACACCCTCCA CCGTGGCCTC CTGCGCACTG CCGTGCAGCG
2521 ACAGACCCG CACCGCACGC TCGCGGACG GCCGGGAGAA GGAGTACGCC ATCGCGATCT
2581 CCTCGAAGTG ACGCAGGGGG AACAGCATCA GGGTGGCCGC GCTGTAGACC GTGACGAGCT
2641 GGCCGACGTC GATGCGGCCG TCCCGGGCGA GCGTCGCCCC GTACCAGACC AGGCAGATCA
2701 GCAGGATCCC CGGCAGCAGC ACCTGCACCG CCGAGATCAG CGCCACATC CTGGCGCTGC
2761 GCACGGCCGC GCGGCGGACC TCCTGGGAGG CGCGGCGGTA GCGGCCGAG AACAGCTCCT
2821 CGCCGCCGAT ACCGCGCAGC ACCCGCAGAC CGGCCACGGT GTCCGAGGCC AGCTCGGTGG
2881 CCTTGCCCGC CTCTCGCGC TGCTCGTCGG CGCGGCGGGT GCGCGCGGC AGCAACGGCA
2941 GCACGGCCAG GGCCAGCACC GGCATGGCGA GCGCCACCAC CAGCCCGAGG GACGGCAGAT
3001 AGACCGCCAG GCCGACGAGC ATCACCACGA GGGCGGTGGC CGCGGCCGCG AACCGGGAGA
3061 GCGCCTCGAC GAACCAGCCG ATCTTCTCCA CGTACCGGT CGACACGGCC ACGACCTCAC
3121 CGGCCGCGAC CCGTCGGGTC AGCGCGGAGC CCAGCTCGGC GGTCTTGCGG GCGAGTAGTT
3181 GCTGGACCCG CGCGGCGGCG GTGATCCAGT TGGTCACGGC GGTCCGGTGG AGCATGGTGT
3241 CGCCGACGGC GATCAGTACG CCGAGGGCCA CGATGAGGCC GCGCGCCAGG GCGAGCCGCC
3301 CTCCGAGCG GTGATGACG GCCTGGACGG CGAGCCCCAC GGTGACCGGC AGACCGGCGA
3361 TGCCGAGCTG GTGACGAGC CCCCAGGAGA GGGACTTCAG CTGCCCGCCG AGCTGATTGC
```

3421 GCCCGAGCCA GAACAGGAAG CGAGGGCCCG AACGTACATC GGGGTCGCCG GGATCCGAAT
 3481 ACGGAAGGTC GCGAATCTGC ATGACGTCCC AGGGCTCGTG AAACGGAGGT CCGGACAGAC
 3541 CTCGAAGACG GGGTGACGTG CAAGGCTCCC TGTTTCGTCCC GTTCCGGGGC AACC GGTTT
 3601 TTTCGGTCCG CCCC GCCCTG CGGGGTCCCG GGCCGAGCAG GCCCGGGACC CCACAGACGT
 3661 CACTCCGCGG GCTTCTCCGA GTCCATGCCG GACCGGGTCT TCTTCCACTC GCCCCGGGTG
 3721 AAGTCCGGGA TCGGCAGGGG CACGCCCTTG GCCTTGATGG ACAGATGGCT CAGCGGCACG
 3781 GGGGCCGTCC AGACCGCCGC GTCGTACACG TCGAAGTCGG GCACCAGACC GAGCCGCATG
 3841 CACTGCATCA GGCGGAACAC CATGATGTAG TCCATCCCGC CGTGGCCGCC CGGCGGATTG
 3901 GCGTGCTCCT TCCACAGCCA GTGGTCCCAC TCGGCGTACT TCTTGAAGTC GTCCCACTGG
 3961 TGGTTGGTGT TCGTGGGCTC CAGATAGATC CGCTCCGGGT AGTCCTCGAA CACGCCCTTG
 4021 GTCCCGCCGA GGCTGTTGAT CCGCGAGTAC GGGTGGGGCG ACGACACGTC GTGCTCCAGG
 4081 CGGATACCCG GGCCCTTGGC GGTCTGCACG AGGCTGATCG TCCGGTCGGC CCCGATGTAC
 4141 GACTCCTTCC AGCTCGGGTC GCCCGCAGGC ATGTGCTCCT CGCGGTAGGC GGCGAGGCCC
 4201 AGGGGGGTGG TGCCGACACT GCTGATGCTG ACGACCCGGT CGCCCCGGTT GACGTCCATG
 4261 TAGTTGGCGA CCGGACCGAA CCCGTGGTTG GGGTAGAGGT CACCGCGCAG CCGGGTGTGC
 4321 CACAGCCGCC GCCACGGACC CTCGTAGTAG TCGGGTCTGA ACATCAGCTC ACGCAGATCG
 4381 TGGTTGTAGG CCCC GGCGCC GTGCTGCAGC TCACCAAGA GACCCGCGTG CGCCATCCGC
 4441 AGCACCCGCA TCTCGTTCTT GCCGTAACAA CAGTTCTCCA GCTGCATGCA GTGCCGGCGG
 4501 GTGCGCTCGG AGAGATCCAC GAGCTGCCAC AGCTCTTCCA GGCGCATCGC GATCGGGCAC
 4561 TCCACCCCGA CGTGCTTGCC GTTCAGCATC GCCGTCTTCG CCATCGGGAA GTGCAGCTCC
 4621 CACGGCGTCA CCACGTAGAC GAAGTCGATG TCCCCGCGCT TGCAGAGGTT CTCGTAGTCG
 4681 TGTCTGTCTT TGGCATAGAT CGCCGGGGCG GGCTGACCGG CGGCCGTAC CTTCTTGGCG
 4741 GCCTTCTCCG CCTTGTCCCG GACCGTGTG CACACCGCCT TGACCTGGAC GCCCCGGAGG
 4801 GCGAGGAAGA GGTGATCAT GCTGTCGCCG CGGTTGCCGA GGCCGATGAT GCGGACCCGG
 4861 ACCGTGGAGC GCCGCTCGAA GGGCACGCCC GCCATGGTGC GGCCCTGCCG GGGAGGGGCG
 4921 GCGGCCACGG CTTCCGCGGC GCGGACGCGC TCCGGGGCGC TCCGCCCCGG CGCCGAAGCG
 4981 GTGCTGCGC CCAGTGCGCC GAGGCCGAGT CCGGCCCCGG CCACGCCCGC CGTGGTCCAC
 5041 AGCACCGAAC GGCGGCTGGG ATCCTGCCGG TTCACCTCGT CGGCCGCGCC GCTGTGCGGG
 5101 GGTATGTCTT GCGGTTCCGG TGCGGGCCGG GCGTCGTCTG TCATCGAGCC TCCAGGTGGG
 5161 GTTTGGGGGT TCAGACGGTG CGCGAGCGCG CCCGGTCCCG CCGTACGGAT ACGGGCGGGC
 5221 GGGACCGGGG CTCGGTAAGG ACCCTGGAGG GTGAGGCTGA TGGTGCAGAA GGAAGTATT
 5281 TGGACTCTTG TCCTCAAACC TTGGACTTTT CTCACGGCAC GCCGAAGCCC GACTGGTG
 5341 AACCAATCGG GGCCGTAAAA CGCTCATCTG TGCAGGCCGG CGGGGGTGCC CGGCCCCGCA
 5401 GTCACCGACT CACGGGAGAG TCGGCTGGGT GGCGTGTCTC AGTTCGATCA GCGCCGAGCG
 5461 GTACGGGTGC CCGGTGGCGC GTTCCATGCC GATCTCGCAC ATCCGGTTCTG CCGACAGATA
 5521 TGGCTCGTAG GGGCGGCGGT CGACCTCGGC GGCTCCTTG GCCGTCGCCG AGTCGGTCAA
 5581 CTCCTTGTGG AGCATGCCCC GGTCGCCCCG GAACGCACAG CACCCCGCGT CGTCCGGGAC
 5641 CACGACCTCC TGCGCGCAGG CCTCGGCCAG CGCCCGCAAC TGCCCCACGT CACCCAGATG
 5701 TTCCATCGAA CAGGTCCGAT GCAGGACCGC CGAGCCGGCC GTCCGGAACA CCGTCAGATG
 5761 CGGCAGCAGC TCCTCGGCCG CCCACACCAG CGAGTCCACG ACGGTCAATT CGCGGTGGAG
 5821 CGCCCGGTTG TCCTCGGTGA GGTAGGGCAC CACCTCCTCG GCGATGCCGA GCGTGCACGA
 5881 GGAGGCGTCC ACGACCAGCG GCAGCGTCCC GCCCCCGGTC CAGCCCCAGG CGGCCTCCAC
 5941 GATCCGGTTC GCCATGATCC TGTTGCCCGC GTCGTATCCC TTGGAATGCC AGATCGTCGC
 6001 GCAGCACGTG CCCGTGACGT CCTCGGGGAT CCACACCGGC TTTCGGGCCG GCCCCGACAC
 6061 GGCGACCACC GCCTCGGCCA GGGAGAGAGC GGGCCCCGCG TCGCCGTCTG CGGGCCCGGC
 6121 GAAGATGCGG TTGACACAGG CCGGGTAGTA GACGGCGCTC GCCCCACGC GTGCGGTGTC
 6181 CGGCAGCCGC CGGGCCGCAG CACCGGGGAT CTGCGGCAGC CACTCCGGTA CGAGATCGGG
 6241 GCGCACGGCC TTGCGGGCGA GGCGCGTCAC GGCCTGCAGC GGTGCGTCCC CCACCCGGTT
 6301 CCCGACGGTG TCGGCCGCG CCACGGCCAG CCGCGCCGAA GCCTCCACCG CGCGGAAGTT
 6361 CTTTCGGGTG AGGGCCGCGA TCCGCTCCTC GCGCGGGGTG TGCTTGCCTG GCCGGAAGCC
 6421 CTTTCATCATC GCCCCGGTGT CGATGCCGAC CGGGCAGGCG AGTTTGCAGG TGAATCCCC
 6481 GGCGCAGGTG TCCACGGCGT CATAGCCGTA CGCGTCCAGA AGGCCGGACT CCACCGGTGA
 6541 GCCGTCGGTC TGCCGCATCA TCTCCCGGCG CAGCACGATC CGCTGGCGCG GAGTGGTGGT
 6601 CAGATCCTCA CTGGGGCAGG TCGGCTCGCA GAAGCCGCAC TCGATGCACG GGTGCGCGAC
 6661 CGCCTCCACC TTCGGAATGG TCTTCAGGCC CCGCAGATGG GCGCGGGGT CCCGGTCCAG
 6721 CACGATGCGT GGAGCGAGCA CCCC GGCGGG GTCGATGACC TGCTTGGTCC GCCACATCAG
 6781 CTCGGTGGCG CGCGGCCCC ACTCGCGCTC CAGGAACGGC GCGATATTGC GTCCGGTGGC
 6841 GTGCTCCGCC TTGAGCGATC CGTCGAACCG GTCCACCACC AGCGCGCAGA ACTCCTGCAT
 6901 GAACGCGTCG TACCGGGCCA CGTCGGCCGG CTTGCGCCGG TCGAACGCGA GCAGGAAGTG
 6961 CAGATTGCCG TGTGCCCGCT GCCCCGCCAC GCGGGCGTCG AAGCCGTGGC GCGACTGGAG
 7021 CTCCAGCAGC GCCGCGCAGG CGTCCGCCAG CCGGGCGGGC GGCAACGCGA AGTCCTCCGT

7081 GATCAGGGTG GTGCCCCGAGG GCCGGGAGCC GCCGACGGCG GTCACGAACG CCTTGCGGGC
7141 CTTCCAGTAC CCGGCGATCG TCCCGGCGTC CCGGGTGAAC GCGTTGGTCA CGGACGCCCG
7201 CGGACGCACG AGGTCCAGAC CGGCCACGAC CGCGTCCGCC GCCCGCTCGA ACGCCGCCCG
7261 GCGCGCCTCG TCGGCCGCCC GGAACCTCAC CAGCAGCGCG GTCGTCTCCC GGGGCAGCGC
7321 CGCCCAGTCC GCCGGAACGC CCGGCACGCT GACGGAGGCG CGCAGGGTGT TGCCGTCCAT
7381 CAGCTCCACG GCGATCGCCC CCGCCTCGTT GAACCGGGG ACGCCGCCCG CGGCGGCGGT
7441 GAGGGAGGGG AAGAACAGCA GGCCGCTGGA GACCCGCCGG TCGAGCGGGA GGGTGTGAA
7501 GACGACCTCG GAGATGAAGC CGAACGTGCC CTCGGAGCCG ACCATCAGCC CGCGCAGGAT
7561 CTGCACCGGC GTCGCCCCGT CGAGGAAGGC GTCCAGGCGA TAGCCATTGG TGTTCTTGAT
7621 CGTGTAATTG GCGCGGATCC GGGCGGTGAG CTCCGCGTCC GCCTCGATCT CCGCCTTCAG
7681 CTCCAGCAGC CCCGCGCACA GCTCCGGTTC GCGGTGGGCC AGCTCCTCGT CGGCGGCGGG
7741 GTGCGCGGTG TCGACGACGG TGCCGCTCGG CAGCACGAAG GTGAGCGAGG CGAGCGTGCG
7801 GTAGGAGTTG CGGGTGGTGC CCGCCGTCAT GCCCGAGGCG TTGTTGGCGA CGACCCCGCC
7861 GAGGGTGCAG GCGATGGCGC TGGCCGATC GGGGCCCAGC AGCCTGCCGT ACCGGGCGAG
7921 GGCGGCGTTG GCCCGCATGA CGGTGGTGCC CGGCAGGATC CGGGCCCGCG CCCCGTCGTC
7981 CAGCACCTCC ACGCCGGTCC AGTGACGCGG TACGTCGACG AGGATGTCTT CGCCCTGGGC
8041 CTGGCCGTTG AGGCTGGTGC CCGCGGCCCG GAAGACCACG GATCGGCCCT TGCCATGGGC
8101 GTACGACAGG ATCGCGGACA CGTCGTCGAG GTCCTCGGGG ACCAGCACGA CCCGGGGGAG
8161 GAAGCGGTAG GGGCTGGCGT CGGAGGCGTA CCGCACGAGG TCGGAGATCT TCCGAGCAC
8221 CTTGTCCGCG CCGAGCAGCG CGGTCACTCG GCTCCGCGAG GCTCCGGGG TGCCGCCCGC
8281 GCTGCCGTCG GTGACCCGGT CGGGGGCGGG TTCCCGCGCC GTTCCGGGGC GCAGCGCTTC
8341 CGGGTCGGGC TCCAGCAGCG GCATGTCGGC CTTCCCTCG GCTCGGCGCT CAGCGGTGGC
8401 ACGCGGCAGC GCGGCTCAGC AGTGGCGCTC CGGCATTCCG TCGACCAGAG CGGACAGCAG
8461 CTCGCCGAAC ACCTCGCGCT GATCGGCGGT CAATGGAGCC AGGATCTCTT CTGCGGCGGC
8521 CCGGCGCGCG CTGCGCAGGG ACCGCAGCGT GCGCGCGCCC TCGTCGGTGA TCTCGATACG
8581 GACCACCCGG CCGCTGTGCG GATCCGGGGC GCGGCGCACC CGGCCGCTCG CCTCCAGGGC
8641 GTCGACCAGC GTGGTCACGG CGCGCGGGAC GACGTCGAGC CGTCGGGCCA GATCCGCCAT
8701 CCGGGGGGCC GCGTCGTAAC TCGCGACCGT CCGCAACAGG CGGAACGGG CCGGAGTGAT
8761 GTCGATCGGC TCCAGCTGGC GGCGCTGGAT GCGGTGCAGC CGCCGGGTGA GCCGCAGCAG
8821 CTGTTCCGGC AGCAAGCCGT CACGGGAGTC CCGGGAATCG CGAGAGTCCC GGGACTCGGG
8881 GGAATCAGGG GAGTCGGGGG AATCCGGGGG GTCCATACGG GAACAATATC AGGACCTTGT
8941 TCATTGTGAG CATAGGTAAC AATGAGCTAG GCTCTCACTG TGCGGGACCG GGAAGTCCCC
9001 GCGCCGCCTC ACGCCCGACG AAGGAGCCCA TGAACCCGA CGAACCCAG TGGAGCCCCC
9061 CGCCCGATCG CCGCCCGGCC GCCGACCGCG GCGCCGCCGA GGTGCGCCGC ATCCTCGGCC
9121 TCTTCCGGGT CGCCTCCCCG TTCCTGCTGC GCGAGATCCT CGACACCGCC ATCCCGCAGG
9181 TGGTCGGGGT CGCCTCCCCG TTCCTGCTGC GCGAGATCCT CGACACCGCC ATCCCGCAGG
9241 GACGCACGGG CCTGCTGACC CTGCTGGCGC TCGGCATGAT CCTCACCGCC GTGATGACCA
9301 GCGTCTTCGG CGTGCTCCAG ACCCTCATCT CGACCACCGT CGGCCAGCGC GTCATGCACG
9361 ACCTGCGCAC CGCCGTCTAC ACCCAGCTCC AGCGGATGCC GCTCGCCTTC TTCACCCGGA
9421 CCCGCACGGG CGAGGTCCAG TCCCGCATCG CCAACGACAT CGGCGGCATG CAGGCGACGG
9481 TCACCTCCAC CGCGACCTCG CTGGTCTCCA ACCTCACGGC CGTCATCGCG ACCGTGCTCG
9541 CCATGCTCGC CCTCGACTGG CGGCTCACCG TCGTCTCGCT GCTCCTGCTG CCGGTCTTCG
9601 TCGCGATCAG CCGCCGCGTC GGCCGGGAAC GCAAGAAGAT CACCACCCAG CGCCAGAAAC
9661 AGATGGCCGC GATGGCCGCC ACCGTACCG AGTCCCTCTC GGTCAGCGGC ATCCTCCTCG
9721 GCCGCACGAT GGGCCGCTCC GACTCCCTCA CCCAGGGCTT CGCCGAGGAG TCCGAGCGCC
9781 TGGTCGACCT CGAAGTGCGC TCCAACATGG CCGGCCGCTG GCGGATGTCC GTGATCGGCA
9841 TTGTGATGGC CGCCATGCCC GCCGTCTATCT ACTGGGCGGC CGGACTCAC TTCGCTCGCG
9901 GAGCCGCCGC CGTCTCCATC GGCACACTCG TCGCCTTCGT CACGCTCCAG CAGGGGCTGT
9961 TCCGCCCGGC GGTACGCTG CTCTCCACCG GTGTGCAGAT GCAGACCTCC CTCGCCCTCT
10021 TCCAGGCTGGA CTTCAATAC CTCGACCTCA CCGGTGGACAT CACCGAACCG GAACACCCCG
10081 TCCGGCTGGA GAGGATCCGC GCGGAGATCG CCTTCGAGGA CGTCGACTTC AGCTACGACG
10141 AGAAGAACGG CCCGACGCTG ACCGGCATCG ACGTGACCGT CCGCGCGGGC GACAGCCTCG
10201 CCGTCTTCGG CTCCACCGGC TCCGGCAAGT CCACCTCAG CTACCTGGTG CCGCGGCTGT
10261 ACGACGTCAC CGGCGGCGCG GTCACGCTCG ACGGCATCGA CGTCCGCGAC CTGGACTTCG
10321 ACACCTTCGC CCGGGCCGTC GCGCTGCTCT CCCAGGAGAC GTACCTCTTC CACGCTCGG
10381 TCGCCGACAA CCTCCGCTTC GCCAAGCCGG AGGCCACCGA CGAGGAGATC GAGGCCCGCG
10441 CCGCGCGCGC GCAGATCCAC GACCACATCG CCTCCCTGCC CGACGGCTAC GACACGATGG
10501 TCGGCGAGCG CGGCTACCGC TTCTCGGGCG GCGAGAAGCA GCGCCTCGCC ATCGCCCGCA
10561 CCATCCTGCG CGACCTCCG GTGCTGATCC TCGACGAGGC GACCAGCGCG CTCGACACCC
10621 GTACGGAACA GGCCGTGCGG GAGGCGATCG ACGCCCTGTC CGCCGGACGG ACCACGCTCA
10681 CCATCGCGCA CCGGCTCTCC ACCGTCCGCG ACGCGGACCA GATCGTCGTC CTGGAGGACG

10741 GCCGGGTCGC CGAGCGCGGT ACGCACGAGG AACTGCTCGA CCGCGACGGC CGCTACGCCG
 10801 CCCTGATCCG CCGCGACTCC CACCCGGTCC CGGTCCCCTG CCCGGCTCCC TGACCACCCT
 10861 TGTCGGGGCCG GCCCTCGATC AGACCGCCCC TGACGTCACC GCCATGGCCC GCATACGGCA
 10921 TGATCGCCGC GCATGAGAGC TCTCCTCGGG GTGGAAGTCC CCGGCTACCG CACCGTCGAC
 10981 ACCGACACCT GGCTGAACGA CCACGGCGAT GTGCTGTCCT TGCATTCTT CGACCTGCCG
 11041 CCGGACCTGC CGGCCGCGCT GGACGACGGC CCGGCCCTGC GGCACGGTCT GACCCACTTC
 11101 ACCGCCAGGG CGGGCGGCGG CCTCATCGAG ACATCGGTGA AGCGGCTGGG CGAGCTGCCC
 11161 GCCCTGCGGC AGATACTCAA ACTGCCGCTG CCGAACCAGC CCAGCGGCCA GCGGTTTCATC
 11221 GGCAGCTTCA CCGTGCCGCG CGCCGGATGC AGCACCCTGG TGAAGATCCA GCGGCGGAG
 11281 CGCGGCATGA CGGGCATGCG GGAAGCCGTG GTGATGGCCA AGCTCGGCCC CGACCAGTAC
 11341 TTCCGGCCGC ACCCTTACGC CCCCAGGTC CAGGGCGGGC TGCCCTTTCA CACGGCGGAT
 11401 CACGTCCAGT GGGACGCGGA GTTCCCGAC CATCCGCTCA CCCGGGTCCG CCGGACGCTC
 11461 GACACCCTCG CGGCGGCGGT GACGGTGGCA CCCGAGTTCG CCGCGCTGCC GCCCTTCACC
 11521 GGACCGGCTC AGGCGAACGG CTGAGCCGAC CGGCTGCGTA CACACAGCAC ACAGCACACA
 11581 GGGCACACGG CGCACACAGC ACACACGGCG GCGCCGCCGC TCCCGTGGGA CGGGGAGCGA
 11641 CGGCGCCGGG CGGAGCAATG GTGAGACGAG CCAACCCACG AAGTGGACGA CGCCGGCAAG
 11701 CAGGTTGGTC AGGAAGTTCA TCTGGTCTTT CTCTTGATC GTGGTGCATC TGTGGGACTG
 11761 CGCAGTAGCG GTCTGCAGCC CGTTGACTGC GCTCTGCAAT CATCACGCCC CGGACGAGTG
 11821 AAGAGCAACG AATCCCCCTGA CGATCACGCG TTCCAGCGAA CACCCGATCT CTTGTTTCGTG
 11881 TGTTCCGGCT ACGGGTGTTC TGTCCGCGTC GTACGGCGTT CGTGTGCGCC GGGCCGACGC
 11941 CGTGGTCGGG CTACCGGCCC TGGCTCGCAC CCGGGTTAA CGTGCCCGCA TGGTGAACGA
 12001 GTCCCGGAC GCCCGACCCC GTCGCAGACT CCGCCCGACC CGCCCGGAA AGATCGTCTT
 12061 GGTCTGTCGGC GCACTGCTCG TCGTACGGC CGCCGTCTTG ATCCCCCTGT CCCTGACCGG
 12121 ATCGGACGAG CCGCCGAAGA AGCAGGAGAC CCCGACAGC ACGCTGATGA TCCCGAAGG
 12181 CCGCCGAGTG TCCAGGTGT ACGAAGCGGT CGACAAGGCG CTCGACCTGA AGCCCGGCAG
 12241 CACGCTGAAG GCCGCGTCGA CCGTGGACCT GAAGCTGCCC GCCCAGGCCG AGGGCAACCC
 12301 CGAGGGGTAC CTCTTCCCGG CCACGTATCC GATCGACGAC ACGACCGAGC CCGCGGGCCT
 12361 GCTGCGCTAC ATGGCCGACA CCGCCCGCAA ACACTTCGCC GCGGACCATG TCACGGCCGG
 12421 GGCCAGCGG AACAACGTCT CCGTCTACGA CACGGTCACC ATCGCCAGCA TCGTCCAGGC
 12481 CGAGGCCGAC ACCCCGGCCG ACATGGGCAA GGTGGCCCCG GTCGTCTACA ACCGGCTGCT
 12541 CAAGGACATG CCGCTCCAGA TGGACTCCAC CATCAACTAC GCCCTCAAGC GCTCCACCCT
 12601 GGACACGTCG ACCGCCGACA CCCAGCTGGA CAGCCGTAC AACAGCTACC GGATCAAGGG
 12661 CCTGCCGCCG ACGCCCATCG GCAATCCGGG AGAGGACGCG CTGCGCGCCG CCGTCAGGCC
 12721 CACGCCCGGC CCCTGGCTCT ACTTCGTCA GGTGCGCCCC GGCAGACCCG GGTTCACGGA
 12781 CAGCTACGAC GAGCAGCAGA AGAAGCTCGA GGAGTTCAAC CGCGCCGTG GCTCCGCCAC
 12841 GACGGGCTGA CCGAATCGGC AGACGGGGCG GGGGGATTCA CACCCCGGC ACGGGCGCGG
 12901 GCACGGAGAC GACCGCCGAG GCCCCCTCGT CGGCGCCCGT CTCCTTCAGC AGCCGCATGA
 12961 CCGACCGGAC CGCCGCGCGG CCGGCGCGGT TCGCGCCGAT GGTGCTGGCG GAAGGGCCGT
 13021 ACCCGACGAG ATGGACGCGC CCGTCCCGTA CCGCACGGGT GTCTTCGGC CGGATGCCAC
 13081 CACCCGGCTC GCGCAGCTTC AGCGGGGCCA GATGGTCCAC GGCGGGCCGG AACCCGGTCG
 13141 CCCAGAGGAT CACGTCGGTC TCGACGGTAC GGCGGTCGTC CCAGGCCACA CCGGTCGGCG
 13201 TGATCCGGTC GAACATCGGC AGCCGGTCCA GCACTCCCCG CTCCCGGGCC CGCCGCACGG
 13261 CATCGTTTCA GCGCAGCCCG GTCACGCTGA CCACGCTCTT CGGCGGCAGC CCGTTGCGTA
 13321 CCCGCTCCTC CACCATCGCC ACGGCCGCCG GCCCCACTC CTCGGTGAAC GGACCTTCGC
 13381 GGAACACCGG TTCGCTGCGG GTCACCCAGA AGGTGTGCGC CGGTGCTCG GCGATCTCCA
 13441 TCAGATGCTG CGTACCGGAA GCGCCACCCC CGACCACGAG GACGCGCTGC CCGGCGAAGT
 13501 CCTCGGGCCC CGGATAGTTC GCCGTGTGCA ACTGCCGCC CCGGAACGTC TCCTGGCCCC
 13561 GATAGCGCGG CCAGAACGGC CCGTCCCAGG TGCCGGTGGC GTTGATCAGA GCCCGCGCGG
 13621 CGTACGTCCC CTCGGACGTC TCCACAGCA CCGGACCGCC GCTTCCCTCC CGTACGGCGC
 13681 TCACCTCCAC GGGCCGGTGG ACCCGCAGGC CGAAGCGGTC CTCGTACGCG GCGAAGTACG
 13741 CTCCGATCAC CTCCGACGAG GGCCGGTTCG GGTGCGCCCC GTTCAGCTCC ATGCCCGGAA
 13801 GCGCGTGCAT CCCGTGGACC TTGCCGTACG TCAGCGAGGG CCAGCGGAAC TGCCACGCAC
 13861 CGCCCGGCCG GGGCGCGTGG TCCAGCACGA CGAAGTCGTT GTCCGGCTCC AGCCCGACGC
 13921 GGCGCAGATG GTAGGCGGCG GACAGTCCCG CCTGACCCGC GCCGATGACG ACCACGTCCA
 13981 GCTCGCGCAC CCCAGAAATG TTCACGCTTC TACTAACTCG TCGGGCGCCC GGGATCATCC
 14041 CGGGCGCCCC ACGAGCGTCA CCGCACGGCT CAGCGACCCC CCGCGAGCAG CAGGGGAGCC
 14101 CCGCCCGGCG CCGTGGCGGT CCGGCTCTCG GCGCCACTCA CACCCAGCAG CGGCGACGCG
 14161 GGCACGGTCG ACAGCAGCCC GCGGCGGGAC AGCTCGGGTG TGACGCCCTC GCCGAACCAG
 14221 TACGCCTCCT CCAGATGCGG ATATCCGGAG AGCAGGAAGT GCTCCACGCC CAGCGCGTGG
 14281 TACTCCTCGA TCCGGTCCGC GACCTCCGCA TGGCTGCCCA CCAGCGCGGT CCCGGCCCCG
 14341 CCGCGACCA GACCGACCCC TGTGTGCGCG AGTGGAGATG ATCCTGCACG GCACGCCCCA

14401 CATGCGATCG AGCGTTTCTG AGGTCTGTGC CGTGGCCTCT CGGACATGGC CACGGCGAGG
14461 GAATCGGTGT TGCAGAGTGA CGTGTGTGCT TGAGTCGGCG GCTCGGGCGG GGGTGGCGTG
14521 GTCAACGGTT CGGTGGGCGG GCGGCTGAGG CTTGGTGGGT GTCCGCGCCG GAGCTGACGT
14581 GGCGGTGTAC TTCTTGTGGG CCGTGCCAGT CCAGGCACAC CACGGTGGCG TCGTCTTCGA
14641 GGCGGCCGCC TCGGCGCTCG CGTACGGCGG AGGTCAGCAT CAGGGTGGTC TCGCGTGGGT
14701 GCAGGCTGCG GGTCTGCCGT AGCAGGGCAG CTACGTCGAT CTCTCTCCG TGGCGTTCGA
14761 GCATGCCGTC CGTCAGCATG AGGAGCCGGT CTCCCGGGTG CAGGTCCAGG GTTTGGACGC
14821 GGTAGGGGCG GGGCGAGACG ACGGCCAGCC CGAAAGGCTG GTCGACCTGG CAGGGGATGG
14881 TTTCCACCAT GCCTGCACGC ATGCGCATGG GCCAGGGATG GCCGGCGTTG ACGAGCTCGG
14941 CCTTTCCGGT GTGGAGGTTG ATGCGCAGCA GTTGTCCGGT GGCGTGGCCC TGTCCGTGGC
15001 TGGTCAGGGC CTGGTCGCCC TGGCGGGCCT GTTCGGCGAG GGGGGCTCCG GCGCGGCGGG
15061 CTCTGCGCAG GCGGCCACC AGGACGGTGG CCGCCAGGGC TCGCCGAGG TCATGGCCCA
15121 TGGGGTCGGT CACCGACAGG TGCAGGGTGT CGCGGTCCAG TGCGTAGTCG AACGTGTGCG
15181 CGCTGAGGTC CTCGGAGGGC TCCAGGCTCC CGCTCAGGGT GAACTGCGCG GCCTCGCAGG
15241 ACAGGGCCTG TGAAGCAGC TGATACTGGA TCTCCGCTGC CAGGGTCCGG GGTCTGGAGC
15301 GTTTGCCCCA GGTGTAGAAG TCGGTGAAGC GCCCGTTGGC GATCACGACG TAGGCCAGCG
15361 CGTGAGCGGC TTCCCCGACA GCGAGCACA CCTCTTCCTC GTCGCTCCTG CCGGCCGGCA
15421 GGAGCAGTTC GAGCAGACCG ATCGCGTCCC CCCGGTTGGT CACGGGAACCT ATTACCCGCT
15481 GTTCTTGTCC GGCGGGCTCG TGATGCGGCC CGTGGGTGCG GATCACCTGC TCGTAGACGC
15541 TCCCCCGAA CAGAGGGATC CGCTCCGTTT CGTTTTCACT GCCCGCGGCA GTCGTGGTGG
15601 AGAGCGCGC GAGCGCTCTA CCGGTGAGAT CCACAATCAG GAATGTGACC TTCGTAGCCG
15661 CGAACC GCCCT TCTGCGACCA CGGCCACGGC CTCAACCGGA GCCGCCGTCT
15721 CCGCCGCCGT CAGCAGTCGG GACAGGTCGC TCGAGCCACG GTCATGGCA GCGGTTCCCT
15781 TCTCTGGATG TTTGGGGCCC GTTGCGCCCC GCCGCCAGT CGCCCTCCT CGTACCTGCC
15841 TTGCGCTCCA CGGTGGTCTGA CAGCGAGCAG CCCGGCACGG GCCCGCCCG TCTCGGCCCA
15901 TTGCGTGACA GTCCTGCATC CACCTGTTCC AGTCTGAACC TCAATCGGCC CTTTGTCCGG
15961 ATGAGGGACC GGGTCGGCCG GAGGCGAGGC GCCACCGGGT GAGGAAGGCG CCGACCGCCA
16021 CCTCGATGGG GTCGGGGCGG ACGATGTCAC CGAACTCGGT GGCGCTTCTC CCGCACCACC
16081 CCGAGCAGAG CGTTGTCCAC GGGCGATGTT CGTCGCCGAC ACCAGCAGGA CACGCCGCC
16141 CTGGGCGATG CGGTCGCCGA TGGCTCGCCG GAGGACGGTG GTCTTCCCTG TACCAGGCGG
16201 CCCACACCA GGTGGACGCC CTCGCCGAGG CATGCTCGAT ACGCGAGCCT GGGCAGGATG
16261 GAAGCCCGGC GGATCGATGG CGTGGGCAGA GCGACCGCCG ATCATCGCGG TCGCCAGAGC
16321 AGTGGCGAGA GGATGCTCAC CCAGCCAGGC GATACCGTCA CGAAGGCCT CGATCAGGAA
16381 GGTGGGCGGC TGCTTGAGCA TCCACAGGTG AGGGTCGTG ATCTGCGAAA TCTGTACCC
16441 GCAATGTGAG CAGGGAGCCG TTCTGTACAG CCTCGAAGAC TGCGAAACCC TCGATCTCGA
16501 TGCCGTCTTT GCCCGTCCCG GCGGCCCTCA GGGAGTCCAG CTGCCCAGGT CCGATGTGCG
16561 AGCCAGCAG ATCGACCACG TACCGGCCCG GGTACCGCT CCTGGCCGCC CGCCGACGA
16621 GCTGCCAGCG TGGCTGCCTG CCCACGCCTC CTTGACAGC GATCCACTCT CCGAGTGCCG
16681 AGGCGATTTT CTCACGCCAT CCCACCTACC GTCCCCCGA TCAGCCTCGG TCCGATCGCC
16741 TGCCCGCTGC TCGCTGTGTC CCTCCGGCTG CGATCCGGTT CGCTCGAAGT GCCTGCGGCC
16801 TGTTACGGG GCCGGTGGAT CCGCTCCGGA TCGCTGTCC TTGCAGGCAC GTTTCGTCCAG
16861 GCAGTCGGCT CCCGAAGCCG TCCAGGGCGC ATCACTCCGC AGGGAGCTAG AGGGCTGTCC
16921 CGTAAAAAAC CTCCGTCTCA GGGGCGTTGG GGGTAGTCAG GGTGATCTGC GTAGGGTGAC
16981 GCGAGACCGA GCAGGTCATC GCATGGCCAG GTCGCGCCGT CGTACCGACA GTCATCGCGG
17041 TTGTCCACG GCAGTTCGGG GTCACCCGTG GCAGGGCTGA GCCCATGTG GGCAGACAG
17101 CGCCGCTTGG CCTCGGCTTC CCGCAGGACG CGAGCCGGGT CGTGCAGCG GACGTGCGAG
17161 GCGATCGTGG AATGGAAGCC GGAGAGTTCT CCCTGGCAGA AGTCGACCG GTGTCCGTGA
17221 GGGGCCCACT CCCCAGGCC GTCACCGTCG CACCGGCCGG CCAGGTTGGC CTCCTCGTCC
17281 AGCCGAGCGT GGAGGAACGT CACAAGGTCT TGGCTCATGG GGTATCCTG GCCGACGGCT
17341 CCGCCGGTGG CCGGCCCACT GTTTGCGAAT TTGCGGGCGG TCTGTGCGAG GGCACGCCCC
17401 TGTCCGTGTT CGGCACGGAC GCGGGAGCGG GAGGCCCTT GGAACCGGAA TGCTCCAGCT
17461 TCGAAGGCAA CGGCGAGCAG CAAGGGGTCC GGCACCATCC CCGCACTCCG TGCGCCACAC
17521 CTGCGCCTCT CGCGCTCTTG TGTACGAGA ACTCCCAGAC CGCAAAGCGC CACACCCACC
17581 TGCAGTCGGA CGCGCATAGT CCGCCCTGGA TGCTGCGGAA TCGCATCCTC AGCCCGCGTC
17641 GCAGTATAGA TAGTCCGGGT TGTCCCAACG GTTGTGGTG GACACCAACA AGGACAATGC
17701 CAACCGTCAG GTTCGGCTGA GCTTCTTCGG AGGGTGAGCC GATCTGGTTG CACGAGAGGC
17761 GGGAAAGTCCG GCCGCACCAA CCGGACGCTC GGACCGTGTG TCCGAACCTG TCACGCAAGC
17821 CTGCACAGGC CCCCGCCGTG CAATCGAAGG GCTGCTCCTG GTGCCCGTGA TGTCCGTACG
17881 CGATCTCGTC GGGATGCCGT GTCACCCGTG CGAACCGCCA CGCCGCGCCG AGGGGCGCCG
17941 GCGCGGCGTG GGGAGGATGA GGTGGTGAA GGGGGTGCTG ATGACGGTTC GGCATCAGGG
18001 GGTGCGGTGG TGGTTCGCTC TTCTCGCTCT CGTCGGGTGC GTGGTCTGTG TCCTCTGCGT

18061 CGTCGCGCTC AGCGGGGCGG GGCACACTT CGGGCTCTCC TTGTGGGCGG GCATCGCGCT
18121 CGTGGTGGTG GGGGCGCTGT TTCCCTCGG GGGGCTGGGC TTCTGTACT GGGTGGACGA
18181 CGGCCGGTCC GAGGACAGCT TCCTCGTGAA GTTCTGTGC TTCGTCGCC ACTCCGCCGT
18241 CCTCGGGCTG GCAGCCGTCT CGTGACCCG GGCTGAGGCG TGGGCCTTTG AGCAGCGCGG
18301 GCGGTGGACG GAGGCGACGG TCGTGGGATA CAGCCCGCCC CGGGTGGTCC CGGGTGATCC
18361 GCCGACGAAG GTGCGGGCGT CCTGCGCGCT GGAGACCGCC GAGGGCGAAC GCGTCCGGCC
18421 CCGGCTGCCG GAGGGCCGCG GCTGCCGCGA CGGGGTGCGG CACGGGTCCC GCCTCGACGT
18481 GCTGTACGAC CCCCAGGGTCT TGCTGGCGCC CCGGGCCACC GAGCCCATGG ACCACGGCGT
18541 CACCGTCCCG GTCTTCGGGG GCGTGGCGAC CCTGTCCGGT TTCTTCGGCT GTGTCCGGCT
18601 CGCCTGGCGG TGGGAAACCC TCCGGGTACG CAGCGCGCGC CGCACGCGAG CGCGCCGAGG
18661 GCGGGAATCC GCAGCCGGTT AGGGGGTGGG GGCGTTCGCC GGCTCTCTT GCCGCCGTGA
18721 CCTGGAGCGC GGCGCGGCTG GAGCCACCT GCCGGGCCGA GTAGTTGCCT GCACTGCGCC
18781 CTTCTCGCCG TGGGAGATCG TGGCTGAGGC GATGGGCGGA AGACACCCG CTTTCCCCG
18841 TTCAGAGGGG AAGGCCGGGT GTCAGGCGCA AGGACCTGCG AGAACCCGGA AGGATCCTGC
18901 TGCCGGGCGG GTCATCATTT CTTGAATGCG CGCATGTACT TTCCGAACCT CTCCAGGCCG
18961 TCGATATTGC GCGGGCTGCT GATGCCCTCG TTGTAGTCGA GGACGAAGAA GTTGTCTTCT
19021 TTCACGGCCG GCAGTTCCTT GGTGTGCGGC GACTTCTTCA GGAACCTCGAT CTTCTTCTCG
19081 GCGGGCTGGT CGCCGTAGTC GAAGATCATG ATGACCTCGG GCTCGGCCTG GGTGACGGCT
19141 TCCAGTTCA CCTGGGTCCA GCGCTCCTCC AGGCCGTCGA AGATGTTCTT CCCGCCCGCG
19201 GTCTTGATGA TGTCGTTGGG CGGCACCTGG TTGCCCGCCG TGAACGGCTG GTCGGTCCCG
19261 GAGTCGTAGA GGAACACGGG CACGGGCTTG CCCTTCGGAG CCTGCTCGG GACGGCGGCC
19321 TCGCGCTTCT TCAAGCCGGC GACGACCTTC TCCGCCTCCT CTTTCGACCT GAAGATGGCT
19381 CCGAGGCGTT CGAGGTCGGT GTAGAGGCC TTGAAAGCG TCAACTTCTC GCGATGGCCC
19441 GGGTAGTTGT AGCAGCTCTC ACTGTGCATG AAGCTCTGTA CGCCGAGCTT GTCGAGGATC
19501 TCCGGGGTGA TGCCCCGCTG GTCGCTGAAG CCCGAGTTCC AGCCGGCGAC GACGAAGTCC
19561 GACTTGCGCT CCACGACGAT CTCCTTGTG AGGAGGTCGT CGCTGAGCAT CTTACCTTG
19621 CCGTAGTCCT TCGCCAGGG AGACTCGCTG ACCGGCGGGT TGGCCGGCGG CATGACGTAG
19681 CCGTGCACGT GGTGCGCCAG GCCCAGACTG AACAGCTTGT CGGCGCTGCC GCCCTCGTAG
19741 GCGACGGCCC GCTTCGGCAC CGTGTACTCG ACGGACTCGC CGCAGCGCTT CACGGTGCTC
19801 TTCCCGGAGC CCTTGCCCTG GGATTCGACC TCGGCGCCAC ACCCCGTGAG CAGGAGCGCG
19861 GACGCGGCGA CGGGGATGGC GAGTTTGGTG AACTTCATGG TCTTCTCAG GAATCGAGTG
19921 AGTAGAGCAA CTGGGGGTG CCCGTCAGCG GATGCGGGAC GACGAGGCG CGGACCCGA
19981 ATACCTCGTC GACGAGTTG GCGGTGAGGA CGTCCCTGGG CGTGCCCGAG GTGATCAGGC
20041 GGCCTTCGCT GAGTACGCCG ATCCGTCGCG ACGCGCGGGC CGCGAGGTT AGGTGCTGGA
20101 GTACGACGAG GACGGTCAGG CCGGCACCGC GCAGCAGGGA CAGGAGCCGC ACCTGATGGC
20161 GTACGTCGAG ATGGTTCGTC GGCTCGTCGA GGACGAGGAT CTTGCGCTCC TGCACGAGG
20221 CGCGGGCGAG CAGGACGCGC TGGCGCTCGC GCCTGAGGAT GGTGAGGATG CCGCGTCGGG
20281 CCAGGTGCAG GATGTCGAGC CGACGCATGG CGTGCTCGCA CAGATCCCGT TCGTGACCGT
20341 TCAACGGGGT GCTGCCGCGC TGGTGGGGTG TGCGGCCGAG GGCGATCACC TCCTCGACGG
20401 TGAAGTCGAG GTCGACGGCG CCGTCTGGG TCATCGCCGC GATGAGCTGG GCGCTGCGGC
20461 GCATGGTCAG CGACGAGAGC TCCTGGCCGT CCACCTTCAC GGTGCCGGAG CTGGGTTTCA
20521 GGGCCCGGTA CACGCACCGC AGGGCGGTGG ACTTGCCGCT GCCGTTGGGG CCGACGAGGC
20581 CGACCACCTG ACCGCTGCCG ACGTCCAGGG AGAGGTCCCG TACCAGGCTC TTGCCGTGCG
20641 TCACCACCGA GAGCCCGTCG AGTTCGAGGT CCATCTCAAC GGCTCCGAA CATGTAGGAC
20701 TTGCGGCGCA TCAGGGTGAT GAACACCGGG ACGCCGACCA GCGCGGTGAT GACGCCGAGC
20761 GGCAGCTCGC GGGGGGCGAC CAGGGTCCGC GACACGAGAT CGACCCAGAC CATGAAGACC
20821 GCCCCGCGCA GTGGTGCGAC GCGGAGCACC CGCGCGTGCG TCGCGCCAC CACCATGCGT
20881 ACGAGGTGCG GCATGACGAG GCCGACGAAG GCGATGGAAC CGCTGACGGC GACCATCACG
20941 CCCGTCACCA GGGAGACGAG CACGAGCAGG GACTTGCGGT GTCGGTGGG GCTGATGCCC
21001 AGGCTGGCTG CGGTCTCGTC ACCGAGAGCC AGGACGTGCA GCGGGCGGCC GTGCCGGTGC
21061 AGGACGAGGA CACCGAGCAG CACGGCGGCG GTGACACCG GTGACGAAACC CCAGGAAGCG
21121 GCGCCGAAGC TGCCCATGGT CCAGTACAGG ACCATGCTGG TCGCTCGGA GCTGGGCGCG
21181 AAGTAGATGA TGACACTCAT CACGGCCTGG AAACCCAGCG ACATGGCGAC ACCGGTCAGT
21241 ACGAGCCGCA CGCCGAGAG CGCCCCCTTG GTGGACGAG CGCCGTACAC CAGGACTGAG
21301 GCCACGAGCG CGCCGAGGAA GGCGCCACG GACACCGCGT AGATCCCGAA CACGGCGAGC
21361 CCGCCCATGA CCGTCACACC GACGGCGCCC ACGGAGGCC CCGAGGAGAC GCCCAGAACG
21421 AACGGGTCGG CCAGCGCGTT GCGCACCAAG GCCTGGATGG CGACACCGAC CGCGCTGAGC
21481 CCGGCCCCCA CGAGCGCCG GAGCAGGACG CGCGGGGTGC GGATCTGCCA GATGATCTGG
21541 TACGTCGTCA CCTCGTCCGC CGAGATCGGC CCGCCACTGA GCGCGGCCCA GAGGAAGCGC
21601 GCGGTCTCGG CCGGGGGGAC CACGGCAGGC CCGAGACCGA TGGCGACGAC GACGGAGACG
21661 ACGAGCGCGG CGAACAGGCT CACGCAGATC GCCACCAGGC CCGTCCGGGA GCCGGTCCGA

21721 ACCGGCTCTT GCGCGGTGGG CGCGGGACGT TGCAGCGCCT CGGGTGGCGC GGGCGGTGAC
 21781 ATGTGGATCG GCCTTCCGGT TTCGGAGCGT TGATGAACGG TGGATGTGCG TCCGTGGGGT
 21841 GCCCGCGACC TTGGGCGGGC GCGCGGTCGG CTTCCGGCTAC GCCGAACCGG GGATCTCGTC
 21901 CTCGGAGCGC AGCACCAGGA GCGCGGCCAC CACGGCCACG GCGACGAGCA GCGCGAACGC
 21961 GGCGGCGATC CCCGGGTACC CCGCGAGCCC GAGTCCCGCT CCGCCGAGGG CGGCGCCGGC
 22021 GAAGACGCCG AGGCTCTGGC CCGCCGCGTT GAGGCTCAGC GCGGAACCCC GCATCGATCC
 22081 GCAGCGCCTG ACCAGCAGAC TGACGGCGCA GGCGGCGACG GCCGCGTGGC TAGCGGCGTG
 22141 CAGCGAAGTA AAGGCCAGGG CGAGCGGCAG CCAGGTCGTG AACCAGAAAC CGGTAGCGGT
 22201 GACCAGGGCC GCCAACAGTC CGACGAGCAA GAGCTGTTTC GTACCCACGG TGGATTTCTC
 22261 GGCGTTGGTG ATGCGGCCCC TGAGCAGGTT GCTGACGAAG AACGAGGCGC CGCTGAGCGT
 22321 CCACACCAGC GAGAACAGGG CGGGGTCGAG GTGGAACCGG TCGTCGTAGT AGACCGCGAG
 22381 GTAGGCGAGG TAGCCCATGA AGACCGCGGT GCGCAGGAAG GAGATGGCGA GCAGCGGCAC
 22441 CGAGCCGCGG ACCTGGGCCA GGGCCTTGAA CGAGGCGAAG TAGCCCGTGC GCGGGCCACC
 22501 CTCGACCACC GGGTCCTCGC CTTTCTGCGC GCGTACGAGG AAGACCGCGG CGAGCAGCAG
 22561 CGAGACGACG GTGACGGCGA GCAGGTCGCC CTCCCATCCC CACAGCAGGG CCGGCAGGGC
 22621 GATCAGGGGC GCGGCGAGCA TCGCCGTCAT CGAGGTCGTC GACGTGACGA GGGTGGCCCG
 22681 ACGGGCGGCG GACTTGCCGT CGCCGAACCG GTCGGCGGCG GCAGCGGTGA GCGCCGGGTT
 22741 GATCACC CGG GTGCCGCGC CGACAGCAG GCAGAACACC GCGGTCAGGA GGAAGTCTCC
 22801 GCTCGCGCCC AGGGCTGAGG AGACGGCGAG TACGACGAGA CCGACCGCGA CCGCCTTCGA
 22861 CTTGGGTACC CGGTCGATCA GGGGGGCCAG GGCCGTGCCC ACGGCGAGCG CCGCGAGGCC
 22921 CCCAGGCCCG CGCAGGCCGC CCACCGCGGC GACACCGCTC CCGGTCTCCT CGGCGATCGG
 22981 CACCAGATAC GTGCTGAAGA CGGTGAACGG CAGCAGGCCG ACGGCGGAGG CCACCAGGAC
 23041 CGGCCACAGG GCTCGCGCCA TCTTCAGGTC GCCGGGCATC TCGGGGGACT TCTCCGGTGC
 23101 GACGGCCGAA CGGGAGGTGC CGGCGCTCAC AGGTACCCGC CTGCGCGGTA GCGGTACATC
 23161 GTCGTCTCGT CGGCGCTGAA CTGTGAGAAC GGGGAAGGGT CCGCGTTTCA GCGGTGACG
 23221 CCCGAGCCGA GGAACGCGCG GCGCAGGCG CTGCCCCTCT GGGCGTACAC GACGAGCGGC
 23281 ACCTTCTGCT CGCGGCAGTG TTCCAGGATC AGGTCTGAAG TGCCGTTGGA GAGTGTCTATC
 23341 CCCGTGGCGA CGACGGCGTG GGCCTCTGCG AGGACCTCGG TCATGTCTGC CGCGACCGGC
 23401 TCTCCCCACT GGGTGGTTTC CAGGTTGAGG TCGCACGGCA GGCAGACGCC GCCCCGCTCG
 23461 CGGATCGCGG CGACGAGCGG GTTGACGACG CCGATGAGCG CGACCTTGGC GCCCTCCTCG
 23521 ATGTCGAGCA GCCCGGCGAT GGACGCGTCC CGCGCCTTCG CCCGCACCTC GGGGGTCCCC
 23581 ACCGGCAGCG GGACGGCCTC CTGCTCCGGG GCTTCCCGAT GCGGCTGTAT CTGTGCGAGG
 23641 TAGGCGTCGA GCGCCGCTAT GCGCACCGGG CCGGACTCGT GGCAGCAAA CTTTCCAGC
 23701 GGGTGCCCGG AGGCGTTCTC GCAGAAGTCC GGGGTGAGTT CGCCTGCCTC GAAGGAGCAG
 23761 CCGCCGAAGG ACCGGCCGAC ACGCAGCACC AGGTAGTGGT TGTGGTACGT CACCGGTCCG
 23821 CCGGCGAGCC GTGTCGTGTG GTAGAGCCAG AACCGGCTGG TGACGGTCAT GTCCTTCGGG
 23881 TCGGGGCCGT AGTCCCCGGC GAGGACGGCA TCGGTGAGCT CGGCGACCGA CTGCGGCGTG
 23941 GGAAAGGGCA TGTCAGAGGG CTTTCTTCTG GTCGGAGGTG GAGTCATCGG TCCACGTCAT
 24001 GGCGGACCAG GGGTGGCTGA GCTCGTCGAG CGAGGTGATC TCGCGCGGTT CGAGGTCTCT
 24061 GATGTCGGGC GCCTCGGTGT GCTTGGCGTA CGCGCTGTGC ACGTAGCGGT GACCTGTGTC
 24121 CGCCGCGATG AAGACGTACG TCCGGGAATC GTCCTTCGAC CGCTCCCACC GGGTGGTTCAG
 24181 GTAGGCGGCG CCCGCGGACA GGCCTGCGAA GATGCCGCTG GAGCGGAGCA GGTGGACGGC
 24241 GCCTGCGAGC GCGGAGTCGA AGCTGACCCA GTGGATCCGG TCGTACAGAT CGTGCCGGAC
 24301 GTTCTCGAAC GGGATGGCGC TGCCGATGCC GGCATGATC ATGTCCGGGT CCGAGACGTG
 24361 CTCCGAGCCG AACGTGACGC TGCCGAAGGG CTGGACTCCG ACGAGGGAGA CGTCTCGGCC
 24421 CGCCTCGCGC AGATACGAGG CGATGGCGTC TGTCGACGCG CCGGAACCCA CGCCGCCAC
 24481 CAAGTCCAGG GGCCCGGCGG GCACCTCGTC GCGCATCGTT TCGGACCACTT CGCGGTAGCC
 24541 GTAGTAGTGG ATGCTGTCGT GGTACTGCCG CATCCAGTGG TACGAGGGGT TCTCCTCCAG
 24601 GATCTCGGCG ATGCGCCGCA CCCGGAGCTC CTGGTCGAGG CGGAGATTCC TGGACGGCCG
 24661 CACCTGCTCG AGCGTGGCAC CGAGAATCTC GAGCTGCGCC TTGAGCGTGC GGTCCACCGT
 24721 GGTGACCCCC ACGATGTGGC ACTTCATGCC GTAGCGGTGG CAGGCGAGGG CGAGGGCCTG
 24781 CGCGTAGATG CCGCTCGAAC TGTCGACGAG GGTGTACCG GGTTTGACGG TGCCCGACTC
 24841 AAGGAGGTGC CGCACCGCCC CCAGAGCCGA GTAGATCTTC ATGGTCTCGA ACCGAGACA
 24901 GACCAGGTCC GGCCGAGTG CTATGAGATC GGGTTTCTTG ATCGCTTCAG CTATGTGCTC
 24961 GTACATCTCC GTCTTCCGGT CGAGCGGGAC ATGAACCGTC TGCCTCGATC AGGTCCGGCT
 25021 GGGCTGGGCC GCGGTGTGGC CGTGAGCCCC GACGAGAGCA TTATGGAAAT GAAAACGATT
 25081 GTCAAAACCG AGTAAGGTGT GCGCCAGTCA TCACCACGGG AGCCGCACAG GCAGCTCTAC
 25141 GCCCCGTGAC GGGCAGCAAG GCTTTTGGAG GAACTCATGC ATCTGCCCCG GGTGGTCCG
 25201 CGATCTGCC TGTCGGGTCG GCGGGGCATG GACACTGGAG TGGGCACCGC CTACGGAACG
 25261 TTCGGGGAAC TGCTCCAGGG TGAAGTCCCG GAGGAGGCG GCGATTCTCT CGTCACGCTG
 25321 CCTGTCGCCC GGTGGGCGAG GCGCTCCTTC CGGTGCACC CCGCCATGGG AGATGTCTATC

25381 GTCAGGCCGT CGCACAAGGA GAAGGCGAGG CGGCTGGCCT GCCTGATCCT GGAGGAGGCA
25441 CCGGGGATGA CCGGTGGGGT GCTGACGGTC AACAGCGTGA TCCCGGAGGG CAAAGGGCTG
25501 GCCAGTTCAT CCGCCGACCT GGTCGCCACG GCGCGCGCGG TGGGCGGGG CCTGCGGCTC
25561 GACATGCCGC CATCGCGGAT CGAGGGGTG CTGAGGCTGA TCGAACCAGC CGATGGTGTG
25621 CTGTACCCGG GAATAGTCGC CTTCCATCAT CGAGCGGTGC GACTGCGCGC GATGCTGGGC
25681 TCGTTGCCCC CCATGTCGGT CGTCGGTGTG GACGAGGGCG GGGCCGTGGA CACGGTCGAC
25741 TTCAACCGCA TACCCAAGCC GTTCACGCCG GCGGACCGGC GTGAGTACGC CGACCTGCTG
25801 AACCGGCTGA GTGGGGCCGT TCGCTCACGC GACCTCGCGG AGGTGGGAG GGTGGCGACG
25861 CGCAGCGCGC TCATGAACCA GCCGCTTCGG TACAAGCGAC TGCTGGAGCC CATGCGGGAG
25921 ATCTGCAGGG ATGCCGGTGG TCTGGGCGTG GCCGTGGGCC ACAGTGGGAC GCGCTCGGC
25981 GTGCTCCTGG ACGCCGCGGA TCCCGCGTAC CCGCACCGGG CCACCGCGGT GGCCCGGGCG
26041 TGGCGGGATC TGGCCGGGGC CGTCGCGGTC TATCGGACCC TCAGTTTCCC GAACGCCGTC
26101 AGCCATGGTG GTCGGACCGT CGGCTGAGGG CGGTTCCCGG AGGCATGCCC CGACGGGGCC
26161 CGATGGCGCG GCAAGCAGGG ATTCGCCTGA CGTTGAGGGT GGCCCGGATC GCTGTATGGT
26221 CACCGCGGTG CCGGTGCGTG GACCGTGTCA CTCCCGGCTC CCTTGTGAAG CCGATCGCCG
26281 GTGCTCCGCG GACGCTGTGA AGGTGGACGG CCTCGACCGG TTCGTCCAAG GGCCCGAGGT
26341 GCCAAGGCCCT CTGCGACCGG TATCGCGGAC GCCCTCGGGC ACGTGGACTT CCTCTCGGCC
26401 GCCGCCGGGC CAACCGTTCC GGACAATCGA AGGGACCCAG GTTCATGCTC ACCGCACAGC
26461 AGCCTGCTCC CGGCGTCTGT CCCGCCGGA TCCACGTAC GAGCAGGTG GAGGCGGCTC
26521 ACCCGCTCGC CGCTGACGGG GCTGCTCGTC TGACAGGCGT CGAGCCCTCC GGTGACGGCC
26581 TGGTCCTCGC CGCCGCGACC GTCCTGGGGG AGCGGCTGCA GCAGGTGTTC CCTCACCGGC
26641 TGGCGGCGTC CGACGGCTCG AACTTCGTCC ACCTTCATGC GGACAGCTTC GACTTCGTGCG
26701 TCAACGTAGG GGGCGTCGAG CATCGCCGAC GTGATCCGGA TGAGGACTAT GTCCTCATCC
26761 AGTGCCTCCG GCAGTCCGAC TCCGGCGGCG ACTCCTTCGT GGCTGACGCC TATCGCTTCG
26821 TGGACCACTG CGCGACGGCC GATCCTGAAC TGTGGGACTT CCTGACCCGA GGGGACGTCG
26881 ACCTGTACGG CGCGTGGTCC GGACTGCGTG GTATGCCCGC AACCCCTTT GTGGGCAGGC
26941 ATGTCGAGTA CACCCGCGCC GGTCGGCGTA TCGTCCGGCG CGGCGACGGG GTGACCCCTC
27001 TGCACCGGGA CCCTGGCGCG GACCACACCC GCGGATGCT CGCCCGTCTG GAGGAAGCCG
27061 TCCATGCGCT GGAGGAGACG CTCCC CGCAT TCCGGCTCGA CAAGGGCGAA ATCCTCGTCC
27121 TGGACAACTA CCGCTGCTGG CACGGCCGCG AGGCTCACAC GGGAGATCGC GCGGTACGTA
27181 TCCTCACGGT GCGCAGCAGC GACGCCCGCT GAGGCGCTGT TGGTTCGCCT CACTCGCCGT
27241 GACACAGGGG CAGGCGTCTG CGGCGGTGCT GTTTCGCGC GGGACGGACC GGGGAGATT
27301 CCCCGGTCGG TAAAGGGGGG GACCGGCGAT CCGCTCACCC CGCCTCGATC ATTGCGCAGG
27361 CTCTTCGAGC GCTTCGTGCT TCACGCCGGC TGCCAGATCC GGGCCAGTGC CTCGGGGGTG
27421 AGTACTTCCT CCGGTGATCC CTGCCC GATC AGTCGTCCGT CGGCCAGGAG CAGGCAGGCG
27481 TCGGCCGAGC GGGCGGCGTC CAGGTCTGTT GTGGCCTGGA CGACGGTGGT GCCGTCGGCG
27541 ACCAGGTCCG TCAGCAGGGC CGTGATCCGC TCCCGCGCCT CGGGGTGCGA TCCGGTGGTC
27601 GGCTCGTCCA GGAGAAGCAG GTCGGACTGT TGGGCGAGGC CCTGCGCGAT CAGCACGCGC
27661 TGACGCTGGC CGCCCGACAG CTCGCCGAGC TGGCGGGCGC CGAGGTGCGC GACCCCGAGC
27721 CTCTCCATGG CGGAGTCGAC CGCGGTCCGG TCCGTGCGGG TCAGCCGCCG CCACAGGCC
27781 CGTGTCCTCC AGCGGCCCAT CTCCACCGTC TGCCGCGCCG TGAGGGGGAG GGTGTGCGCC
27841 ACGGCACCGC GCTGCGGGAC GAAAGCCGGC GGGGAGCCCT CTGCGTACCG GAGTTGTCCG
27901 GATGTGGCGG TGATCACTCC GGCCAGGACG CCCAGCAGCG TCGACTTGCC GCTTCCGTTG
27961 GGTCCGACCA GGGCGGTCAT GGCCAACGGC GGTATTGCGG CGCTGAGTTG GTGGAGCACG
28021 GGGCGGCCGG GGTAGCCGGC GCTCAGCCGC TGGAAACCGA CGCGTTTATT CCGCAGTTCC
28081 GTGGCCGGCG GGAACGGAGG GTTGTTATTG AACATGTTG TCATTATATG GTCCCTCGTAT
28141 GGAGTGGTTG ACGGCCCCCT TCGAGTTGGC CTTTGTGTCAG AGGGCCCTAT GGGCCGGGAT
28201 CCTGGTGTGCG GCGATATGCG CCCTCGCGGG AACGTGGGTG GTGCTGCGCG GGATGGCCTT
28261 CCTCGGTGAC GCGATGTCGC ACGGGCTGCT GCCCGGCGTC GCGGTGCGCT CCCTGCTGGG
28321 AGGCAACCTG CTGGTGGGGG CGGTGGTGTG CGCGGCCGTG ATGGCGGCGG GCGTCACGGC
28381 CCTCGGGCGG ACTCCGCGAC TGTCCAGGA CACCGGCATC GGCCTGCTGT TCGTGGGCAT
28441 GCTGTCGCTC GCGTCATCA TCGTGTGCGG GTCGAGTCC TTCGCGGTGG ACCTCACCGG
28501 CTTCTGTGTT GGAGACGTCC TCGCCGTGCG GGGGAGCGAT CTGCTGCTTC TTGGAGTAGC
28561 CCTGCTGCTG GCGCTGGCCG TCTCGGTGCT CGGCTACCGG GCTTTCCTGG CCCTCGCGTT
28621 CGACGAGCGC AAGGCCCCGA CACTCGGGCT GCGTCCCCCG CTCGCCCATG CCGTGTGCT
28681 CGGCCTGCTG GCGCTGGCCA TCGTGGCCTC CTTCCACATC GTGGGCACGC TGCTCGTCCT
28741 CGGTCTGCTC ATCGCCCCGC CCGCGGCGGC CATGCCCTGG GCGCGAAGCG TCCAGGCGGT
28801 CATGGTCCTC GCGGCGCTCC TCGGCGCCGC CGCCACCTTC GGCGGCCTGC TCCTGTCTCTG
28861 GCATCTGCGC ACCGCGGCCG GAGCGACCGT CTCGGCCCTC GCGGTCGCTC TCTTCTTCCT
28921 GTCCACCTG GCATCCGGAC TTCGGCACCG CCGCCGTGCG CGCCGGGGCG GTCTTGCCGA
28981 ACCGGCGGTC GCCCCGGGCC GCGACCTCCT CCACGTCCTG ACCGAGAGAA ACCTGAGGCG

29041 ATCTCCTTGC TCGTCCGAAA AAACGTCACA TCGCTGGCTC CGGCGCTTGC GGCCGTGATC
29101 CTCCTGACCG CCGGATGCGG GGGCGGGGAC GAGGCCAAGT CCGGTTCCGG GCCCGCTCT
29161 TCGTCCCCCA CTCCGCACCG CTATGTCGAA GGCGCCACCG AGGCGGGCCGA GCAGCAGTCC
29221 AGACTTCTGC TCGGCGACCC CGGGAGCGGT GAGACCCGCG TGCTGGACCT GATCACCGGC
29281 AAGGTGTACG ACATCGCCCG CAGCCCCGGT GCCACCGCAC TCACCACGGA CGGCCGCTTC
29341 GGCTACTTCC ACGGCCCGGA CGGCATACCG GTGCTCGACA GCGGTGCGTG GATGGTGGAC
29401 CACGGCGACC ACGTCCACTA TTACCGCGCG AAGATCAAGG AGGTCGGCGA ACTCCCGGGC
29461 GGCACCGGTA CGAGCATCCG CGGCGACGCG GCGGTGACCG TGGCCTCGTC GCGGACGGG
29521 AAGGCGAGCG TGTATCGCAG GCGGACCTG GAGAAAGGCG CCCTGGGCAC GCCGTCCCCG
29581 CTGCCCCGGA CGTTCCGCGG CGCCGTGCTG CCGTACGCGG AACACCTGGT GACACTCACC
29641 GCTGAGAGCG GGGCTCCGGC GAAGGTCGCC GTGCTGGACC GTTCCGGCAA GCGCGTCGCC
29701 GCTCCGAGG CGGAGTGCGA GGAGCCTCAG GGCGACGCGG TCACCCGGCG CGGGGTGTGTC
29761 CTCGGCTGCG CCGACGGCGC TCTGCTCGTC CATGAGGACG ACGGCGCCTT CACGGCGGAG
29821 AAGATTCCGT ACGGCGAGGA CGTGCCGAAG ACCGAGCGGG CCGTGGAGTT CCGGCACCGC
29881 CCGGGCAGCA GCACCTCAC GGCACCCGCC GGCAAGGACG CTGTCTGGGT CCTGGATGCC
29941 GGCGAGGGCG CCTGGACCCG GGTGAAGACC GGCCCCGTGG TCGCCGCCAA CACGGCCGGC
30001 GAAGGCTCGC CGCTGGTCTG CTGGAGACC GACGGGGCCC TGCACGGCTA CGACATACCC
30061 ACCGGCAAGG AGACCGGCGT GACCGATCCC CTGCTCAAGG AACTGCCCCG AACCGGTGCG
30121 GGCGGCGGCG CGGCTCCGGT GATCGAGGTG GACCGCAGCC GGGCCTACCT CAACGACCCC
30181 GAGGGCAAGC GCGTGTACGA GATCGACTAC AACGACGATC TCCGCGTGGC CCGTACGTTC
30241 GACGTCGACG TACGGCCGTC CCTGATGGTG GAGACGGGCC GATGAGCGCG GCGCTGGGCG
30301 CTCCACGGAT GCGTGCCCTG CTGGTGTCCT TGGCCGGATT CTTCGTCGTC CCGGTGCGG
30361 CGACCGGCTG CGCGGGCGGC GGAGACGAAC GGCCCCGGGT CGTGGTGACC ACCAACATCC
30421 TCGGCGACAT CACCCGGGAG ATCGTCGGGG ACGAGGCCGG CGTCAGTGTC CTGATGAAGC
30481 CCAACCCGA CCCGCACTCC TTCGGCTCTT CGGCCGTGCA GGCCGCTGAG TTGGAGAACG
30541 CCGACCTGGT CGTCTACAAC GGGCTCGGCC TGGAGGAGAA CGTGTTCGCG CACGTGGAGG
30601 CTGCCCCGGA GTCCGGAGTG GCCGCCTTCG CCGCGGGTGA GGCGGCCGAC CCGCTCACCT
30661 TCCATGCCGG ACAGGACGGC GGCCCCGAAG AGGACGCCGG CAAGCCCGAT CCGCACTTCT
30721 GGACCGACCC CGACCGCGTA CGCGAGGCCG CCGGCCTGAT CGCCGACCAG GTCGCCGAGC
30781 ATGTGGAGGG CGTCGACGAG AAGAAGGTCC GGGAGAACGC CGAGCGGTAC GACGGACAAC
30841 TCGCCGACCT CACGGGATGG ATGGAGAAGT CCTTCGCCGC CATCCCCGAG GACCGGCGTG
30901 CCCTGGTGAC CAACCACCAC GTCTTCGGCT ACCTCGCCGA CCGCTTCGGC CTCCGCGTCA
30961 TCGGCGCGGT CATCCCCAGC GGAACCACGC TCGCCTCGCC CAGCTCCTCC GACCTGCGCT
31021 CTCTCACCCA GGCCATGGAG AAGGCCAAGG TGCGCACCGT CTTCGCCGAC TCCTCCCAGC
31081 CCACCCGGCT CGCCGAGGTC CTGCGCCAGG AGATGGGCGG CGACGTGGAC GTCGTCTCGC
31141 TCTACTCCGA GTCGCTGACC GAGAAGGGCA AGGGCGCCGG AACCTACCTG GAGATGATGC
31201 GCGCCAACAC CTCGCCCATG GCCGAGGGCC TCACCGGCGA CTGAACGAGC TTCCCCGCGG
31261 CACGGCAGTT CGAGCGCCGG CCGCTCCACC CCACAAACCC GCGCCTGAGG GCCGGAGAGG
31321 AAACACCGAT CATGAACAAG CCCACCCGCG CCAGAGTCTT CACGGGCACG GCGCTGGTCG
31381 TGGCGGCGTC GATGGCGCTG ACCGCCTGCG GCGGCAACGG CAACGACGAC GCCCTTCCG
31441 GCAAAGAGCC CAAGGAGCAG AAGAGCAGCG AGGCCGCGG GGTGCGGAAC CCGATCGTCG
31501 CCTCGTACGA CGGGGGACTG TACGTCTCTG ACGGCGAGAC CCTGAAGCTC GCGAAGACGA
31561 TCGCACTGCC CGGCTTCAAC CGGGTCAACC CGGCGGGCGA CAACGAGCAC GTCGTCTGCT
31621 CCACGGACTC CGGCTTCCGC GTGTTTCGACG CCACCCGACA GGAGTTCACC GACGCCGAGT
31681 TCAAGGGTTC CAAGCCGGGG CACGTCGTCC GGCACGGCGG CAAGACGGTC CTGTTACCCG
31741 ACGGCACGGG AGAGGTGAAC GTCTTCGACC CCGCCGACCT GTCCGACGGG AAGAAGCCGG
31801 ACGGCCGCAC CTACACGTCC GCGAAGCCCC ACCACGGTGT CGCCATCGAA CTGGCCGGCG
31861 GAGAACTCGT CACCACCCTC GGCACCGAGG AGAAGCGCAC CGGAGCCCTC GTCCTGGACA
31921 AGGACAACAA GGAGATCGCA CGCGCCGAGA ACTGCCCCGG AGTGACGGG GAGGCCGCCG
31981 CCCAGGGCGA GGTGGCCGGC TTCGGCTGCG AGGACGGCGT CCTGCTCTAC AAGGACGGCA
32041 AGTTCACCAA GGTCCGACGC CCCGCTGACT ACGCCCGCAC CGGCAACCAG GCCGGCAGCG
32101 ACGCCTCCCC GATCCTCCTC GCGCACTACA AGACCGACCC CGACGCCGAA CTGGAACGCC
32161 CCACCCGCAT ATCCCTGATC GACACCCGTA CGGCGAAGAT GAAGCTGGTC GACCTCGGCA
32221 CCAGCTACTC CTTCCGCTCC CTCGCCCAGG GCGCGACGG CGAAGCCCTC GTGCTCGGCA
32281 CCAACGGCAC CCTCCACGTC ATCGACCCGG AGACCGGAAA GGTCGAGAAG AAGATCGACG
32341 CGGTCGGCGA CTGGACCGAG CCCCTGGACT GGCAGCAGCC CAGGCCACC CTGTTCTGTC
32401 GGGACCACAC GCGGTACGTC TCCGAACCGG GCAAGCGCCA ACTCCACTCC ATCGACCTGG
32461 AATCGGGGAA GAAGCTGGCA TCCGTACCCC TGCCGAAGGG CACCAACGAA CTGTCCGGCA
32521 CGGTCGCCGG TCACTGACCT GTCCCGTTCC CTCTTTTCTT CCGGCCCCGA GGAGCGCAAC
32581 GCCTGCCCGA TTCGTGTTCC GGCAGGCGTT GCTGTGCTCG GAGCCTGCAA CCTTGACGAC
32641 CCTGCCGAGG AGAACCGTTT CACCACGGAG GCCTGGGGTG CGCAGATGGA ACTGTGCGCG

32701 CTCCACTCCA GGGACCGTGA CGCCACCGTC AAGACCTGTG CCGCCGGCCG CCCGAAACGC
 32761 AAGCCGTCTG ACGGCTTCCT GGGCCGTCCT ACAGCCGCCG AGGAGCTCGC CGCGGTACAG
 32821 AGCTGCGGCG GCGGTGCCTG CGCCGCCACC ACACGATCGC GAGCGTGAAG GCGGCCGCAA
 32881 CGCCCAGCAG GGGCCACAGG ATGGTGGAGA GCACGCTCTC GGCTTCGCGC AGGGAGGTCTG
 32941 AGACCACTGT TCCCGCGGAC ACGTAGAGCG CGGACCACAT CGCGGCTCCG GCGAGGGAGG
 33001 CGGGCAGGAA GCGGAGGTAG CGCACGGAGC CGACGCCGGC GGTCGCGGGG GTGAGGGTGC
 33061 GTACCACGGG CAAAAGGCGG GTCAGGAAGA CGGCGCGCGC CCCGTACCGG TGGCAGAGCT
 33121 CTTGCGCGCG GTCCCACTGG TGCTGCCCAA TCCGCCGTAC CAGGCGCGTC TCCCGCATCC
 33181 GCTGCCCGTA GCGGATGCCG AGGAAGTAGC CGATGTGGTC GCCGGCCGAG CTGCTGAGTG
 33241 TGACGACGAG GAAGAGGGCC AACAGCGGGC GTGTCCCTCT CGTTCCGGCG CTCAGGGCCA
 33301 GTACCGCGAC CTCGCCGGGG ACGGCCATGC CGGCCCAAG GCCGGATTCC GCGAACCGCA
 33361 ATACGGAGGC CAGCGCGAAT CTGGTGACCG GGTTCATGTC CGACACCGCT GTCAGTACAT
 33421 CGTTCATCCA CGACACGGCA GCGCGCTCT GTCTCTCTCT GTTCGTGGAG CCTTCCCGAC
 33481 GCGGCCACGG GGATTCCCGC GCCCTTCTTC CGAGAACACA CCGAAGAGAA CAGCGGAACG
 33541 ACTTCCCGGC GTCACCGGAC GCATACCGGG GCGGCCGGTG GGAGCGCCTG AAAAAGAACG
 33601 AAGGGACACC AACCTACCAG GGAACCGCTG GACGACTCCT CCTTCCCGGC CACGACCACC
 33661 CCGCGACGGA CCGCGCAGAC CGCCCCCGGC AACCATTCCC CTTCACCCAC CCCGTCCGCC
 33721 GACGGAGCAC GGGGGCTCGC CGTACAGATC CGGGCCTCGT TGATCCACTG GGTGAGAACG
 33781 GCGGGGCCGG CCGCGGCCGC GAGGGCGGCC CGGTAGTGAG ACAGACGCTT CTCGCCCTTT
 33841 CTCACCGCCC GCCGGGCTTG CTCGACCTCC GGGGCGCGGC CATCGGATCG GGCAGCCGCG
 33901 TGCGTCAGGG CGGTGAGGGT GGCGGTGAGA CGTTCGGCG CGAAGGCACG GCGATCCAC
 33961 TGGTCGAGTG CCGGGCAGAT CATGTCTCC CGCAGGCACA TCATGTCCTC CCGCGGGCAC
 34021 ACGGTGCGGG GGTGACCGAG TCCGGGGTGG AGGGCCTTGT TCCTGGGACC CGTCTCTGAC
 34081 CGTGTACGGG CGTCCGAGGT CGGCTCAGGC GATCGCGGTC AACTACCCCG TGGGCTACAG
 34141 TGCGTTGACT GCGGGCAGTG CACACGCCCA CCGGCACCGA CGACGCGGAG AAGCATGGGC
 34201 GGGAGCGCGA TCAGGACCCG GCAGCTGACC AAGCACTTCG GTGCGGTGCA GGCCTGGTTC
 34261 GCGGTGGATC TGGAGGTGCC CGCGGGGAGC GTGCTGGGGC TCCTGGGACA CAACGGTGCC
 34321 GGGAAAGACCA CGCTGATCCA GATCCTCTCG ACGGTGCTCC CCGCGTCCGG TGGGTCCGCC
 34381 GAGGTCGCGG GCTTCGACAT CGTGCGCGAT GCGCGACGGG TACGCGCCTG TATCGGGGTG
 34441 ACGGGGCGAG TCGCTGCCCT GGACGAGCAT CTGTCCGGGC TCGCCAATCT GGTGCTGATC
 34501 TCCCGGCTGC TGGGTGCCCG GCGGAGGGAG GCCAGACGCC GGGCGGCCGA ACTGGTCGAA
 34561 CAATTCCGTC TCACCGAGGC AGCGGACAGA CCGATGCGGA CCTACTCCGG CGGAATGCGG
 34621 CGGCGCATCG ACCTGGCGGC GAGTCTGGTG GCCAGGCCCT CGGTGCTGTT CCTCGACGAG
 34681 CCCACCACCG GGCTGGACCC GGTGAGCCGC ACCGCACTCT GGGAGACGGT GGAAGGGCTG
 34741 GTCGCGGAGG GCACGACGGT TCTGCTGACC ACCAGTACC TCGACAGGGT CGACCGGCTG
 34801 GCGGACCGGA TAGCGGTGCT GTCGTCCGGC CACGTGGTGA CGGTGCGCAC GGCGGCGGAG
 34861 CTCAAGGCGG CCGGCACCCG GTCCGTCCGC CTGACCTTCG GGTCCGCGGC GGATCTGGAG
 34921 AGCGCGGAAG GAGCGCTGCG CCTGGAGGGC CTCGGCCTCA CAACGGATCC GGTGTCCCGG
 34981 ACGGTGTCAC TGCCGCTGGC GGCAACGGCC GAGCTGGCCG GGATCTTCCG GATTCTCGGC
 35041 GCGGCGGGCG TGGAGCTCGC CGAACTGGCG CTCAAGGAGC CCACGCTGGA CGACGTGTAT
 35101 CTGAGCCTGG CGGAGAGCTG GGAGACCACG AGCGGGGGAA CCGTCCGGTG CTGACCACAC
 35161 GACGTACGGG TCCGGGGACC TCGCCGGTGG CGGACGGGCC CCGGTGGCGC GCGGGGGGTG
 35221 CGGGGATCGG CACCCAGTTC CCGGTGCTGA CCGGCCGGCA GTTCCGGATC ATCTACGGGG
 35281 ACCGGCGGAT CGCGCTGTTC AGCCTGCTCC AGCCGATCAT CATGCTCATG CTGTTCACTG
 35341 AGGTGCTGGG CCGCATGGCC AATCCGGAGA TCTTCCCGCC GGGTGTGCGC TACCTCGACT
 35401 ACCTGGTGCC GGCTCTGCTG CTGACGACCG GGATCGGTTT CGCGCAGGGC GCGGGCTGG
 35461 GTCTCGTCAG GGACATGGAG TCCGGGATGA TGGTCCGGCT GCGGGTGATG CCGGTACGGC
 35521 TGCCGCTGGT CCTGGTGGCC CCGTCGCTGG CCGATCTGGC GCGGGTCGCC CTGCGCTCG
 35581 TGGCGTTGCT CGCCTGTGCG ATGGGGCCCG TGGGCTACCG GCCGGCCGGG GCGGTGTCCG
 35641 GGATCGTCGG CGCGACGCTG CTCGCGTTGC TCGTCGCGTG GTCGCTGATC TGGGTGTTCC
 35701 TGGCCCTCGC CGCGTGGCTG CGGAGCATCG AGGTGCTGTC CAGCATCGGG TTCCTCGTCA
 35761 CCTTCCCCCT GATGTTTCGG TCGAGTGCCT TCGTCCCGCT CGACATTCTG CCGGGATGGC
 35821 TCAGGGTCAT CGCGACGGTC AATCCCTCA CGTACGCGGT GGAGGCGTCC CGCGATCTGG
 35881 CGCTGGACCA CAGCGCGCTG GCGCGGCGC TCGCGGCCGT CGGCACCAGT CTTGCGCTCT
 35941 TGGCGGTGAC CGGTCTGCTG GCGGTACGCG GGCTGCGGCG CCCGCCGGGT GCGGGCGGCC
 36001 CGCACCGGAC GCCCTGACCC CTCCCCACCA CCTGCCAGT GTGACGTTTG CGCAGATGAG
 36061 AACGTGCGTA AACGCCGCAT ACGCAAAGAT CGTCCCTGCC GGGACCCATT GACGTTTCGA
 36121 GGGGCGTGGA ACATACTGGC GATCAAGTCG CACAGGAACC AACAGGCACA CCAACCACAG
 36181 GCGTTACAGG GGGGGTTGGT GTTTCGTCCA TATCAAGTGG TTTGGTCCGC CGAAGCGGTT
 36241 GGACCTCACA TGACGGCAAC AGGGCATTCG CACATGCCTG ATGACGGGAC GGCACACCTC
 36301 ACGCAGCGGC GACCGGTGCG AAGCCGGACG CGGAATGACT CCCTGCCTTA CAGGTATGCG

36361 AGCGCGGATG CGTCGTTCGA CCGGAGTCAG GAGGGGGAGT GCCTGCCGTG AGTGAGAGCC
 36421 GCTGTGCCGG GCAGGGCCTG GTGGGGGCAC TGGCGACCTG GGCACGGACA CGTGCCCGGG
 36481 AGACTGCCGT GGTTCCTCGTA CGGGACACCG GAACCACCGA CGACACGGCG TCGGTGGACT
 36541 ACGGACAGCT GGACGAGTGG GCCAGAAGCA TCGCGGTGAC CCTCCGACAG CAACTCGCGC
 36601 CGGGGGGACG GGCACCTCTG CTGCTGCCGT CCGGCCCCGA GTTCACGGCC GCGTACCTCG
 36661 GCTGCCTGTA CGCGGGTCTG GCCGCCGTAC CGGCGCCGCT GCGCGGGGGG CGCCACTTCG
 36721 AACGCCGCCG TGTCGCGGCC ATCGCCGCCG ACAGCGGAGC CGGCGTGGTG CTGACCGTCG
 36781 CGGGTGAGAC CGCCTCCGTC CACGACTGGC TGACCGAGAC CACGGCCCCG GCTACTCGCG
 36841 TCGTGCCCGT GGACGACCGG GCGGCCGTCG GCGACCCGGC GCAGTGGGAC GACCCGGGCG
 36901 TCGCGCCCCG CGACGTGGCT CTCATCCAGT ACACCTCGGG CTCGACCGGC AACCCCAAGG
 36961 CGGTGGTCGT GACCCACGCC AACCTGCTGG CGAACGCGCG GAATCTCGCC GAGGCCTGCG
 37021 AGCTGACCGC CGCCACTCCC ATGGGCGGGT GGCTGCCCCAT GTACCACGAC ATGGGGCTCC
 37081 TGGGCACGCT GACACCGGCC CTGTACCTCG GCACCACGTG CGTGCTGATG AGCTCCACGG
 37141 CATTTCATCA ACGGCCGCAC CTGTGGCTAC GGACCATCGA CCGGTTTCGGC CTGGTCTGGT
 37201 CGTCGGCTCC CGACTTCGCG TACGACATGT GTCTGAAGCG CGTCACCGAC GAGCAGATCG
 37261 CCGGGCTGGA CCTGTCCCGC TGGCGGTGGG CCGGCAACGG CGCGGAGCCC ATCCGGGCGAG
 37321 CCACCGTACG GGCCTTCGGC GAACGGTTTCG CCCGGTACGG CCTGCGCCCC GAGGCGCTCA
 37381 CCGCCGGCTA CCGGCTGGCC GAGGCCACCC TGTTCGTGTC GAGGTCGCGAG GGGCTGCACA
 37441 CGGCACGAGT CGCCACCGCC GCCCTCGAAC GCCACGAATT CCGCCTCGCC GTACCCGGCG
 37501 AGGCAGCCCC GGAGATCGTC AGCTGCGGTC CCGTCGGCCA CTTCGCGGCC CGCATCGTCG
 37561 AACCCGGCGG GCACCGTGTT CTGCCGCCCG GCCAGGTCGG CGAGCTGGTC CTCCAGGGAG
 37621 CCGCCGTCTG CGCCGGCTAC TGGCAGGCCA AGGAGGAGAC CGAGCAGACC TTCGGCTCA
 37681 CCCTCGACGG CGAGGACGGT CACTGGCTGC GCACCGGCGA TCTCGCCGCC CTGCACGAAG
 37741 GGAATCTCCA CATCACCGGC CGCTGCAAG AGGCCCTGGT GATACGAGGA CGCAATCTGT
 37801 ACCCGCAGGA CATCGAGCAC GAACTCCGCC TGCAACACCC GGAATTGAG AGCGTCGGCG
 37861 CGCGCTTCAC CGTCCCGGCG GCACCTGGCA CGCCGGGCTT GATGGTGGTC CACGAAGTCC
 37921 GCACCCCGGT CCGCGCCGAC GACCACCGCG CCCTGGTCAG CGCCCTGCGG GGGACGATCA
 37981 ACCCGCAATT CGGACTCGAC GCGCAGGGCA TCGCCCTGGT GAGCCGCGGC ACCGTACTGC
 38041 GTACCACCAG CGGCAAGGTC CGCCGGGGCG CCATGCGTGA CCTCTGCCTC CGCGGGGAGC
 38101 TGAACATCGT CCACGCGGAC AAGGGCTGGC ACGCCATCGC CGGCACGGCC GGAGAGGACA
 38161 TCGCCCCCAC TGACCACGCT CCACATCCGC ACCCCGCGTA ATCGCCGGAG GCGGGCCCTG
 38221 CCCTGGAACG GGCACCGCGG TGCCCCCGCA CAGCGAGGAG TAGCTCCACA TGAACCCGCC
 38281 CGAAGCGGTC AGCACGCCCC GCGAGGTCAC CGCGTGGATC ACCGGACAGA TCGCCGAGTT
 38341 CGTGAACGAG ACACCCGACC GGATCGCCGG TGACGCACCC CTGACCGACC ATGGCCTCGA
 38401 CTCCGTCTCC GGAGTTGCCC TCTGCGCGCA GGTGAGGAC CGTACGGGA TCGAGGTGCA
 38461 CCGGAGCTG CTGTGGAGCG TCCCCACACT CAACGAGTTC GTCCAGGCAC TGATGCCCCA
 38521 GTTGGCCGAC CGCACCTGAG GGGACTCCG AGAGATGGAC ATGCAGTCGC AGCGCCTCGG
 38581 CGTCACCGCC GCCAACAGA GCGTCTGGCT CGCCGGCCAG CTGGCGGACG ACCACCGCCT
 38641 GTACCACTGT GCGGCGTACC TGTCACCTAC CGGGTCCATC GACCCGCGGA CACTCGGCAC
 38701 GCGGTTCCGG CGGACCCTCG ACGAGACCGA GCGGCTGCGT ACCCGGTTTCG TACCGCAGGA
 38761 CGGGGAACTG CTGCAGATCC TCGAACCCGG TGCCGGACAG CTCCTGCTGG AAGCCGACTT
 38821 CTCCGGCGAC CCGGACCCCG AGCGGGCGGC ACACGACTGG ATGCACGCGG CGCTCGCCGC
 38881 ACCGGTCCGC CTCGACCGCG CCGGGACCGC CACCCACGCC CTGCTACCC TCGGCCCGTC
 38941 CCGCCACCTG CTGTACTTCG GCTACCACCA CATCGCGCTC GACGGCTACG GTGCCCTGCT
 39001 CCACCTGCGC CGCCTCGCCC ACGTCTACAC CGCCCTCAGC AACGGGGACG ACCCCGGCCC
 39061 CTGCCCCGTT GGGCCCCGCG CCGGTGTCTT CACGGAGGAG GCGGCCTACC GTGACTCCGA
 39121 CAACCATCGG CGCGACGGGG AATTCTGGAC CCGGTCCCTC GCGGTTGCGG ACGAGGCCCC
 39181 CGGGCTGAGC GAGCGGGAGG CCGGCGCTCT CGCCGTCCCG CTGCGCCGCA CCGTGGAGCT
 39241 GTCCGGCGAA CGGACGGAGA AGCTGGCCGC CTCGGCCGCG GCCACTGGAG CTCGCTGGTC
 39301 GTCATCGGTC GTCGCCGCCA CCGCCGCGTT CGTACGCCGC CACGCTGCCG CCGACGACAC
 39361 CGTCATCGGC CTGCCCCGTA CCGCCCGGCT CACCGGGCCG GCGCTGCGTA CCGCGTGCAT
 39421 GCTCGCCAAC GACGTGCCGC TGCGCCTCGA CGCCCGGCTC GATGCCCCGT TCGCCGCGCT
 39481 CCTTGCCGAC ACCACCCGCG CCGTCGGCAC GCTGGCGCGC CACCAGCGGT TCCGCGGGGA
 39541 AGAACTCCAC CGGAACCTGG GGGGCGTCCG CCGCACCGCG GGCCTGGCGC GGGTCACCGT
 39601 CAACGTCTCG GCGTATGTCG ACAACATCCG GTTCGGCGAC TGCCGGGCGG TGGTCCACGA
 39661 GTTGTCCTCG GGACCGGTCC GCGACTTCCA CATCAACTCC TACGGCACCC CCGGCACCCC
 39721 CGACGGCGTC CAGCTGGTCT TCAGCGGTAA CCGCGCCCTG TACACGGCCA CCGATCTGGC
 39781 CGACCACCAG GAGCGGTTCC TGCGCTTCCT CGACGCTGTG ACCGCCGACC CGGACCTGCC
 39841 GACCGGAAGA CACCGCCTCC TGTCGCCGGG CACCCGCGCC CGGCTGCTCG ACGACTCCCG
 39901 CGGCACGGAA CGCCCCGTAC CGCGTGCCAC CTTGCCGGAA CTCTTCGCCG AACAGGCCCC
 39961 GCGCACCCCC GACGCGCCCC CCGTCCAGCA CGACGGCACC GTCCTCACCT ACCGCGACCT

40021 GCACCGGAGT GTCGAACGGG CGGCCGGACG GCTGGCCGGC CTCGGCCTGC GTACCGAGGA
 40081 CGTCGTCGCC CTCGCCCTCC CCAAGTCCGC CGAGAGCGTC GCGATCCTGC TCGGCATCCA
 40141 GCGGGCCGGC GCCGCCCTACG TGCCGTGGA CCCACCCAT CCGGCCGAGC GGCTGGCCCCG
 40201 TGTACTCGAC GACACCCGAC CCCGGTACCT CGTCACCACC GGACACATCG ACGGCCTGTC
 40261 CCACCCACAG CCGCAGTTGG CCGCCGCCGA CCTCCTCCGT GAGGGCGGCC CAGAGCCCGC
 40321 CCCGGGCCGC CCGGCACCCG GCAACGCGGC GTACATCATC CAGACCTCCG GCTCCACCGG
 40381 ACGGCCGAAG GGTGTCTGTC TACTCACGA AGGGCTGGCC ACCCTCGCCG CCGACCAGAT
 40441 CCGGCGCTAC CGCACGGGAC CGGACGCCCG CGTACTGCAG TTCATCTCCC CGGGGTTCTGA
 40501 CGTCTTCGTC TCCGAAGTGA GCATGACCTT CCTGTCCGGC GGCTGCCTGG TGATACCGCC
 40561 GGACGGCCCTG ACCGGCCGTC ACCTCGCCGA CTTCTTGCC GCGGAGGCCG TCACCACCAC
 40621 ATCCCTCACC CCCGGCGCAC TCGCCACCAT GCCCGCCACA GATCTCCCGC ACCTGCGGAC
 40681 TCTGATCGTC GCGGAGAGG TCTGCCGCC GGAGATCTTC GACCAGTGGG GCCGGGGCCG
 40741 GGACATCGTC AACGCGTACG GGCCACCGA GACAACCGTC GAGGCGACCG CCTGGCACCG
 40801 TGACGGTGCC ACCCACGGCC CCGTCCCGCT CGGCCGCCCC ACCCTCAACC GCGCGGGCTA
 40861 CGTCTCGAC CCGGCGCTCG AACCCTGCCC CGACGGGACG ACCGGCGAAC TGTACCTGGC
 40921 CCGCGAGGGC CTCGCCCGGG GCTACGTCGC TGCTCCCGGG CCCACCGCCG ACGTTTTCTG
 40981 CGCCGACCCG TTCGGCCCCG CCGGCAGCCG CATGTACCGC ACCGGTGACC TGGTGCGGCG
 41041 GCGCTCCGGC GGCATGCTGG AATTCTGTCG ACGAGCCGAC GGACAGGTCA AACTCCGCGG
 41101 CTTCCGCATC GAACTCGGCG AGGTCCAGGC CGCGCTCACC GCTCTCCCCG GGGTACGTCA
 41161 GGCCGGCGTC CTGATCCGCG AGGACGCCCC CGGGGACCCC CGGCTCGTCG GGTACATCGT
 41221 GCGCGCGCCC GCGCGCCGAAC CGGACGCCCG TGAGTCCGT GCGGCCCTGG CCCGTACCTT
 41281 CCCGCCCCAC ATGGTGCCCT GGGCGCTCGT CCCCTCCCC GCACTGCCGC TGACGTCCAA
 41341 CGGCAAATG GACAGGGCG CCCTTCCCGT CCGCGCCGCC CGCGCCGGCG GATCCGGGCA
 41401 ACGCCCGGTC ACCCCACAGG AGAAGACACT CTGCGCCCTG TTCGCCGACG TCCTCGGCGT
 41461 AACGGAGGTC GCCACGGACG ACGTGTCTT CGAGCTCGGC GGCCACTCCC TCAACGGCAC
 41521 CCGGCTGCTC GCGCGGATCA GGACCGAGTT CGGCACCGAC CTCACCCTCC GCGACCTGTT
 41581 CGCCTTCCCC ACCGTGCGCG GCCTTCTCCC GCTCCTGGAC GACAACGGAC GGCAGCACAC
 41641 CACCCGCGCG CTGCCTCCGC GCGCGGAGCG CCTCCCCCTG TCGCACGCGC AGCAGCGACT
 41701 GTGGTTCTCT GACCAGGTCT AAGGCCCCAG CCGCGCGTAC AACATCCCCA CCGCCGTCCG
 41761 GCTCGAAGGC CCGCTCGACA TCCCGGCCCT CGCTGTCGCC CTGCAGGACG TCACCAACCG
 41821 CCACGAGCCC TTGCGTACTC TCCTCGCCGA GGACTCCGAA GGCCCCCACC AGGTTCATCT
 41881 GCGCCCCGAG GCGGCCCGCC CCGAAGTAC CCACAGCACC GTCGCGCCCG GCGATCTCGC
 41941 CGCAGCCCTC GCGGAAGCCG CACGCCGCC CTTCGACCTC GCGGTGAGA TCCCACTCAA
 42001 AGCCACCTG TTCGGCTGCG GCGCGGACGA CCACACCCTG CTGCTCCTCG TCCACCACAC
 42061 CGCCGGCGAC GGAGCCTCCG TCGAGGTCCT GGTACGCGAT CTCGCCACG CCGTACGGCGC
 42121 CCGCCGCGCC GCGGACGCC CGCATTCGA CCGGCTGCC CTGCAGTACG CCGTACCGAC
 42181 CCTGCGCCGA CGGCACCTGC TGGACGATCC GTCGGACAGC ACACAGCTCG ACCACTGGCG
 42241 GACGCGCCTG GCGGCGCTGC CCGAGCAGCT CGAAGTGCC ACCGACCACA CCGGCCCCG
 42301 CGTTCCCAAC CGCCGGGGCG AGGCGATCGC CTTACCCGTG CCGGAGCACA CGCACCACAC
 42361 GCTGCGGGCC ATGGCCAGG CCCACGGCGT CACCGTGTTC ATGGTCATGC AGGCCGCGCT
 42421 CGCCGCCCTG CTGTGCGGGC ACGGCGCGGG CCACGACATC CCCCTCGGAA CACCCGTGCG
 42481 GGGCGCTCC GACGACGGCA CGGAAGACCT CGTCGGGTTC TTCGTCAACA CGCTCGTACT
 42541 GCGCAACGAC GTCTCCGGGG ACCCGACGTT CGCGGAATC GTGTGCGGGG TGCGGGCCCG
 42601 CAACCTGGAC GCGTACGCCT ACCAGGACGT TCCCTTCGAG CGTCTCGTCG ACGTACTCAA
 42661 ACCGGAGCGG TCCCTGTCTT GGCACCCGCT CTTCCAGATC ATGATCGCGT ACAACGGCCC
 42721 GGCGACGAAC GACACCGCCG ACGGTCCCG CTTGCGGGC CTCACCAGCC GCGTCCATGC
 42781 CGTCCACACC GGCATGTCCA AGTTCGACCT GTCGTTCTTC CTCACCAGC ACGCGGACGG
 42841 CCTCGGCATC GACGGCGCTC TCGAGTTCAG CACCGATCTC TTCACGCGGA TCACCGCGGA
 42901 GCGCCTGGTC CAGCGCTACC TCACCGTCTT GGAGCAAGCC GCGGAGCAC CCGACCGCCC
 42961 CATCAGTTTC TACGAATCC TCGGCGACGA CGAAGCGCA CTCCTCGCCC AATGGAACGA
 43021 CACCGCCAC CCCACCCCC CAGGCACGGT GCTCGATCTC CTCGAAAGCC GTGCGGCGCG
 43081 GACCCCCGAC CCGCGGCCCG TCGTCGAGAA CGACCACGTC CTCACCTACG CCGACCTGCA
 43141 CACCGGGGCC AACCGGCTCG CCCGCCACCT GATCACCGCC CACGGCGTCG GTCCCGAACG
 43201 TCTCGTCGCC GTCGCCCTGC CCGGTCCGC CGAGCTGCTG GTGGCACTTC TCGCGTCTCT
 43261 CAAGACCGGA GCCGCCTACG TCCCTCTCGA CCTACCCAC CCGCGGAGC GCACCGCCGT
 43321 CGTCTCGAC GACTGCCGGC CGGCCGTGAT CCTACCGAC GCGGTGCGG CCCGTGAACT
 43381 GCGCGGCGC GACATCCAC AGCTCCGCT CGACGAACCC GAGGTCCACG CCGCGATCGC
 43441 GGAACAACCG GGGGGTCCGG TCACCGACCG GGACCGCACG TGCGTCACTC CGGTACGCGG
 43501 CGAGCACGTG GCATACGTGA TCTACACATC CGGCTCCACG GGCCGGCCCA AGGGTGTGGC
 43561 GGTGGAACAC CGTTCACTGG CCGACTTCGT GCGGTACTCC GTGACCGCGT ACCCGGAGC
 43621 CTTGACGTC ACCCTGCTGC ACAGCCCCGT GACCTTCGAC CTCACCGTGA CCTCGCTGTT

43681 CCCGCCACTG GTCGTCGGTG GCGCCATCCA TGTCGCGGAC CTGACCGAGG CGTGCCCACC
43741 GAGCCTGGCC GCGGCGGGCG GGCCGACGTT CGTCAAGGCC ACACCGAGCC ATCTGCCACT
43801 GCTCACGCAC GAGGCGACAT GGGCCGCGTC CGCGAAGGTG CTGCTCGTCG GGGGCGAGCA
43861 GTTGCTGGGA AGGGAGCTGG ACAAGTGGCG GGCCGGGTGCG CCGGAGGCCG TCGTCTTCAA
43921 CGACTACGGC CCCACCGAGG CCACGGTCAA CTGCGTGGAC TTCCGTATCG ATCCGGGACA
43981 ACCGATCGGT GCGGGGCGCG TGGCGATCGG CCGCCCGTTG CGGAACACGC GGGTGTTCTG
44041 GCTCGACGGT GGGTTGCGGG CGGTGCCCGT CGGTGTGGTC GGTGAGCTCC ATGTGGCGGG
44101 CGAGGGGCTG GCGCGGGGTT ATCTCGGGCA GCCGGGTCTG ACGGCGGAGC GGTTCTGTGGC
44161 GTGTCCGTTC GGTGATGCCG GGGAGCGGAT GTACCGCACG GGTGACCTGG TCGGGTGGCG
44221 TCGGATGGG ATGCTGGAGT TCGTCGGCCG GGTGACGAT CAGGTCAAGG TCGGGGGTTT
44281 CCGGATCGAG CTGGGCGAGG TGGAGGCCG TGTCGCGGCC TGCCCGGGTG TGGACCGCTC
44341 CGTGGTGGTG GTACGGGAGG ACCGACCGGG AGACCGCCGG CTGGTGGCGT ATGTGACGGC
44401 CGCCGGTGAC GAGGCGGAGG GGCTGGCACC GCTGATCGTG GAGACGGCCG CGGGCCGTCT
44461 GCCCGGGTAC ATGGTGCCGT CGGCCGTGGT CGTACTGGAC GAGATTCCCC TGACGCCGAA
44521 CGGCAAGGTG GACCGTGCCG CGCTGCCCGC GCCGCGCGTC GCCCCGGCCG CGGAGTTCCG
44581 CGTCAACCGA TCACCCCGTG AAGAGGCTCT GTGCGCCCTG TTCGCGGAAG TGCTGGGCGT
44641 GGAACGGGTC GCGGTGGACG ACGGGTTCTT CGACCTCGGC GGAGACAGCA TTCTGTCCAT
44701 TCAACTGGTG GCGCGGGCGC GCCGGGCGGG TCTGGAGGTG TCGGTGCGGG ACGTTTTCGA
44761 GCACCGCACC GTACGGGCGC TGGCCGGTGT GGTGCGGGAG TCCGGAGGCG TCGTGGCCG
44821 CGTCGTGGAC TCCGGTGTGT GTGCGGTGTA GCGGTGGCCG GTGGTGGAGT GGTGCGGGA
44881 GCGTGGTGGC GGTGGGCTCG GCGGTGCGGT CAGGGCCTTC AACCAGTCCG TCGTGGTCGC
44941 CACACCGGCC GGTATCACCT GGGACGAAC GCGGACGGTC CTGGACGCGG TACCGGAACG
45001 CCACGACGCC TGGCGGCTAC GGGTAGTGGA TTCCGGTGAC GGCGCCTGGT CCCTGCGCGT
45061 CGACGCGCCC GCCCCCGGCG GTGAGCCCGA CTGGATCACC CGGCACGGCA TGGCCAGCGC
45121 CGACCTGGAG GAGCAGGTGA ACGCCGTGCG GGCCGCGGCC GTGGAGGCCC GGAGCCGGCT
45181 CGATCCACTG ACCGGACGGA TGGTCCGCGC GGTATGGCTG GACCGTGGAC CCGACCGCCG
45241 GGGAGTCCTG GTCCTGGTGG CGCACCACCT GGTGTCGAC GCGCTCTCCT GGCGCATCGT
45301 CCTCGGCGAC CTCGGCGAAG CCTGGACACA GGCACGCGCT GGCGGGCATG TGCGGTGGA
45361 CACGTCGCGC ACATCGCTGC GCGGTGGGC GCGGCGCGTG GCGGAACAGG GCCGCCACGG
45421 CGCCCGCGCC ACCGAAGCAA ACCTGTGGGC ACAGATGGTC CACGGCTCGG ACCCTCTGGT
45481 CGCCCCACGC GCGGTGGACC CTTCGGTGGA CGTCTTCGGC GTGGTGGAGT CCGTGGGTTT
45541 ACGGGCGTCG GTGGGGGTGT CGCGTGCCCT GCTGACGGAG GTCCCGTCGG TCCTGGGTGT
45601 GGGCGTGACG GAAGTGCTGC TGGCGGCATT CGGCCTGGCA GTGACGCGCT GCGCGGCGCC
45661 CGGCGGAAGC GTCGTCGTGG ACGTCGAGG TCACGGCCGC AACGAGACG CGCTACCCGG
45721 CGCGGACCTC TCCCGCACCG TGGGTGGTT CACCAGCATC TACCCCGTCC GCCTCCCCCT
45781 CGAGCCGGCG GCCTGGGACG AGATACGCGC CGGCGGTCCC GCGTCGGAC GCACCGTCCG
45841 CGAGATCAAG GAATGCCTCC GCACCCTGCC CGACCAGGGC CTGGGCTACG GCATCCTGCG
45901 CTACCTCGAC CCCGAAAACG GACCCGCCCT CGCCAGCAC CCCACCCCGC ACTTCGGCTT
45961 CAACTACCTC GGACGGGTCT CCGTCTCGGC GGACGCTGCC TCACTGGACG AAGGCGACGC
46021 CCATGCCGAC GGGCTCGGCG GCCTCGTCGG CGGCAGGGCA GCGGCGGACT CCGACGAGGA
46081 ACAGTGGGCC GACTGGGTTC CCGTGTGCGG TCCGTTTCGG GTGGGCGCGG GTCAGGACCC
46141 CGTTCTGCCG GTGGCCACG CCGTGGAGTT CAACGCGATC ACCCTGGACA CACCCGACGG
46201 CCCCCGCTC AGCGTGACAT GGTGTCGGC GACGACACTG CTGTCCGAAT CCCGATACG
46261 AGAACTCGCC CGCTTCTGGG ACGAAGCCCT CGAAGGGCTG GTCGCACACG CCCGCCGTCC
46321 CGACGCGGGC GGAAGTACCC CCTCGGACCT GCCGCTGGTC GCCCTCGACC ACGCGGAAC
46381 GGAGGCCCTG CAGGCCGACG TCACCGGTGG CGTGACGAC ATCCTGCCCG TATCACCCTG
46441 TCAGGAAGGA CTGCTCTTCC ACAGCTCCTT CGCCGCCGAC GGGGTGACG TCTACGTGGG
46501 ACAACTCACG TTCGACCTGA CCGGACCACT CGACGCCGAC CACTGACAG TCTGGTTCGA
46561 AAGCCTGGTG ACACGCCACG ACGTCCTGCG CACCGGCTAC CGCCAGGCAC AGTCCGGCGA
46621 ATGGATCGCC GTCGTGGCAC GACAAGTCCA CACCCCTGG CAGTACATCC ACACACTCGA
46681 CACGGACGCC GACACCCTCA CAAACGACGA GCGCTGGCGG CCGTTCGACA TGACGAGGG
46741 CCCACTCGCA CGATTACCC TCGCACGCAT CAACGACACC CACTTCCGCT TCATCGTCAC
46801 GTACCACCAC GTCATCCTCG ACGGTGGTTC CGTGGCGGTT CTCATACGCG AACTCTTCAC
46861 CACCTATCGC GACACCGCCC TCGGCCGCG GCCGGAGGTT CCGTACTCCC CACCGCGCCG
46921 TGACTTCATG GCGTGGCTCG CCGAACGCGA CCAGACCGCT GCGGGACAGG CATGGCGTTC
46981 CGCGTGGCC GGAATCGCGG AGCCACAGT GCTCGCCCTC GGAACGGAGG GCAGTGGGGT
47041 GATTCCCGAA GTCCTTGAGG AAGAGATCAG CGAGGAACG ACCTCGGAAC TGGTGGCGTG
47101 GGCGCGTGGG CGTGGTGTGA CCGTCGCGTC GGTGGTGCAG GCGGCCTGGG CGTTGGTGTG
47161 GGGGCGGCTG GTGGGCCGGG ACGACGTGGT GTTCGGCCTG ACCGTGTGCG GCCGGCCCGC
47221 CGAAGTGGCG GGTGTGGAGG ACATGGTGGT TCTGTTCTG AACACCATTC CGTTGCGGGC
47281 CCGGATGGAC CCGGCGGAGT CACTGGGCGC CTTGCTGGAG CCGCTGCAGC GGAACAGAC

47341 GGAAGTCTC GAGCACCAGC ACGTCCGGCT GGCCGAGGTC CAGCGCTGGG CCGGACACAA
 47401 GGAAGTCTTC GACGTCGGAA TGGTCTTCGA GAACTACCCG ATGGATTCCC TGCTGCAGGA
 47461 TTCACTGTTT CACGGCAGTG GCCTGCAGAT CGACGGAATA CAGGGTGCCG ATGCGACGCA
 47521 TTTCGCTTTT AACCTGGCAG TGGTTCCTCT TCCCGCCATG CGATTCCGGC TCGGCTATCG
 47581 GCCGGACGTG TTTGACGCGG GTCGGGTGCG TGAGCTGTGG GGTGTGGATCG TCCGGGCCTT
 47641 GGAGTGCGTG GTCTGCGAGC GTGATGTGCC GGTGTCCGGT GTCGATGTGC TGGGTGCCCG
 47701 TGAGCGGGAG ACGCTGCTGG GCTGGGGTGC GGGCGCGGAA CCCGGCGTGC GTGCGCTGCC
 47761 GGGTGCGGGT GCGGGTGCGG GTGCGGGGCT GGTGCGGGTG TTCGAGGAGC GGGTGCGGAC
 47821 CGACCCGGAC GCGGTGGCCG TGCGCGGGCG GGGAGTGGAA TGGAGTTACG CGGAGCTGAA
 47881 CGCGCGGGCG AATGCGGTGG CCCGGTGGCT GATCGGCCGG GCGGTGGGAC CCGAGCGCGG
 47941 TGTCGGGGTG GTGATGGACC GCGGCCGGGA CGTGGTGGCC ATGCTCCTCG CCGTGCCTAA
 48001 AAGCGGCGGC TTCTACCTGC CCGTCGACCC GCAATGGCCC ACCGAACGCA TCGACTGGGT
 48061 ACTCGCCGAC GCCGGCATCG ACCTGGCCGT CGTGGGCGAG AACCTGGCCG CTGCGGTGCA
 48121 GGCCGTCCGC GACTGCGAGG TGGTCGACTA CGCGCAGATC GCGCGCGAAA CACGGCTGAA
 48181 CGAGCAGGCG GCCACCGACG CCGGTGATGT GACGACGGG GAGCGCGTGT CCGCTCTGCT
 48241 GTCCGGGCAT CCGCTGTATG TCATCTACAC CTCCGGCTCG ACGGGCCTGC CCAAGGGCGT
 48301 GGTGGTCACC CACGCTCTCG TCGGCGCCTA TCTGCGGCGC GGCCGCAACG CCTACCGCGG
 48361 CGCCGCCGAC GGCTGGGGC ACGTGCACTC CTCCTCGCG TCCGACCTGA CCGTGACCGT
 48421 TCTGTTTACC CCCCTGGTCT CCGGCGGCTG CGTCACCTC GCGGATCTCG ACGACACCGC
 48481 CAACGGCCTG GCGGCCACCT TCCTCAAGGC CACTCCTTCC CACTGCCCC TGCTCGGCCA
 48541 ACTCGACCGG GTACTCGCCC CCGACGCCAC CTCCTCCTC GCGGCGGAAG CCTCACCGC
 48601 CGGCGCCCTG CACCATGTGC GACCCACCA CCCCCACACC ACGTCATCA ACGCCTACGG
 48661 CCCGACCGAA CTCACCGTCA ACTGCGCCGA ATACCGCATC CCCCCCGGCC ACTGCTCTCC
 48721 CGACGGCCCC GTCCCCATCG GACGCCCCCT CACCGGCCAC CACTCTTTCG TCCTCGACCC
 48781 CGCCCTCCGC CTCACACCCC CCGACACCAT CGGCGAAGTG TATGTGGCCG GTGACGGCCT
 48841 GGCGCGGGGC TATCTCGGGC GCGCGGACCT GACCGCCGAA CGCTTCGTGG CCTGCCCCCT
 48901 CCGCAGCCCC GCGGAACGCA TGTACCGCAC CGGCGACCTC GCACGTGGC GCAGCGACGG
 48961 AACACTCGAA TTCATCGGCC GTGCCGACGA CCAGGTCAAG ATCCGCGGCT TCCGCATCGA
 49021 ACTCGGCGAA GTCGAGGCGG CTGTGCGGCG GCATCCGCAC GTGGCGCGGG CCATCGCCGT
 49081 CGTACGCGAG GACCGGCCCC GCGACGAGCG CCTGGTCCGC TACGTGACAG GCAGCGACCC
 49141 GAGCGGCCTG TCCTCGGCGG TGACGGACAC CGTCGCGGCG CGCTGCCCC CGTACATGGT
 49201 GCCGTGCGCC GTCGTGCTAC TGGACAGAT CCCCCACACC CCGAACGGCA AGGTGACCGG
 49261 CGCCGCCCTC CCCGCGCCCC GGACCGCCTC CGGAACCACC TCCCGAGCAC CCGGCACAGC
 49321 CCGTGAAGAG ATCCTGTGCA CCTGTGTCG CGACGTACTC GGTCTGGATC AGGTGCGGCT
 49381 GGACGAGGAC TTCTTCGACC TCGGCGGCCA TTCCCTGCTC GCCACCCGCC TCACCTACAG
 49441 GATCCGCTCG GCCCTCGGCA TCGACCTCGG TGTCCGAGCC CTCTTCAAAG CCCCCACCGT
 49501 CGGCGGCTCG GACGAGCTGC TCCAGCAACA GACCACCAGC CTCCGGGCAC CCTGGTTCGC
 49561 GCGGGAGCGC ACCGGTTGTG AGCCGCTGTC GTTCGCGCAG CAGCGCCTGT GGTTCCTCCA
 49621 CCAGCTCGAA GGCCCCAAG CCGCGTACAA CATCCCCATG GCTCTGCGAC TCACCGGCCG
 49681 CCTGGACCTG ACCGCGCTGG AAGCGGCCCT GACGGATGTG ATCGCCCGCC ACGAAAGCCT
 49741 GCGAACGGTC ATCGCCAGG ACGATTGCGG CCGCGTGTGG CAGAACATCC TGCCACCGA
 49801 CGACACCGCG ACCACCTCA CCTCGACAC CATGCCGGTC GACGCGCACA CCCTGCAGAA
 49861 TCGGTGGGAC GAGGCCGCCC GCCATCCGTT CGACCTACC ACCGAGATCC CCCTCCGCGC
 49921 CACCGTCTTC CGCGTCACCG ACGACGAGCA CGTCCTCTTG CTGCTGCTCC ACCACATCGC
 49981 CGGCGACGGC TGGTCCATGG CCCCCCTGGC CCACGACCTG TCCGCGCCT ACACCGTCAG
 50041 ACTCGAGCAC CACGACCGC AACTGCCCCG TCTGGCCGTC CAATACGCG ACTACGCCGC
 50101 CTGGCAACGC GACGTCCTGG GCACCGAGAA CAACACATCG AGCCAACTCT CCACCAACT
 50161 CGACTACTGG TACAGCAAAC TCGAAGGCCT CCCCCCGAA CTGACCTCTC CCACCACTG
 50221 CGTCCGGCCC GCGGTGGCCT CCCACGCACT CGACCGGTC GAGTTACCG TGCCCCACGA
 50281 CGTGACCAAA GGCCTGACCG CACTCGCCCC GACCCAGGGC GCCACCGTCT TCATGGTGGT
 50341 GCAGGCGGCC CTGGCGGCC TGCTGTCCCG ACTCGGCGCC GGCACCGACA TCCCCATCGG
 50401 CACCCCCATC GCCGGCCGCA CCGACAGGC GATGGAGAAC CTGATCGGAC TCTTCGTCAA
 50461 CACCTCTGTA CTGCGACCG ACGTCTCCGG GGACCCGACC TTCGCGGAGC TCCTGGCCCCG
 50521 TGTGCGCACC ACTGCTCTCG ACGCATACGC ACACCGAGC ATCCCTTTCG AACGCTGGT
 50581 AGAAGCCATC AACCCTGAAC GATCCCTCAC CCGGCACCCC CTCTTCCAGG TCATGCTCGC
 50641 CTTCAACAAC ACGGACCGCC GATCCGCGCT CGACGCGCTC GACGCCATGC CCGGCCTTCA
 50701 CGCAGACCG GCGACGTCC TGGCTGTGAC CAGCCCCCTAC GATCTCGCGT TCTCGTTTCT
 50761 GGAGACACCC GCGAGCACGG AGATGCCCCG CATCCTGGAC TACGCAACCG ACCTGTTTCA
 50821 CCGCTCCACG GCCGAGGCCA TGACCGAAGC TCTGGTGGCG CTCCTCGCGG AGATCGCCCCG
 50881 CCGGCCCCGAG CTGTCCGTGG GCGACATCGG CATCCTGTG GCGGACGAGG TGAAGGCCCT
 50941 CAGCCCCGAG GCTCCCCCGG CAGCCGAGGA ACTTCACACC TCCACACTGC CTGAGCTGTT

```

51001 CGAGGAGCAG GTGGCGGCTC GGGGCCATGC GGTCGCGGTG GTGTGCGAAG GAGAGGAGCT
51061 GTCGTACAAG GAGTTGAACG CGCGGGCGAA TCGCCTGGCC AGGGTGCTGA TGGAGCGCGG
51121 CGCAGGCCCC GAACGGTTTCG TGGGCGTGGC ACTACCGCGT GGCCTGGACC TCATCGTGGC
51181 ACTCCTGGCC GTGACCAAAA CCGGCGCCGC ATACGTTCCG CTCGACCCCG AATACCCAC
51241 CGACCGCCTC GCGTACATGG TCACCGACGC CAACCCACC GCGGTCTGA CCTCAACGGA
51301 CGTACACATC CCCCTGATCG CCCCCGCGAT CGAGCTCGAC GACGAGGCAA TCCGCACCGA
51361 ACTCGCCGCC GCTCCCGACA CAGCCCCCTG TGTGCGGAGC GGCCCCGCC ACCCCGCTTA
51421 CGTCATCTAC ACCTCCGGCT CCACCGGTG CCGCAAGGGC TCGTCATCA GCCACGCCAA
51481 TGTCTGACGC CTGTTACACG CATGCTCCGA CAGTTTCGAC TTCGGACCG ACCACGTCTG
51541 GACGCTCTTC CACTCGTACG CCTTCGACTT CTCGGTCTGG GAGATCTGGG GCGCGCTGCT
51601 TCACGGCGGG CGGCTCGTCG TCGTGCCGTT CGAGGTGACT CGTTCTCCCG CCGAATTCTT
51661 CGCGCTGCTC GCCGAGCAGC AGGTACGCT GCTGAGCCAG ACACCGTCCG CGTTCCATCA
51721 GCTGACGGAG GCCGCCCGCC AGGAGCCGGC GCGCTGCGCC GGGCTGGCCC TGCGACATGT
51781 GGTCTTCGGC GCGGAGGCGC TCGACCCGTC GCGACTGCGC GACTGGTTCG ACCTGCCGCT
51841 CGGCTCACGG CCGACGCTCG TGAACATGTA CGGCATCACC GAGACCACC TCCACGTCAC
51901 GGTGCTCCCG CTGGAGGATC GCGCGACGAG TCTTTCGGC AGCCCGATCG GTCGGCCCTT
51961 GGCCGATCTG CAGGTGTACG TCCTCGACGA ACGGCTCCGC CCGTGCCCC CAGGCACCGT
52021 CGGCGAGATG TACGTGGCAG GCGCCGGTCT GGCCCGCGGC TATCTGGGAC GCCCGCTCT
52081 GACCGCCGAG CGGTTCTGTT CCGACCCGAA TTCCCGTTCC GGCGGCCGTC TGTACCGCAC
52141 AGGCGACCTG GCCAAGGTGC GGCCCGACGG GGGACTGGAG TATGTGGGCC GCGGGGACCG
52201 GCAGGTGAAG ATCCCGGGCT TCCGGATCGA ACTCGGCGAG ATCGAGGCC CGCTGGTCAC
52261 ACACGCGGGT GTCGTCCAGG CGGTGGTCTT GGTGCGGGAC GAGCAGACCG ACGACCAACG
52321 GCTTGTGCGC CACGTGGTGC CCGCGCTGCC GCACCGGGCG CCGACCTGG CCGAACTCCA
52381 CGAGCACCTC GCGGCGACCC TGCCGGCGTA CATGGTGCCG TCCGCGTACC GGACCCTGGA
52441 CGAGCTGCCG CTGACGGCCA ACGAAAGCT CGACCGCGCG GCGCTGGCCG GGCAGTGGCA
52501 GGGCGGAACC CGCACCCGGA GACTGCCTCG GACGCCGCG GAAGAGATCC TGTGCGAGTT
52561 GTTCGCCGAC GTCCTCCGGT TGCCCGCCGC CGGGGCCGAC GACGACTTCT TCGCCCTGGG
52621 AGGCCATTCC CTGCTGGCGA CGGCCTCTCT GTCGGCTGTC AGGGGCACCC TGGGTGTGGA
52681 ACTCGGCATC CCGACCTCT TCGCCGCGCC CACGCCTGCC GGGCTCGCGA CCGTACTGGC
52741 GGCTCCGGC ACCGCCCTGC CACCTGTGAC CAGGATCGAC CGGCGCCCTG AACGGTCCC
52801 GCTGTCTTTC GCACAGCGGC GACTGTGGTT CCTGAGCAAG CTGGAAGGGC CCAGCGCCAC
52861 CTACAACATC CCGGTGCGCG TCCGGCTCAC CGGCGCCCTC GACGTCCCG CTCTCCGGGC
52921 CGCCCTGGGG GACGTACCCG CACGGCACGA ATCACTGCGT ACGGTCTTCC CCGACGACGG
52981 GGGCGAACCC CGCCAGCTGG TGCTCCACAC CGCCGAACCC CCCTTCTCTA CGCAGAGGT
53041 GACCGTCGGA GAGGTGGCGG AACAGCGCGC GTCCGCCACC GGGTACGACT TCGACATCAC
53101 CAGCGATACG CCGTGCGGG CCACCTGTTT GCGCGTCTCA CCGAGGAAC ACGTCTCTG
53161 GGTGGTTCAT CACCACATCG CCGGCGACGG CTGGTCCATG GGGCCGTTGG TCGTGAACCT
53221 GGTACCGGCC TACCGGGCCC GAACGCGGGG CGACGCCCCG GAGTACACCC CGCTTCCCGT
53281 GCAGTACGCC GACTACGCCC TGTGGCAACA CGCTGTTGCG GCGACGAGG ACGCCCCGGA
53341 CGGCCGGACG GCGCGTCGGC TCGGGTACTG GCGCGAGATG CTGGCCGGGC TGCCCGAGGA
53401 GCACACGCTG CCCGCCGACC GGCCCCGGCC CGTTCGGTCC TCGACCCGG GCGGCCGGGT
53461 ACGGTTTCGAA CTGCCCGCCG GCGTGACCCG GAGTCTGCTG GCCGTGGCGC GTGACCGTCG
53521 GGCCACGCTG TTCATGGTGG TGCAGGCTGC GCTCGCCGGT CTGTTGTCCC GGCTCGGCGC
53581 GGGCGACGAC ATCCCCATCG GCACCCCGGT CGCCGGGCGG GGCGATGAAG CGCTGGACGA
53641 CGTCGTCGGG TTTTTCGTCA ATACCCTGGT CCTTCGGACG AATCTCGCG GGGATCCGTC
53701 CTTCCGCCGAC CTGGTGGACC GGGTCAGGAC CGCCGACCTC GACGCGTTCC CGCACCAGGA
53761 CGTGCCCTTC GAACGGCTCG TGGAGGCGCT TGCGCCACGG CGTTCCCTCG CCCGCCACCC
53821 GCTGTTCCAG ATCTGGTACA CCCTACCAA CGCCGACCAG GACATCACCG GCCAGGCACT
53881 CAACGCCCTC CCGGGCCTGA CCGGGACGGA GTACCCGCTG GGGGCCAGTG CCGCCAAGTT
53941 CGACCTGTCT TTCACCTTCA CTGAACACCG CACCCCGAC GGAGACGCC CCGGCCTGTC
54001 CGTCTGCTC GACTACAGCA GCGACCTGTA CGACCACGGC ACTGCCGCC CACTGGGCCA
54061 CCGGCTGACC GGATTCTTCG CAGCACTGGC CGCCGACCCC ACCGCCCCC TGGGCACCGT
54121 CCCGCTCCTC ACCGACGACG AGCGGGACCG CATCCTCGGT GACTGGGGCA GCGGTACGCA
54181 CACCCGCTG CCCCCGCGCA GCGTGCCGGA GCAGATCGTC CGCCGGGCCG CGCTGGACCC
54241 GGACGCCGTC GCCGTCATCA CCGCGGAAGA GGAACCTCTG TACCGGGAAC TGGAAAGGCT
54301 CAGCGGTGAG ACGGCGCGGC TGCTGGCCGA CCGGGGGATC GGCCGCGAGA GCCTCGTCGC
54361 CGTCGCCCTG CCGCGACGG CCGGCCTGGT CACCACTCTG CTCGGCGTCC TGCGCACCGG
54421 CGCCGCTTAC CTCCCGCTCG ACACCGGGTA CCCC GCCGAG CGACTCGCGC ACGTGCTCTC
54481 CGACGCCCGT CCGGACCTCG TCCTCACCCA CGCCGGCCTC GCCGGACGGC TGCCGGCCGG
54541 CCTCGCGCCG ACCGTCTCTG TCGACGAGCC GCAGCCGCC GCGCGAGCC CCGCGCGGT
54601 TCCACGTCC CCGTCGGGCG ACCACCTCG GTACGTCATC CACACCTCC GCTCCACCGG

```

54661 CAGGCCCAAG GGCCTCGCGA TCGCCGAGTC CTCCCTGCGC GCCTTCCTCG CGGACGCGGT
 54721 CCGGCGCCAC GACCTGACCC CGCACGACCG GTTGCTCGCG GTGACCACCG TCGGCTTCGA
 54781 CATCGCCGGC CTCGAACTGT TCGCCCCGCT CCTCGCCGGT GCCGCGATCG TGCTGGCCGA
 54841 CGAGGACGCC GTACGCGACC CCGCCTCGAT CACCTCCCTG TGCGCACGCC ACCACGTCAC
 54901 CGTCGTCCAG GCCACGCCCA GTTGGTGGCG GGCCATGCTC GACGGAGCAC CGGCCGACGC
 54961 CGCCGCCCGG CTCGAGCACG TACGGATCCT GGTCGGCGGC GAACCGCTGC CCGCCGACCT
 55021 GGCCCGTGTC CTGACCGCAA CCGGCGCCGC CGTCACCAAC GTGTACGGAC CCACCGAAGC
 55081 CACCATCTGG GCCACCGCCG CCCACTCAC CGCCGGCGAC GACCGCACAC CCGGCATCGG
 55141 CACCCCCCTG GACAACTGGC GCGTCCACAT ACTCGACGCG GCCCTCGGAC CCGTTCCCCC
 55201 GGGTGTTCGG GCGGAGATCC ACATCGCCGG GTCCGGGCTC GCGCGCGGCT ATCTGCGCCG
 55261 CCCGGACCTC ACCGCCGAAC GCTTCGTGCG CAACCCGTTT GCGCCCGGCG AGCGGATGTA
 55321 CCGCACCGGC GACCTCGGCC GGTTCGCCCC GGACGGCACG CTCGAACACC TCGGCCGCGT
 55381 GGACGACCAG GTCAAGGTAC GGGGCTTCGG CATCGAACTC GGCGACGTCG AGGCCGCCCT
 55441 CGCCCGGCAT CCCGACGTGG GGCGCGCCGC CGCCGCGGTC CGCCCCGACC ACCGCGGCCA
 55501 GGGCCGCCTT GTCGCGTACG TCGTCCCCCG TCCCGGCACC CGGGGACCGG ACGCCGGCGA
 55561 ACTGCGCGAG ACGGTACGCG AACTTCTGCC TGACTACATG GTCCCCTCCG CCCAGGTGAC
 55621 TCTCACCACC CTGCCTCACA CCGCGAACGG CAAACTCGAC CGCGCCGCGC TGCCCCCCCC
 55681 CGTGTTTCGG ACCCCTGCCG GACGCGCCCC CGCCACCCGC GAGGAAAAGA TCCTCGCCGG
 55741 GCTCTTCGCG GACATCCTGG GCCTGCCCCG CGTGGGAGCC GACAGCGGCT TCTTCGACCT
 55801 CGGCGGCGAC AGCGTGCTGT CCATCCAGCT CGTGAGCCGC GCGCGGAGGG AAGGACTGCA
 55861 CATCACCGTA CGAGACGTGT TCGAGCACGG GACGGTCGGC GCACTCGCCG CCGCGGCCCT
 55921 TCCGGCACCG GCCGACGACG CGGACGACAC CGTCCCCGGC ACGGACGTAC TGCCTTCGAT
 55981 CAGCGACGAC GAATTCGAGG AGTTCGAGCT GGAGCTCGGA CTCGAGGGGG AGGAAGAGCA
 56041 GTGGTGAACC GCCGGTCGAA GGTAGTCGAG GAGATCCTGC CTGTCTCGGC GCTCCAGGAA
 56101 GGAGTCTGT TCCACAGCTC CTTGCGCCGC GCCGACGGAG TCGACGTGTA CGCGGGACAG
 56161 CTCGCGTTTC ACCTGGTTCG CGCGGTGGAC ACCGGTCGGC TGCGGGCCGC CGTCGAAAGC
 56221 CTCGTGGCGC GGCACGGCGT CCTGCGCTCA AGCTACCGTC AGGCGCGCTC CGGGGAGTGG
 56281 GTCGCGGTTC TGGCGCGGCG CGTCGCGACG CCATGGCGCG CCGTCGACGC CCGCGACGGT
 56341 GCCACGGACG CTGCCGCCGT GGCCCGGGAG GAACGCTGGC GCGCGTTTCG CCTGGGCCGG
 56401 GCGCCGCTGG CTCGGTTTCG GTCGTACGCG ACCGACGACG ACCGTTTCGG GTTCGTGATC
 56461 ACGTACCACC ACGTCATCCT CGACGGCTGG TCGCTGCCCG TACTGCTGCG CGAACTCCTT
 56521 GCCCTGTACG GAAGCGGCGC CGACCCGTCG GTGCTGCCGC CCGTCCGCCC CTACGGCGAC
 56581 TTTCTCCGGT GGGCCGCGCG GCGCGACGAC GCCCGCGCCG AAACCGCCTG GCGCGACGCG
 56641 CTCACCGGCC TGGACGAGCC CTCCCTGGTC GCACCCGGCG CTCCCCCGA CGGCGTCGTG
 56701 CCGGCCTCCG TCCACGCCGA ACTCGACAAG GCCGGCACCG AGAACCTCGC CGCCTGGGCC
 56761 AGGCACCGCG GCATCACCCA GGCCACCGCC GTCCGCGCCG CGTGGGCCCT GCTTCTCGGC
 56821 CAGCACACCG GCCGCGACGA CGTCTGTGTT TCCATCAACA CCGTCCCCCT GCGCACGGTC
 56881 CTCGCGCGCG CCGAGCACAT GGTCCGACAT TCCATCAACA CCGTCCCCCT GCGCACGGTC
 56941 CTCGACCCCG CCGACACCCT CGGCACGTTT GCCGCTCGCC TCCAGGCCGA ACAGACCACC
 57001 CTCCTCGAAC ACCAGCACGT GCGGCTCTCC GACATCCAGC GCTGGGCCGG ACACAAAGAA
 57061 CTCTTCGACA CCATTGTTCG CTTCGAGAAC TACCCCATCG GCCACAGCGG CCGCGGCTCC
 57121 ATCCGACCG ACGACTTCAC CGTCACCGCC ACCGAAGGCT CCGACGCCAC CCACTACCCC
 57181 CTCACCCTCA CCGCCGTACC CGGCGAAACC CTGCGCCTCA AGCTCGACCA CCGCCCCGAC
 57241 CTCGTGACGA CCACCACCGC CACCGCCCTG CTGCGCCGCG TGACCCGCGT CCTGGAAACC
 57301 GCCACCGACG ACACCGGGCA CACCCTCGCC CGCCTCGACC TCCTCGACGA CGACGAACGC
 57361 CACCGCCTGC TGCGCGGCTG GAACGACACC ACGCGCGAGC AGCCGCCCAC CTACTACCAC
 57421 CAGGAATTTC AGGAACAGGC GCGGAGGCGG CCGCACGACA CGGCCCTTGT CTTACCACAG
 57481 ACCTCCTGGA CGTACGAAGA ACTCAACGAC CGCGCCAACC GGCTCGCCCC CCTGCTCGTC
 57541 GCGCGCGGCG CCGGCTCCGA CGACTTCGTC GCGCTCGCCT TCCCCGTTT CCGGGAATCC
 57601 GTCGTGCGCA TCCTCGCCGT ACTCAAAGCG GGCGCCGCTT ACCTGCGGCT CGACATGGAC
 57661 CAGCCCGCCG AACGGCTCAC CGGCATCCTC CGCGACGAC ACCCGACCGT CGTCCTCAGC
 57721 ACCACCACCG CCACCCCGCT CCGCGACCCC GGCGCGACCC TCGTCTTCGA CAGCCCCACC
 57781 ACCGCCCCG GCGCTCGTGC GGCACCCGCA CACAACCTCA CCGACGCCGA CCGCCGTACC
 57841 CCGCTCAACG CCGCAACGCG CGCCTACATC ATCCACACCT CCGGCTCCAC CGGACGCCCC
 57901 AAGGGCGTCG TCATCGAACA CCGCAGTCTC GCCAACCTCT TCCACGACCA TCGGCGCGCC
 57961 CTCATAGAAC CCCATGCCGC CGGAGGATCA CGGCTCAAGG CCGGCCTCAC CGCCTCCCTC
 58021 TCCTTCGACA CCTCCTGGGA AGGTCTGATC TGCCTGGCCG CCGGCCACGA ACTGCACCTT
 58081 ATTGACGACG ACACCCGCGG AGACGCGGAA CGCGTCGCCG AACTCATCGA CCGGCAGCGC
 58141 ATCGACGTCA TCGACGTCAC CCCCTCCTTC GCCCAGCAAC TCGTAGAGAC CGGAATCCTC
 58201 GACGAGGGCC GCCACCACCC CGCCGCTTTC ATGCTCGGCG GTGAAGGCGT CGACGCGAAA
 58261 CTCTGGACCA GGCTCTCCGA CGTCCCCGGC GTCACCTCGT ACAACTACTA CGGCCCCACC

58321 GAATTCACCG TCGACGCCCT CGCCTGCACG GTCGGCATCG CACCCCGCCC CGTCATCGGC
58381 CACCCCTCG ACAACACGGC CGCCTACATC CTCGACGGCT TCCTGCGTCC CGTACCCGAA
58441 GGCGTCGCCG GCGAGCTCTA CCTCGCCGGC ACCCAGCTCG CCCGCGGCTA CGCCGGCCGG
58501 CCCGGCCTGA CGGCCGAACG CTTCGTGGCC TGCCCTTTCG GCGCGCCGGG CGAACGCATG
58561 TACCGCACCG GCGACCTCGT CCGCGCAGT CCCGGCGGCG TGGTCGAATA CCTCGGACGC
58621 GTGGACGATC AGATCAAACCT CCGCGGCTTC CGCATCGAAC CCGCCGAGAT CGAGCTCGCC
58681 CTGGCCGGCC ACCCCGCCGT CGCCAGAAC GTCGTCCTCC TGCACCGCTC CGCCACCGGA
58741 GAGGCTCGCC TCGTGGCGTA CGTCGTCCCC GGCACACCCG TGCACCCGCG GGAACTCACC
58801 GGGCACCTCG CGCCCGGGCT GCGCGCGTAC ATGGTGCCCT CGGCTTTCGT TCTCCTCGAC
58861 ACCCTCCCGC TCACCCCAA CGGCAAACCTG GACCGCGGCG CCCTGCCGGA GCGCGCTTC
58921 GGTACCGCGC CCCGCCCGA GCGCCCCCGC ACACCCGTCG AGGAGATCCT CTGCGGCCTG
58981 TACGCCGACG TGCTCGGGCT TCCCTCGTTC GCGCGCGACG ACGACTTCTT CGACGCCGGC
59041 GGGCACTCGC TGCTGGCCAG CAAACTCGTC AGCCGTATCC GTACGAACCT GAAAACCGAA
59101 CTCAACGTCC GCGCCCTCTT CGAGCACCGC ACGGTCTCCT CCCTGGCCAC CGCCCTCCAC
59161 CGGGCCGGCG AGGCCGGCCC CGCGCTCACC GCCGGACCGC GCCCCGCACG GATCCCGCTG
59221 TCGTACGCCC AGCGCCGCCT GTGGTTCCTC AACCGGCTCG ACCGCGACAG CGCCGCGTAC
59281 AACATGCCCC TCGCACTCCG CTGCGTGGC CCCCTGGACA GCACCGCCAT GTGCGCCGCA
59341 CTCACCGACG TCGCCGAACG CCACGAGGCG CTGCGCACCG TGTTCGAGGA GGACCGGGAC
59401 GGTGCCACCG AGATCGTGCT GCGCGCGACC GGCCTCGGCC CTCTGCTCAC CTTGACCGGG
59461 GCCGACGGGA CGACCCTGCG TGCCCTCATC ACCGAGTTCG TACGACGGCC GTTCGACCTG
59521 GCGGCGGAGA TCCCCTTCG CGCCGCACTG TTCCGCGTCG GCGACGAGGA ACATGTACTG
59581 GTCGTCTGTC TGCAACACAT CGCCGGGGAC GGCTGGTCCA TGGGACCGCT CGCACGCGAC
59641 GTGGCCGAGG CCTACCGGGC GCGGGCGGCC GGGAGGGCAC CCGACTGGGA ACCGCTGCCC
59701 GTGCAGTACG CCGACTACGC GCTCTGGCAG CGGGAGGTGC TGGGCGCGGA GGACGACGAG
59761 ACCGGCGAAC TCTCCGCCA ACTCGCCAC TGGCGCACCC GCCTCGCAGG GGCCCCCGCA
59821 GAACTCACGC TGCCACCGA CCGCCACGC CCCGCTGTCG CCTCCACCGC CGGAGACCGC
59881 GTCGAATTCA CCGTGCCCGC CGGACTCCAC CAGGCCCTCG CCGACCTGGC ACGGGCCAC
59941 GCGCGACGG TCTTCATGGT CGTCCAGGCC GCCCTCGCCG TCCTGTGTC ACGTCTCGGC
60001 GCCGGCGACG ACATCCCCAT CGGCACCCCG GTCGCCGGCC GCACCGACGA GGCCACGGAG
60061 GAACTGATCG GGTTCCTTCG CAACACGCTG GTGCTGCGCA CCGACGTGTC CGGCGACCCG
60121 ACGTTCGCCG AACTCCTCGC GCGGTGCGG GCCACCGACC TGCACGCGTA CGCACACCAG
60181 GACGTGCCAT TCGAACGTCT GGTGAGGTG TTGAACCCGG AGCGGTCACT GGCACGGCAT
60241 CCACTGTTCC AGGTCATGCT GACGTTCAAC GTCCCGGACA TGGACGGGT CGGAAGCCGG
60301 CTGGGGAATC TGGGGGAAC GGAGGTCTCC GGTGAGGCCA TCCGGACGGA TCAGACCAAG
60361 GTGGATCTCG CTTTCACGTG CACGGAGATG TACGCCCGCG ACGGTGCGGC CTCGGGAATG
60421 GCGGGGTGCG TGGAAATACG GCTTGATGTG TTCGGTGC GG TACAGGCCCG GAAACGACG
60481 GAGCGGTTGG TGCGGGTGTT GGAGGGTGTT GTTCTGGTG GGGGTGGGGT GTCTGTGTGCG
60541 GGGGTTGATG TGTTGGGTGT GGGTGAGCGG GAGAGGTTGT TGGGTGGGGT TGTGGGTGGG
60601 CCGGTGCCTG TGGTGCCGGG TGGTGGGTG GTGGGGTGT TCGAGGAGCG GGTGCGGGCC
60661 GACGCGGACG CCGTGCCCGT GCGTGCGCG GGGGTGGTGT GGAGTTATGG GGAGTTGAAT
60721 GCGCGGGTGA ATGTGGTGGC GCGGTGGTTG GTGGGTGCGG GTGTGGGGGC GGAGTGTGGT
60781 GTGGGTGTGG TGATGGGCCG CGGGGTGGAT GTGGTGGTGA TGTGCTGGC GGTGGCGAAG
60841 GCGGGTGGGT TTTATGTGCC GGTGGATCCG GAGTGGCCGG TGGAGCGGGT GGGGTGGGTG
60901 CTGGCGGATG CCGGGGTGGG GCTGGTTGTG GTGGGGGAGG GGTGTGCGCA TGTGGTGGGG
60961 GATTTTCCTG GGGGTGAGGT TTTGAGTTT TCGCGGGTTG TTCGTGAGTC GTGCTTGTG
61021 GAGTTGGTGG CTGCGGATGG GGTGAGGTT CGGAATGTGA CCGATGGTGA GCGGGCGTGC
61081 CGTCTGTTGC CGGGGCATCC GTTGATGTG GTTTATACGT CCGGTTCGAC GGGGCGGCCG
61141 AAGGGTGTTG TGGTGACGCA TGCTTCGGTG GGTGGGTATT TGGCGCGTGG TCGGGATGTG
61201 TATGCGGGTG CCGTTGGTGG TGTGGGGTTT GTGCATTCTG CGCTTGCGTT CGATCTGACG
61261 GTGACGGTTC TGTTACGCC TTTGGTGTCT GCGGGTTGTG TTGTGTTGGG TGAGTTGGAC
61321 GAGTCGGCGC AGGGGGTGGG TGCCTCGTTC GTGAAGGTGA CTCCGTCGCA TCTGGGTTTG
61381 CTGGGTGAGC TGGAGGGTGT GGTGGCGGGG AACGGCATGC TGCTGGTGGG GGGTGAGGCG
61441 TTGTGCGGTG GTGCGCTGCG TGAGTGGCGT GAGCGTAATC CGGGTGTGGT GGTGGTGAAT
61501 GCTTATGGTC CGACGGAGCT GACGGTGAAC TGTGCCGAGT TCCTTATCGC GCCTGGTGAG
61561 GAGGTTCCGG ATGGGCCTGT GCCGATCGGG CGTCCTTTCG CGGGTACGCG GATGTTTGT
61621 CTGGATGCGG CGCTGCGGGT GGTGCCGGTC GGTGTGGTGG GTGAGTTGTA TGTGGCGGGT
61681 GTGGGTCTGG CGCGGGGCTA TCTCGGGCGT GCGGGTCTGA CGGCGGAGCG GTTCGTGGCC
61741 TGCCCTTTCG GTGCGCCGGG TGAGCGTATG TACCGTACGG GGGATCTGGT GCGGTGGCGG
61801 GTGGACGGCG CGCTTGAGTT TGTTGGTCTG GCGGATGATC AGGTGAAGGT CCGTGGTTTC
61861 CGTGTGGAGT TGGGTGAGGT GGAGGGTGCT GTTGCGGCGC ATCCTGATGT GGTGCGTGCG
61921 GTTGTGTG TGCGTGAGGA CCGGCCGGT GATCACCGGT TGGTTGCGTA TGTACCGGT

61981 GTTGACACGG GTGGACTGTC CTCTGCGGTG ATGCGTGCCG TTGCTGAGCG TCTGCCTGCG
62041 TACATGGTGC CGTCGGCGGT GGTGGTTCCTG GATGAGATCC CGTTGACGCC GAATGGGAAG
62101 GTGGACCGGG CGGCGCTTCC GGTGCCGGGG GTGGAGGCGG GCGCGGGCTA CCGGGCGCCT
62161 GTTTCGCCGC GGGAGGAGGT GTTGTGTGGT CTGTTGCGCG AGGTGCTGGG GCTGGAGCGG
62221 GTGGGGGTGG ACGATGATTT CTTCGGGTG GGTGGTCATT CTCTTCTGGC GACTCGTCTG
62281 ATTTTCGCGTG TCCGTGCGGT GTTGGGTGTT GAGGCGGGTG TCGGGCGGTT GTTCGAGGCG
62341 CCGACGGTGA GCCGTTTGA GCGGTTGCTG CGGGAGCGGT CGGCTTTGGG GGTGCGGGTG
62401 CCTCTGGTGG CACGGGAGCG GACGGGTGCG GAGCCGTTGT CGTTCGCTCA GCAGCGTCTG
62461 TGGTTCCTTG AGGAACCTGA AGGGCCCGGT GCTGCGTACA ACATTCCGAT GGCGCTGCGT
62521 CTGGCCGGTG TTCTGGACGT CGAAGCGCTG CACCAGGCGC TCATTGATGT CATCGCCCGC
62581 CACGAAAGCC TCCGCACCCT CATCGCGCAG GATGCGGGTA CTGCCTGGCA GCACATCCTG
62641 CCCGTTGACG ACCCTCGCAC CCGTCCCGGT CTCCCTCTTG TGGACATCGG TGCCGACGCC
62701 CTTCAGGAGC GGCTCGACGA AGCCGCCGGC CGGCCCTTCG ATCTCGCGGC CGATCTCCCG
62761 GTCCGGGCCA CAGTCTTCCG CCTCACCGAC AACGACCACA TCCTCCTGGT CGTGGCCCAT
62821 CACGTGGCCT TCGACGCGAT GTCCCGTGTG CCGTTCATCC GGAACGTCAA GCGCGCCTTC
62881 GAGGCCCCGA CGAACGGCGC GGCCCCCGAC TGGAGGCCGC TGCCCGTGCA GTACGCGGAT
62941 TATGCGGCCCT GGCAGCGCGA CGTACTCGGC ACGGAGGACG ACGAGTCGAG CGAGCTGTGCG
63001 GCCCAGCTCG CCTACTGGCG CACCCTACTA GCCTCACTAC CGGCCGAGTT GGCGCTCCCG
63061 ACGGACCGGG CCCGGCCCCG CGTCGCCTCG TACGAAGGCG GCAAGGTGCA GTTCACCGTC
63121 CCCGCCGGGG TGTATGACGG CTTGGTGGCT CTCGCCCGTG CCGAGGGTGA CACGGCTTTC
63181 ATGGTCGTGC AGGCGGCGCT GGCCGCGCTC CTCTCCCGGC TCGGCGCCGG CGACGACATC
63241 CCCATCGGCA CCCCAGTCGC CGGCCGCACC GACCAGGCCA CCGAAGATCT CATCGGCTTC
63301 TCCGTGAACA CCCTCGTCCT GCGCACCGAC GTGTCCGGCG ACCCGACGTT CGCCGAACTC
63361 CTCGCGCGCG TCCGGGCCAC CGACCTCGAC GCCTACGCCC ACCAGGACAT CCCCTTCGAA
63421 CGACTGGTCG AAGCGGTCAA CCCCAGCGC TCCCTCGCCC GCCACCCCTT CTTCCAGGTC
63481 ATGCTGACCT TCGACAACAC GATTGACCGT GAGGTCACGG AGGGCTTCGC GGGCTCGGG
63541 GTGGAAGGCC TGCCGCTGGG TCGGGGAGCG GTCAAATTCG ATCTGCTCTT CGGTCTCTCC
63601 GAGGTGGGCG GCGAGCTGCG CGGAGCCGTG GAGTACCGCT GCGATCTCTT CGACCACCCG
63661 ACGGTGGCGC AGCTCGCGGA GCGCCTGGTG CGGGTACTGG AGCGCGTGGC TTCCGACGCT
63721 TCGGTACGCA CGGGTGAAC TCCGGTTCGT GGCAGAGCGG AGCGCGCCCG TGCTCTGACG
63781 GAGTGGAATG ACACGGGCGT CCCCAGGTGT CCGGAAACAT TCCTGGAGTT GTTCGAGGCG
63841 CAGGTCGCGG CCCGGGGTGA CGCGCCGGCG GTCGTGTACG AGGGTGAGGT TCTGTCTGAC
63901 CGGGAACCTG ACGCGCGGGC GAACCGCCTG GCCGGGCTGC TGGTGGGGCG CGGTGCGGGC
63961 CCGGAGCATT TCGTGGGGGT GCGGTGCTCG CCGGTGAGGT ATCTGATCGT GGCCCTGCTG
64021 GCCGTGCTCA AGTCCGGTGC CGCGTACGTT CCCCCTGGAC CGGAGTACCC GGCCGAGCGG
64081 CTGGTCCACA TGGTCACCGA CGCCGCCCCC GTCGTGGTCG TGACCTCCAC CGACGTACGT
64141 ACTCTGCGGA CCGTTCCCCG GGTGAGCTG GACGACGAGG CGACCCGCGC CACCCTGGTC
64201 GCAGCCCCCG CCACAGGGCC CGACGTGAAG ATGTCCGCCT CCCACCCCG GTACGTGATC
64261 TACACCTCCG GGTCCACGGG CCGCCCCAAG GCGTCTGTC TACGCCACGG CAGCCTGGCC
64321 AACTTCCTCG CCTGGGCGCG GGAAGACCTG GGTGCCGAGC GGCTCCGGCA CGTCTGTTG
64381 TCCACGTCCC TCAGCTTCGA CGTCTCCGTG GTCGAACTCT TCGCCCCGCT GTCCTGCGGC
64441 GGCACCGTCG AGATCGTCCG GAATCTGCTG GCCCTCGTCG ACCGCCCCGG CCGATGGTCC
64501 GCGAGCCTGG TCAGCGGCGT GCCGTGCGCC TTCGCGCAGC TGCTGGAAGC CGGCCTCGAC
64561 CGGGCCGACG TGGGCATGAT CGCCCTGGCC GGCAGGCGC TGTCGCTCG CGACGTGCGC
64621 CGCGTCCGCG CTGTGCTGCC CGGGGCCCGC GTGGCCAAT TCTACGGCCC GACCGAAGCC
64681 ACCGTCTACG CCACGGCCTG GTACGGCGAC ACCCCCATGG ACGCCGCGG CCCCATGGGC
64741 CGGCCCCCTG GCAACACGTG TGTGTATGTG CTGGACGACG GGCTGGGCTA GGTGCCGGTC
64801 GGTGTGGTGG GTGAGCTGTA TGTGGCGGGT GTGGGTCTGG CGCGGGGCTA TCTCGGGCGT
64861 GTGGGTCTGA CGGCGGAGCG GTTGTGCGCG GTTCCGTTCC GTGCGCGGGG TGAGCGTATG
64921 TATCGCACGG GGGATTTGGT GCGGTGCGGG GTGGACGGCA CGCTTGAGTT TGTGTTGTCG
64981 GCGGATGATC AGGTGAAGGT CCGTGGTTTT CGTGTGGAGT TGGGTGAGGT GGAGGGTGCT
65041 GTTGCGGCGC ATCTGATGT GGTGCGTGCG GTTGTGTTGG TGCGTGAGGA CCGGCCGGGT
65101 GATCACCGGT TGGTTGCGTA GTTCACCGGT GTTGACACGG GTGGACTGTC CTCTGCGGTG
65161 ATGCGTGCCG TTGCTGAGCG TCTGCCTGCG TACATGGTGC CGTCGGCGGT GGTGGTTCTG
65221 GATGAGATCC CGTTGACGCC GAACGGGAAG GTGGACCGGG CGGGTCTTCC GGTGCCGGTG
65281 GTGTGCGTGG CGGGGTTCTG TGCGCCGTCG TCGCCGCGGG AGGAGGTGTT GTGTGGTCTG
65341 TTCGCGGAGG TGCTGGGTGT TGAGCGGGTG GGGGTGGACG ATGGGTTCTT CGATCTGGGC
65401 GGGGACAGCA TTCTGTGCGT TCAGTTGGTG GCGCGGGCTC GTCGGGCGGG TCTGGAGTTG
65461 TCGGTTTCGG ATGTTTTTGA GGGCCGTACG GTACGTGCTC TGGCGGCTGT GGTGCGTGGT
65521 TCGGACGCTG GGGCGGTTGG TGTGGTGGGG GGTGCTGAGA TTGTGCTGCC GGGTGTGGGT
65581 GAGGTGGAGC GGTGGCCGGT GGTGGAGTGG CTGGCGGAGC GTGGTGGGGG GTCGCTGGGT

65641 GGTGTGGTTC GGGGTTTCAA TCAGTCTGTT GTGCTTGCTG TGCCTGCTGG GTTGGTGTGG
65701 GAGGAGTTGC GGGTGTGTGT GGGTGCCGTG CCGGATCGGC ATGAGGCGTG GCGGTTGCCG
65761 GTGCTGGATT CCGGGGCGTT GTGTGTGAT GGTGTGTTC CCGATGACGG GTCGTGGATT
65821 GTCCGGTGTG ACCTGAGCGG TATGGGTGTG GATGGTCAGG TGGATGCTGT GCGGGCTGCG
65881 GCTGTGGAGG CGCGTGCGTG GCTGGATCCG TCGGTGGGCC GGGTGGTGGG GCGGTGTGG
65941 CTGGAGCGTG GTGGTGATCG TTCGGGGGTG TTGGTGCTGG TGGCGCATCA CCTGGTGGTG
66001 GACGGTGTGT CGTGGCGGGT GGTGCTGGGG GATCTGGCGG AGGGGTGGGC GCAGGTGCGT
66061 TCGGGTGGCC GTGTGGAGTT GGGTGTGGTG GGGACGTCGT TGCGGGGTTG GCGGGCGGCG
66121 TTGGCGGAGC AGGGCCGGCG GGGCGAGCGT CCGGGGAGG TGGAGTTGTG GTCGCGGATG
66181 GTTCGGGGTG CCGATGTTCT GGTGGGGTCG CGTGCTGTGG ATGGTGCGGT GGATGTTTTT
66241 GCGGGGGTGG TGTCGGTTGA TTCGCGGGCG TCGGTGTCGG TGTCGCGTGC GTTGCTGACG
66301 GAGGTGCCGT CGGTTCTGGG GTTGGTGTG CAGGAGGTGT TGCTGGCGGC ATTCGGGCTG
66361 GCGGTGCGCG GGTGGCGCGG CCGGGGTGGG CCGGTGTGG TGGATGTTGA GGGGCACGGG
66421 CGTAATGAGG ACGCTGTGCG GGGTGCTGAT CTGTCTCGTA CTGTGCGTTG GTTCACCAGT
66481 GTGTATCCGG TCCGTGTGCC GGTGGAGTCC GCTTCGTGGG ACGAGGTGCG TGCGGGTGGT
66541 CCGGTGGTGG GCCGTGTGGT GCGTGAGGTG AAGGAGACTC TGCGTTCGCT GCCTGACCAG
66601 GGTCTGGGTT ATGGCATCCT GCGCTATCTC GATCCCGAGC ACGGTCCTGC TCTGGCCCGG
66661 CATGCCACCC CGCAGTTCGG TTTCAACTAC CTCGGCCGCT TCACCACCGG AACCCACGAC
66721 ACCGGTGACG AGGGGATGAC GGACTGGGTG CCCGTGTCAG GGCCGTTCGC GGTGGGAGCC
66781 GGCCAGGACC CCGAAGTCCG CGTGGCGCAC GCGGTCGAGT TCAACGCGAT CACGCTGGAC
66841 ACCCCGGAGG GCGCGCGCCT GGGCGTGACA TGGTCGTGGC CGACGACGCT GCTGCCGGAG
66901 TCCCGGATAC GGGAGCTGGC CCGCTACTGG GACGAGGCCC TGAAGGGCT GGTGGAACAC
66961 CCCCCGACAC CCGAAGCCGG CGGCCTCACG CCGTCCGACG TGACGCTGGT GGAAGTGAAC
67021 CAGGTGGAGC TCGACCGTCT GCAGGCGGGG GTCGCCGGTG GTGCGGAGGA GATTCTGCCG
67081 GTGTGCGCCC TGCAAGAGGG GCTGCTGTTT CACAGCGCGT TGGCCTCTGG TGGGGTGGAC
67141 GTGTATGTGG GGCAGCTGGT GTTCGATCTG GTCGGTCCGG TGGACGTCGA CCGGCTGCGC
67201 GCGGCTGTG AAGGTCTGGT GGC CGCGCAC GGGGTGCTGC GGTCCGGGATA CCGCCAACCTG
67261 CGGTCGGGCG AATGGGTTGC GGTGCTCGCA CGACAGGTGG ATCTGCCGTG GCAGTCCATC
67321 GACGTGCGCG ACGGCGGTAT CGACGGGTTG GTGGAAGAGG AGCGTGCGG CCGGTTCCGAC
67381 ATGGGCCCGG GTCCACTGGC GCGCTTCGTG CTCATCCGGA CGCACGACGA TCGTTTCCGG
67441 TTCGTCATCA CGTACCACCA CGTCGTCCTC GACGGCTGGT CCGTCCCGGT GCTGCTGCGT
67501 GAGCTGCTGG CCCTGTACGG CAGCTCGGGG GACGTATCGG TTCTGCCGGG GGTCCGCTCG
67561 TACGGCGATT TCCTGCGATG GGTGCGCGCG CGAGACGCCG CAGCCGCCGA AGGCGCATGG
67621 CGGCGGGCGC TGACGGGCGT CCAGGAGCCG TCGCTCGTCG CGCCAGGCGT TTCCCGAGAC
67681 GGGGTGCTCC CGGCGGCGTT CCACGGTGCG GTCGACGGCG ACCTCTCGCA GAAGATCGTG
67741 GCGTGCGCGC GCGGGCGTGG TGTGACGGTT GCGTCCGGTG TACAGGCGGC GTGGGCTTGG
67801 GTGCTGGGGC GGTGTGATGG TCGGGACGAT GTGGTGTTCG GGGTGACGGT GTCGGGTGCG
67861 CCTGCCGAGG TGGTGGGTGT GGAGGACATG GTCGGTCTGT TCGTGAACAC CATTCGGTTG
67921 CGGGCGCGGC TGGATCCGGC GGAGTCGCTG GGTGGTTTTG TGGAGCGGCT GCAGCGGGAG
67981 CAGACGGAGC TGCTGGAGCA TCAGCATGTC CCGCTGGCGG AAGTCCAGCG GTGGGCCGGG
68041 CACAAGGAAC TCTTCGATGT CCGAATGGTC TTCGACAAC ACTCCGGTTTC TTCTGAATCC
68101 CCGGAAGCGG AATTCCAGAT CTCACGAACA GCGGATACA ACGGAACCCA CTACGCACTG
68161 AACCTCGTTG CTTCCATGCA CGGCCTGGAG CTGGAACCTG AAATCGGTTA TCGGCCGGAT
68221 GTGTTTGATG CGGGTCGGGT GCGTGAGGTG TGGGGATGGT TGGTGCGGGT GTTGGAGGGT
68281 GTGGTTTCTG GTGGGGGTGG GGTGTCTGTG TCGGGGGTTG ATGTGTTGGG TGTGGGTGAG
68341 CGGGAGAGGT TGTGGGGTG GGGTGTGGGT GGGCCGGTGC CTGTGGTGCC GGTGGGTGGG
68401 TTGGTGGGGT TGTTGAGGA GCGGGTCGG GCGGACGCGG ACGCGGTGGC CGTGCCTGGC
68461 GCGGGGGTGG TGTGGAGTTA TGGGGAGTTG AATGCGCGGG TGAATGTGGT GGCGGGGTGG
68521 TTGGTGGGTG GGGGTGTGGG GCGGAGTGT GGTGTGGGTG TGGTGATGGG CCGCGGGGTG
68581 GATGTGGTGG TGATGTTGCT GCGGTTGGCG AAGGCGGGTG GGTTTTATGT GCCGGTGGAT
68641 CCGGAGTGGC CGGTGGAGCG GGTGGGGTGG GTGCTGGCGG ATGCCGGGGT GGGGCTGGTT
68701 TGGTGGGGG AGGGGTGTGC GCATGTGGTG GGGGATTTTC CTGGGGGTGA GGTTTTTCGAG
68761 TTTTCGCGGG TTGTTGCTGA GTCGTGCTT GTGGAGTTGG TGGCTGCGGA TGGGGTTGAG
68821 GTTCGGAATG TGACGATGG TGAGCGGGCG TCGGCTCTGT TGCCGGGGCA TCCGTTGTAT
68881 GTGGTTTATA CGTCGGGTTG GACGGGGCGG CCGAAGGGTG TTGTGGTGAC GCATGCTTCG
68941 GTGGGTGGGT ATTTGGCGCG TGGTCGGGAT GTGTATGCGG GTGCCGTTGG TGGTGTGGGG
69001 TTTGTGCATT CGTCGCTTGC GTTCGATCTG ACGGTGACGG TTCTGTTTAC GCCTTTGGTG
69061 TCTGGCGGTT GTGTTGTGTT GGGTGAGTTG GACGAGTCGG CGCAGGGGGT GGGTGCCCTC
69121 TTCGTGAAGG TGAATCCGTC GCATCTGGGT TTGCTGGGTG AGCTGGAGGG TGTGGTGGCG
69181 GGAACGGCA TGCTGCTGGT GGGGGGTGAG GCGTTGTGCG GTGGTGCGCT GCGTGAGTGG
69241 CGTGAGCGTA ATCCGGGTGT GGTGGTGGTG AATGCTTATG GTCCGACGGA GCTGACGGTG

69301	AACTGTGCCG	AGTTCCTTAT	CGCGCCTGGT	GAGGAGGTTT	CGGATGGGCC	TGTGCCGATC
69361	GGGCGTCCTT	TCGCGGGTCA	GCGGATGTTT	GTTCTGGATG	CGGCGCTGCG	GGTGGTGCCG
69421	GTCGGTGTGG	TGGGTGAGTT	GTATGTGGCG	GGTGTGGGTC	TGGCGCGGGG	CTATCTCGGG
69481	CGTGTGGGTC	TGACGGCGGA	GCGGTTTGTG	GCGTGTCCGT	TCGGTGTGCC	GGGTGAGCGT
69541	ATGTATCGCA	CGGGGGATT	GGTGCGGTGG	CGGGTGGACG	GCGCGCTTGA	GTTCCGTTGGC
69601	CGTGCGGATG	ATCAGGTGAA	GGTCCGTGGT	TTCCGTGTGG	AGTTGGGTGA	GGTGGAGGGT
69661	GCTGTTGCGG	CGCATCCTGA	TGTGGTGCGT	GCGGTTGTTG	TGGTGCGTGA	GGACCGGCCG
69721	GGTGATCACC	GGTTGGTGCG	TTACGTGACT	GCGGGTGGTG	TTGGTGGGGA	TGGTCTTCGT
69781	TCCGCGATCT	CTGGTTTGGT	GGCTGAGCGT	CTGCCTGCGT	ACATGGTGCC	GTCGGCGGTG
69841	GTGGTTCTGG	ATGAGATCCC	GTTGACGCCG	AACGGGAAGG	TGGACCGGGC	GGCGCTTCCG
69901	GTGCCGGAGG	TGGAGGCGGG	CACGGGCTAC	CGGGCGCCTG	TTTCGCCGCG	GGAGGAGGTG
69961	TTGTGTGGTC	TGTTCCGCGA	GGTGCTGGGT	GTTGAGCGGG	TGGGGGTGGA	CGATGACTTC
70021	TTCGAGTTGG	GTGGTCATTC	TCTTCTGGCG	ACTCGTCTGA	TTTCGCGTGT	CCGTGCGGTG
70081	TTGGGTGTTG	AGGCGGGTGT	GCGGGCGTTG	TTTCGAGGCG	CGACGGTGAG	CCGTCGTGAG
70141	CGGTTGCTCC	GGGAGCGGTC	GGGTTTGGGG	GTGCGGGTGC	CTCTGGTGGC	ACGGGAGCGG
70201	ACGGGTCGGG	AGCCGTTGTC	GTTTCGCTCAG	CAGCGTCTGT	GGTTCCTTGA	GGAATCGAA
70261	GGGCCCGGTG	CTGCGTACAA	CATTCCGATG	GCGCTGCGTC	TGGCCGGTGT	TCTGGACGTC
70321	GAAGCGCTGC	ACCAGGCGCT	CATTGATGTC	ATCGCCCGCC	ATGAAAGCCT	CCGACCCTTC
70381	ATCGCGCAGG	ATGCGGGTAC	TGCCTGGCAG	CACATCCTGC	CCGTTGACGA	CCCTCGCACC
70441	CGTCCCGGTC	TCCCTCTTGT	GGACATCGGT	GACCGACGCC	TTTCGAGCGG	GCTACGAGAA
70501	GCCCGCGGCC	GGCCCTTCGA	CCTCGCGGCC	GATCTCCCGG	TCCGGGCCAC	AGTCTTCCGC
70561	CTCACCAGCA	ACGACCACAT	CCTCCTGCTG	GTCCTGCACC	ACATCGCCGG	CGACGGCTGG
70621	TCGATGGGCC	CGCTCGCCCG	CGATCTCTCC	ACGGCGTACA	GCGCACGCGC	CGCAGGAGCC
70681	GCCTCGGCCT	GGCGGCCCTT	CTCCGTGCAG	TACGCGGATT	ATGCGGCCTG	GCAGCGCGAC
70741	GTAATCGGCA	CGGAGGACGA	CGAGTCGAGC	GAGCTGTGCG	CCCAGCTCGC	CTACTGGCGC
70801	ACCAACTAG	CGTCACTCCC	AGCCGAGTTG	GCGCTCCCGA	CGGACCGGGC	CCGGCCCGCC
70861	GTCGCCACCT	ACCGGGGCGG	ACGCATCGAG	TTCACCATCC	CCGCCGACGT	CCACCGCAGC
70921	CTCGCCGACC	TCGCCCGTGC	CGAGGGTGTG	ACGGTCTTCA	TGGTCGTGCA	GGCGGCGCTG
70981	GCCGCGCTCC	TCTCCCGGCT	CGCGCGCCGG	GACGACATCC	CCATCGGCAC	CCCGATCGCC
71041	GGCCGCACCG	ACCAGGCCAC	CGAAGATCTC	ATCGGCTTCT	TCGTGAACAC	CCTCGTCTTG
71101	CGCACCGACG	TCTCCGGCGA	CCCAGCGTTC	GCCGAATCC	TCGCGCGCGT	CCGGGCCACC
71161	GACCTCGACG	CCTACGCCCA	CCAGGACATC	CCCTTCGAAC	GACTGGTCGA	AGCGGTCAAC
71221	CCCGAGCGCT	CCCTCGCCCG	CCACCCCTCT	TTCCAGGTCA	TGCTCGCCTT	GACCAACGCC
71281	GAGACGAGCA	CCCCGCTGCC	CATGGCCGAA	GCCCTGGCTG	CCTCCCGGCA	GGACATCGAA
71341	CCGGGCGTGG	CGAAATTCGA	TCTGGCCCTG	TATTGCAACG	AATCCCGCGG	TGAGACGGGC
71401	GACCACCAGG	GCATCAGAAG	TGTCTTCGAG	TACCGCCGCG	ACCTGTGGGA	CGAGGACACC
71461	GTGCGGCAGC	TCGCCGACCG	GTTCTGTCAT	GTTCTCGCTG	CTTTTTCGCG	AGCCCCGGAG
71521	CAACGTGCGA	GCAGCGTCGA	CGTGCTCCGG	GCGGGCGAGC	GCGACCAACT	GCTGCACGAG
71581	TGGAACGACA	CGGCTGCCGC	TCTCCCCCGG	GCACTGCTGC	CCCAGCTGTT	CGAGGAGCAG
71641	GTGCGGCGCA	CCCCGCACGA	TGTCGCTCTC	GTCTCGGGGA	ACATCCGGCT	CACGTACGCG
71701	GAGCTGGACG	CGCGCGCGAA	CCGCTGGGCC	CACCTGCTGC	TCGCCCCGGG	CGCGGCCCCC
71761	GAGACGTTTC	TCGCGGTGGC	CCTGCCCCGG	ACCGAAGAGC	TCCTGGTGGC	CCTGCTGGCC
71821	GTACAGAAAA	CAGGTGCCCG	ACATCTGCCG	CTGGATCCCG	GCTTCCCGGC	CGAGCGGCTC
71881	AGCTACATGC	TGGATGACGC	CCGCCCTGCG	GTGGTCCTCA	CCACGGAGGA	CATCAGCGCC
71941	CGCATACCCG	GCGGAAGCCA	TGTGGTACTC	GACTCCGAGC	AGGTGACCGG	CGAGCTCCAC
72001	GACCACCCGG	CCACGTCCCC	CGCCGGCCGG	GGCAACCCCG	CCGGCCCGGC	GTACGTGATC
72061	TACACCTCCG	GATCCACCCG	CCAGCCCAAG	GGCGTCGTGC	TACCGTCGGC	CGCCCTGGTG
72121	AACTTCTCTG	CCGACATGGT	GCCAGGGCTC	GGGCTCCGCG	GTGGCGACCG	GTCGCTGTCC
72181	GTGACCACCG	TGGGCTTTCGA	CATCGCGGCC	CTCGAGCTCT	TCGTCCCGCT	ACTGAGCGGC
72241	GCCACCGTCG	TCCTCGCGGA	CGGGGAGACG	GTCCGCGACC	CGGCGCTGGC	CCGCCAGACG
72301	TGCGAGGACC	ACGGCGTCAC	CATGGTCCAG	GCGACACCGA	GCTGGTGGCA	CGGCATGCTC
72361	GCCGACGCGG	GCGACAGCCT	GCGCGGCGTG	CACGCCGTCG	TGGGCGGTGA	GGCCCTGAGC
72421	CCCGGGTTGC	GCGACGCGCT	GACACGAGGC	GCGCGGTCCG	TCACGAACAT	GTACGGCCCC
72481	ACGGAGACGA	CCATCTGGTC	CACCAGCGCC	GGGAGGCGCG	CCGGGGACAG	CGCTCCCCCT
72541	TCGATCGGCA	CACCCATCCT	CAACACTCGC	GTGTATGTGC	TCGACGCTGC	TTTGTGTGTC
72601	GTGCCACCGG	GCGTCGCAGG	CGAGCTGTAC	ATCGCGGGCG	ACGGCTTCGC	GCGGGGCTAT
72661	CTCGGGCGTG	CGGGTCTGAC	GGCGGAGCGG	TTCGTGGCCT	GCCCCCTTCG	TGCGCCGGGT
72721	GAGCGTATGT	ACCGTACGGG	GGATCTGGTG	CGGTGGCGGG	TGGACGGCGC	GCTTGAGTTT
72781	GTTGGTCGTG	CGGATGATCA	GGTGAAGGTC	CGTGGTTTCC	GTGTGGAGTT	GGGTGAGGTG
72841	GAGGGTGCTG	FTGCGGCGCA	TCCTGATGTG	GTGCGTGCGG	TTGTTGTGGT	GCGTGAGGAC
72901	CGGCCGGGTG	ATCACCGGTT	GTTTGCGTAT	GTCACCGGTG	TTGACACGGG	TGACTGTCC

72961 TCTGCGGTGA TCGGTGCCGT TGCTGAGCGT CTGCCTGCGT ACATGGTGCC GTCGGCGGTG
73021 GTGGTTCTGG ATGAGATCCC GTTGACGCCG AATGGGAAGG TGGACCGGGC GGCGCTTCCG
73081 GTGCCGGGGG TGGAGGCGGG CGCGGGCTAC CGGGCGCCTG TTTCGCCGCG GGAGGAGGTG
73141 TTGTGTGGTC TGTTCGCGGA GGTGCTGGGT GTTGAGCGGG TGGGGGTGGA CGATGATTTC
73201 TTCGGGTTGG GTGGTCATTC TCTTCTGGCG ACTCGTCTGA TTTCGCGTGT CCGTGCGGTG
73261 TTGGGTGTTG AGGCGGGTGT GCGGGCGTTG TTCGAGGCGC CGACGGTGAG CCGTTTGGAG
73321 CGGTTGCTGC GGGAGCGGTC GGGTTTGGGG GTGCGGGTGC CTCTGGTGGC ACGGGAGCGG
73381 ACGGGTCGGG AGCCGTTGTC GTTCGCTCAG CAGCGTCTGT GGTTCCTTGA GGAAC TGGA
73441 GGGCCCGGTG CTGCGTACAA CATTCGATG ATCGCCCGCC ACGAAAGCCT CCGCACCCCTC
73501 GAAGCGCTGC ACCAGGCGCT CATTGATGTC CACGCCCCGG CAGCAGGTGC TGCCCGTCGG TGACCCCGCC
73561 ATCGCCCGCG ACAGTGACGG CACGCCCCGG CAGCAGGTGC TGCCCGTCGG TGACCCCGCC
73621 GCGCGACCGG CTCTTCCGGT CGTACAGACC GACGCCGACA CCCTCGTCGC GAAACTGAAC
73681 GAGGCCGTGC GCCGCCCTT CGACCTCACG GCCGAGATGC CCCTGCGTGC CACCGTCTTC
73741 CGGGTGGCCG ACGAGGACCA CGCGCTGCTG CTGGTGTTCC ACCACATCGC CGGCGACGGC
73801 TGGTCGACGG GCCTGCTCGC CCGCGACCTG TCCACCGCGT ACGCAGCCAG GCTCGAAGGC
73861 CGGGACCCCC AACTGCCACC CCTCCCCGTG CAGTACGCGG ACTACGCGGC CTGGCAGCGC
73921 GACGTACTCG GCACGGAGGA CGACGAGTCG AGCGAGCTGT CGGCCAGCT CGCCTACTGG
73981 CGCACCCAAC TTGCCGACCT CCCAGCCGAG TTGGCCCTCC CGGCGGACCG GGTCCGGCCC
74041 GCCAGGGCCT CGTACGAAGG AGGCCGGGTC GGCTTACCCG TCCCCGCCGG GGTCCCTCCG
74101 GACCTCACGC GCCTGGCCCG TGTCGAGGGT GTCACGGTCT TCATGGTCTG GCAGGCGGCG
74161 CTGGCCGCGC TCCTCTCCCG GCTCGGCGCC GCGCAGACA TCCCCATCGC CACCCCGATC
74221 GCCGGCCGCA CCGACCAGG CACCGAAGAT CTCATCGGCT TCTTCGTGAA CACCTCGTCTC
74281 CTGCGCACCG ACGTCTCCGG CGACCCGACG TTCGCCGAAC TCCTCGCGCG CGTCCGGGCC
74341 ACCGACCTCG ACGCCTACGC CCACAGGAC ATCCCTTCG AACGACTGGT CGAAGCGGTC
74401 AACCCCGAGC GCTCCCTCGC CCGCCACCCC CTCTTCCAGG TCATGCTCGC CTTGACAAC
74461 ACGGCCGACG GAGGCCCGT AGAAGACTTC CCCGGACTGT CCGCAGCCGG GCTGCCGTTG
74521 GGTGCGGGCG CGGCGAAGTT CGATCTGCTC TTCGGTCTCT CCGAGGTGGG CGGCGAGCTG
74581 CGCGGAGCCG TGGAGTACCG CTGCGATCTC TTCGACCACC CGACGGCCGC ACGGATCGCG
74641 GAGCGCCTGG TGCGGGTGCT GGAGCGGGTC GCCGCCGACG CGTCGGTACG CCTGGGCGAG
74701 CTGCCCCGTG TGAGCGACGC CGAGCGGGCC TGCGTCCTGA CGGAGTGGA CGACACCGCC
74761 GTCCCCGGCG TGACGGGAAC GCTGTGCGCG CTGTTGAGG CACGGGCCGC AGCCCCGGGC
74821 GACGCGCCGG CGGTCTGTGA CGAGGGTGAA GAACTGTCTG ACCGTGAACT GAACACACGC
74881 GCCAACCGCC TCGCCCATGT CCTGGCCGAG CACGGCGCAG GCCCCGAGCG GTTCGTGGT
74941 GTGGCCCTGC CCCGAGTCC GGACCTCGTA GTGGCACTGC TGGCGGTCTG GAAATCGGGC
75001 GCGGCCTACG TACCGCTCGA CCCCAGTAC CCGGCCGACC GGCTCGCGTA CATGGCCGGC
75061 GACGCTGCCC CCGTGGCGGT CCTGACCCGC GGGGACGTCG AACTCCCCGG GTCCGTCCCC
75121 CGGATCGGGC TGGACGACAC AGAGATCCGC GCGTACATGA TCTACACCTC CGGATCCACC
75181 CCCGGCACGC CGGTGACCGA GGCCACCCC GCGTACATGA TCTACACCTC CGGATCCACC
75241 GGCCGCCCCA AGGGCGTCGT CGTCTCCAC GCGGCCATCG TCAACCGGT CGCCTGGATG
75301 CAGGCGGAGT ACCGTCTCGA CGCGACCGAT CGTGTCTTGC AGAAGACTCC GGCCGGTTTC
75361 GACGTGTCGG TCTGGGAGTT CTTCTGGCCG CTGCTCGAGG GCGCGGTCTT CGTGTTCGCC
75421 CGGCCCGGCG GCCACCGGGA CGCGGCGTAT CTGGCCGGAC TCATCGAGCG CGAGCGCATC
75481 ACCACGGCAC ATTTCTGTGC CTCCATGCTG CGCGTCTTCC TCGAAGAGCC CGGCGCGGCA
75541 CTCTGCACCG GACTGAGGCG GGTGATATGC AGCGGCGAGG CCCTCGGCAC GGACCTGGCC
75601 GTGGACTTCC GCGCGAAACT GCGCGTCCCC CTGCACAATC TGTAACGGCC GACCGAAGCG
75661 GCTGTGATG TCACCCACCA CGCGTATGAG CCCGCCACCG GCACGGCCAC GGTCCCCATT
75721 GGCCGCCCCA TCTGGAACAT CCGCACCTAC GTCCTCGACG CCGCCCTGCG TCCTGTGCCA
75781 CCGGGCGTGC CCGGCGAGCT GTATCTGGCC GCGCGCCGGC TGGCCCGCG CTACCACGGC
75841 CGCCCGGCAC TGACGGCGGA GCGGTTTGTG GCGTGTCCGT TCGGTGTGCC GGTGAGCGT
75901 ATGTATCGCA CGGGGGATTT GGTGCGGTGG CCGGTGGACG GCACGCTTGA GTTTGTTGGT
75961 CGTGCGGATG ATCAGGTGAA GGTCCGTGGT TTCCGTGTGG AGTTGGGTGA GGTGGAGGGT
76021 GCTGTTGCGG CGCATCCTGA TGTGGTGCGT GCGGTTGTTG TGGTGCCTGA GGACCGGCCG
76081 GGTGATCACC GGTGTTGGT TTAGGTGACT GTGGGTGGTG TTGGTGGGGA TGGCCTTCGT
76141 TCCGCGATCT CTGGTCTGGT GGCTGAGCGT CTGCCTGCGT ACATGGTGCC GTCGGCGGTG
76201 GTGGTTCTGG ATGAGATCCC GTTGACGCCG AACGGGAAGG TGGACCGGGC GGGTCTTCCG
76261 GTGCCGGTGG TGTGCGTGGC GGGGTTCTGT GCGCCGTCTG CGCCGCGGGA GGAGGTGTTG
76321 TGTGGTCTGT TCGCGGAGGT GCTGGGTGTT GAGCGGGTGG GGGTGGACGA TGGGTTCTTC
76381 GATCTGGGCG GGGACAGCAT TCTGTGATT CAGTTGGTGG CCGGGGCTCG TCGGGCGGGT
76441 CTGGAGTTGT CGGTTCCGGA TGTTTTCGAG GGCCGTACCG TACGTGCTCT GCGGCTGTG
76501 GTGCGTGGTT CGGACGCTGG GCGGTTGGT GTGGTGGGGG GTGCTGAGAT TGTGCTGCCG
76561 GGTGTGGGTG AGGTGGAGCG GTGGCCGGTG GTGGAGTGGC TGGCGGAGCG TGGTGGGGGG

76621	TCGCTGGGTG	GTGTGGTTTCG	GGGTTTCAAT	CAGTCTGTTG	TGCTTGCTGT	GCCTGCTGGG
76681	TTGGTGTGGG	AGGAGTTGCG	GGTGTGTGTTG	GGTGCGGTGC	GGGATCGGCA	TGAGGCGTGG
76741	CGGTTGCGGG	TGCTGGATTTC	CGGGGCGTTG	TGTGTTGATG	GTGTTGTTCC	GGATGACGGG
76801	TCGTGGATTG	TCCGGTGTGA	CCTGAGCGGT	ATGGGTGTGG	ATGGTCAGGT	GGATGCTGTG
76861	CGGGCTGCGG	CTGTGGAGGC	GCGTGCGTGG	CTGGATCCGT	CGGTGGGCCG	GGTGGTGCGG
76921	GCGGTGTGGC	TGGAGCGTGG	TGGTGATCGT	TCGGGGGTGT	TGGTGCTGGT	GGCGCATCAC
76981	CTGGTGGTGG	ACGGTGTGTC	GTGGCGGGTG	GTGCTGGGGG	ATCTGGCGGA	GGGGTGGGCG
77041	CAGGTGCGTT	CGGGTGGCCG	TGTGGAGTTG	GGTGTGGTGG	GGACGTGCTT	GCGGGGTTGG
77101	GCGGCGGCGT	TGGCGGAGCA	GGGCCGCGCG	GGCGAGCGTG	CGGGGGAGGT	GGAGTTGTGG
77161	TCGCGGATGG	TTCGGGGTGC	GGATGTTCTG	GTGGGGTCGC	GTGCTGTGGA	TGGTGCGGTG
77221	GATGTTTTTCG	GCGGGGTGGT	GTCGGTTGAT	TCGCGGGCGT	CGGTGTGCGT	GTCGCGTGCG
77281	TTGCTGACGG	AGGTGCCGTC	GGTTCTGGGT	GTTGGTGTGC	AGGAGGTGTT	GCTGGCGGCA
77341	TTCGGGCTGG	CGGTCGCGCG	GTGGCGCGGC	CGGGGTGGGC	CGGTTGTGGT	GGATGTTGAG
77401	GGGCACGGGC	GTAATGAGGA	CGCTGTGCGG	GGCGCTGATC	TGTCTCGTAC	TGTCGGTTGG
77461	TTCACCAAGT	TGTATCCGGT	CCGTGTGCCG	GTGGAGTCCG	CTTCGTGGGA	CGAGGTGCGT
77521	GCGGGCGGTC	CGGTGGTGGG	CCGTGTGGTG	CGTGAGGTGA	AGGAGACTCT	CGGTTGCGTG
77581	CCTGACCAGG	GTCTGGGTTA	TGGCATCCTG	CGCTATCTCG	ATCCCGAGCA	CGGTCTGCTG
77641	CTGGCCCGGC	ATGCCACCCC	GCAGTTCGGT	TTCAACTACC	TCGGCCGCTT	CACCACCGGA
77701	ACCGACGAAA	CCACCACGGC	CGACGCCCTC	GACCGGGCCC	CCGCGTGGAG	CCTTCTCGCC
77761	CGCAGCGCCG	CCGGCCAGGA	CCCCGAAGCT	CCCGTGGCGC	ACGCGTTCGA	GTTCAACGCG
77821	ATCACGCTGG	ACACCCCGGA	GGGCCCCGCG	CTGGGCGTGA	CATGGTCTGT	GCCGACGACG
77881	CTGCTGCCGG	AGTCCCGGAT	ACGGGAGCTG	GCCCCGCTACT	GGGACGAAGC	CCTGGGAAGG
77941	CTGGTTCGAAC	ACGCCCGGCA	CCCCGAAGCC	GGCGGCCTCA	CGCCGTCCGA	CGTGGGCCCTC
78001	GCGGAACCTCT	CCTTTGCTGA	GATCGAACTG	CTCGAAGACG	ACTGGAGGAC	ACAGGGATGA
78061	CGCAGCGCGC	GATGGAGGAC	ATACTTCCTC	TCACTCCGCT	GCAGGAGGGA	CTGCTGTTCC
78121	ACAGTGTTTA	CGACGAGCAG	TCCGTCGACG	TGTACACCGT	GCAGGTGGTC	GTCGACCTCG
78181	AGGGGCCCGT	CGACCCCGAA	GCACTGCGCG	CCGCCGCGGC	CGCCCTGCTG	CGTCGGCACG
78241	CCAACCTGCG	GGCGGCCTTC	CGGTACGAGC	GGCTGCAGCG	CCCCGTGCAG	ATCATCCCGC
78301	GCGAGGTTGC	GGTGCCGTGG	GAGCACACCG	ACGTCGCGAA	GCTCGAGGGC	GCCGAGCAGA
78361	AGGCCGAGAT	CGAACGCCTG	CTGCACGACC	AGCGGTGGCG	CCGCTTCGAT	CTGACGGCTC
78421	CGCCCTTGCT	GCGGTTCCCTG	CTCGTGCGCA	CAGGCCACGA	CCGGCACCGT	TTCGCGCTGA
78481	CTTTCATCA	CATCCTCATG	GACGGCTGGT	CGATGCCCGT	CCTGCTGCGG	GAACTCATCA
78541	CCCTCTACCG	CACCGGCGAC	GAGACCGCCC	TGCCCTGGGT	CCGGCCGTAC	CGGGACTACC
78601	TGGCTGGAT	CTCCCGCCGC	GACCGGGACG	AGGCCGGGCG	GGCCTGGTCC	AAGGCCTGG
78661	CTGGGGTTGA	CGAGGCCACC	CTCGTCGCCC	CGGGTGCCGA	CCGGGCCGCG	GAGCCGCGCG
78721	TGTGGACCGA	GTCCCGGCTC	GAACCGGACC	TGGCGGCGAC	GCTCGCCGCG	CGCGCCCGCG
78781	AGTTCGGCGT	CACCCTCAAC	ACCCTCGTCC	AGGCCGCGCTG	GGCGCTCGTC	CTCGGCCGCG
78841	TCACCGGCCG	CGACGACGTC	GTGTTGCGCG	TGACCGTGTC	CGGCCGGCCG	CCGGAGCTCG
78901	CAGGTGTGCA	GGACATGGTG	GGCCTCTTCA	TCAACACCGT	GCCGCTGCGT	GCCGAGCTGC
78961	TGCCGCACGA	GAGCCTCCGG	GACTTCACCG	TCCGCTCCA	GCGCGAACAG	ATACAGCTCC
79021	TCGACCACCA	GTACGAACGA	CTGGCGGTCA	TCCAGCGGCT	CGCCGGCCGG	ACAGAACTCT
79081	TCGACACGGT	GATGGTCTTC	GAGAACTACC	CCGTGCGCCG	CGCATCCTCC	GCCGGCGCCG
79141	ACGGCCCCGC	GGCCGAACCC	CGGGTCGCCG	ACGTCCACGT	ACGCGACGCC	ATGCACTACC
79201	CCCTCGGTCT	GCTGGTCCCTG	CCCGGCCCGC	CGCTGCGCCT	GCGCTTTGGC	CACCGGCCGA
79261	GCGCCCTGCC	CGCCGAACGC	GTCACGACGA	TCCGCGACAG	CCTCGTGCGA	GCCCTGGAGC
79321	TCATGGCCGA	CCAGCCGGAC	CTCGCCGTCG	GCAGGGCCGA	CATCCTCGGC	GAGGAGGAGA
79381	AACAGCATCT	CCTCACCGGC	CTCAACGACA	CCCACCGCGA	CGTGCCCCCG	CTCACCGTGC
79441	CCGGAATGAT	CGAGGCCAG	GCGGCCCGCA	CCCCCGGACG	GCCGGCGGTC	CATGCCCGCG
79501	ACGGCGAACT	CTCCTACGCC	GAACTCAACG	CGCGCGCCAA	CCGGCTGCGA	CGCCACCTCG
79561	CCGCGGCCGG	CGTGGGCCCG	GAGCAGTACG	TCACCTGCT	GCTCCCGCTC	TCCGCCCGCA
79621	TGGTCGTGGC	CGCTCTCGCC	GTGATGAAGA	CCGGCGCCCG	GTACGTTCCC	GTGGACCCGG
79681	AGTATCCGGC	CGACCGCATC	GCGTACATGC	TTGGCGACAT	CGGCCCGCGG	CTCGTCCTCA
79741	CCGACTCCCG	CTCGGCCGCG	GCCATGCCCG	CCGGCCCGGC	CCGCGTCTC	ACCCTCGACG
79801	ACGACGCCCT	CGACACGGGC	GTTGCGGCCG	TGCCCGAACA	CGACCTCGGC	ACCGACGGTA
79861	TCGCGCCGCT	TCCCGACCAG	CCCGCGTACG	TCATCTACAC	CTCGGGCTCC	ACCGGCCGCG
79921	CCAAGGGCGT	CGTGATCCTG	CACCGTTCCG	TCACCGGCTA	CCTCCTGCGC	ACGATCGAGG
79981	AATACCCCGA	AGCCGCCGCG	AAGGCATTCTG	TGCACTCGCC	CGTGCTCTTC	GACCTCACCG
80041	TCGGAGCGCT	GTACGCACCC	CTGGTGAGCG	GTGGCTGCCT	GCGCCTCGGA	TCGTTACCG
80101	ACGACAAGAT	CCTCGACCTG	GGCGAGGACA	GCCCCACCTT	CATGAAGGCC	ACCCCCAGCC
80161	ATCTCGCCGT	CCTCGACTCC	CTCCCCGACG	AGATCTCCCC	CACCGGGGCC	ATCACCTCG
80221	GCGGTGAGCA	ACTCCTGAGC	GAGACCTCG	ACCCGTGGCG	CGCCCGCCAC	CCCGGCGTGA

80281	CCGTCTTCAA	CGTGACGGC	CCCACCGAGA	CCACGATCAA	CTGCGCCGAA	CACCGCATCG
80341	CCCCCGGCAC	CACCCTGCCT	CCCGGCCCCG	TCCCCATCGG	CCGGCCCTTG	TGGAACACCC
80401	GCCGTGACGT	CCTCGACGGC	GGCCTGCGCG	TCGTGCCAC	GGGCGTCGCC	GGCGAGCTGT
80461	ACGTGGCCGG	CGCGGGCCTG	GCCCGCGGCT	ATCTCGGACG	CCCCGGCCTG	ACGGCCGAAC
80521	GCTTCGTGGC	CTGCCCCCTC	GGCGCACC GG	GCGAACGCAT	GTACCGCACC	GGTGACCTGG
80581	TGCGGTGGAG	AACCGACGGC	ACGCTGGAGT	TCGTCCGGCCG	CGTCGACGAC	CAGGTCAAGG
80641	TACGCGGCTT	CCGCATCGAG	CTCGGTGAGG	TCGAGGCCAC	CGTCGCCGCC	ACCCCGGGTG
80701	TGGCGCGCGC	GATCGTCGCT	GTCCGCGAGG	ACCGCCCCGG	CGACCAGCGG	CTCGTGCGGT
80761	ACGTGACACC	TGCCGACGTC	GACCCACCG	GCGGCCCTGCC	GTCGGCGGTG	ACCGCCCATG
80821	CCGCCGCCCC	CCTGCCCCGG	TACATGGTGC	CGTCCGCCGT	CGTGGTACTG	CACGAGGTAC
80881	CCCTCACCCC	CAACGGCAAG	ATCAACAGGG	CGGCCCTGCC	CGCGCCCGAG	GCCGTCTCCG
80941	GCGCCGGCTT	CCGTGCCCCC	GGCACGGCCC	GTGAGGAAGT	TCTGTGCGGC	CTGTTCGCCC
81001	AAGTCCTCGG	CCTCGAACGG	GTGCGCACGG	CCGACGACTT	CTTCGAATC	GGCGGCCACT
81061	CGCTGCTCGC	CACCCGCCTG	GTGTCCCGCG	TCCGTTCCGT	CCTCGGCGTC	GAACTCGGCG
81121	TCCGCGCCCT	CTTCGACGCC	CCTACCCCCG	GCCGCCTCGA	CCGGCTCCTG	GGGGAACGCT
81181	CCGGCGCCCC	CGTCCGCGCC	CCCCTGACCG	CGCGGGAACG	CACCGGGCGG	GACCCCTGT
81241	CGTACGCCCA	GCAGCGCCTG	TGGTTCCTCC	ACGAATCGA	GGGCCACGGC	GCCACATACA
81301	ACATCCCTCT	CGCGCTGCGC	CTCACC GGTC	CTCTCGACGT	GACCGCCCTC	GAAGCCGCCC
81361	TGACGGATGT	CGTCGCCCCG	CACGAGAGCC	TGCGCACACT	CATCGCCCCG	CACGGCACCG
81421	GCACCGCGTG	GCAGCACATC	CTGCCCCACC	GCAGACCCTCG	CGCCCCGAATC	ACCCTTGAGG
81481	CCGTACCCCT	GCACAGGGAC	GAAGTGGCCG	GGCGCCTCGC	CGAAGCGGCC	CGCCACCCCT
81541	TCGACCTCAC	CGCCGAGATC	CCCGTCCGCG	CCACCGTCTT	CCGCACCGAG	CGCGACGACC
81601	ACACCTTGCT	CGTCGTCACC	CACCACATCG	CAAGCGACCG	TTGGTCCCGC	GAGCCGTTCC
81661	TCCGTGACCT	GTCCGCCGCC	TACGCAGCCC	GGCGCGCACA	CTCCGCGCCG	GAAGTCCCCC
81721	CGCTGTCCGT	GCAGTACGCT	GACTACGCCG	CCTGGCAGCG	CGACGTACTC	GGCACCGAGG
81781	ACGACGGGAC	GAGCGAGATG	GCCGGCCAGC	TCGCCCCTG	GCGGGGACAG	CTCGCCGGCC
81841	TCCCGCAGGG	CCTGGACCTG	CCCACCGACC	GCCCCCGACG	CCCCGACGTC	GGCCGCCGCG
81901	GCGGCCGGTG	CCGGCTGGAG	ATCCCCGCCG	CGCTGCACCG	CGACATCGTC	ACCCTCGCCC
81961	GCCTCACCAG	TACCACCGTG	TTCATGGTGG	TCCAGGCGGC	CCTCGCCGGT	CTGCTGTCCG
82021	GGCTGGGCGC	GGGCACCGAC	ATCCCCATCG	GCACGCCGAT	CGCGGGCCGC	ACCACGAGG
82081	CCACCGAGCA	CCTCATCGGG	TTCTTCGTGA	ACACCCTCGT	CCTGCGCACC	GACGTCTCCG
82141	GCGATCCGAC	GTTCCGCCAA	CTCCTCGCGC	GCGTGCGGGC	CACCGACCTC	GACGCTACG
82201	CACACCAGGA	CGTGCCCTTC	GAACGCCTGG	TGGAGGTCCT	CAACCCGAA	GCGTCACTGC
82261	TGCGCCACCC	CCTCTTCCAG	ATACTGCTCG	CCTTCCAGAA	CACCGAGGAC	CGCAGCATCT
82321	CCGACCGCCC	CGGGACCCCTG	CTGCCCCGACC	TGCAGGTCAC	CGAACAGCCC	CTCGACGCCG
82381	GGACGGCCAA	GTTGACCTTC	GCGTTCGCGT	TCACCGAGCG	GCCCCCGGAG	AAGGGCGAAC
82441	CCTCCGGCAT	CACCGGAATC	GTGCAATACC	ACGCCGACCT	GTACGACGAG	GGCACCGTCC
82501	GGCAGATCGC	GGACTGCTTC	GTGCAGTTCC	TCGACGCGGC	CGTCCACGCC	CCGGGCACCC
82561	GCGTCGACGC	GGTCGGGCTG	CTCCCGGAAC	ACACCCTCCA	CAAAGTGTG	ACCCGCAGCC
82621	GCGGCACTGT	CACCGGCCTG	CCGCCCCCCA	CCCTGCCCCG	GCTGTTTCGAG	GCCCGGGTGG
82681	CGGCGCACCC	CGGTCACATC	GCGGTCGAGG	TCGCCGGCCG	CCGGCCCCGC	ACTACGACGT
82741	ACGACGCACT	GAACCGGCGG	GCCAACCGGC	TCGCCCGGCT	GCTCACCGAC	CGGGGCGTAC
82801	GGCCCGAACA	GCGCGTGGCG	ATCGCCCTGC	CCCGCTCCGC	GGACCTGGTG	ACGGCCTGGC
82861	TCGGGATCCT	CAAGGCCGGC	GCCGTGTGCG	TGCCCCGTCG	CCCCGCCTAC	CCCGACGACC
82921	GCATCGCCCA	CATGGCCGCC	GACGCGGCC	CGGCGCTCCT	CATCGCCTCC	GCAGCCACCC
82981	GCGACCGCAT	GCTCCCCACC	GGCATCCCCG	TACTGGACCT	CGACGACCCG	CCCGTCACCC
83041	CCGCACTCGC	CGCCGCGCCC	GACGGCAATC	CGCGCGGCAC	GGGACTGCTG	CCCCGCCATC
83101	CCGCCTACGT	CATCTACACC	TCCGGCTCCA	CCGGCACACC	CAAGGGCGTC	GTCGTACCCC
83161	ACGAAGGCAT	CCCGGCGCTG	GCCGCCACCC	AGCAGGAGGC	ACTGCGCGCG	GGCCCCGGAG
83221	ACCGGGTCCT	GCAACTGGTG	TCGACCAGCT	TCGACGCCTC	CGTCTGGGAC	CTGTGCTCCG
83281	CGCTGCTGTC	GGGCGCGACC	CTCGTCTCTG	CCCCGGACGC	GGACCTCTTC	GGTGACGAAC
83341	TCGCCGCCGC	GCTCACCGCA	CACCGCATCA	CGCACGTCAC	CCTGCCCCCG	GCCGCGCTGG
83401	CCGCTGTCCC	GGCAGGCGCG	GCACCCCCC	GGCTGACGGT	CACCGTCACC	GGCGACGTGT
83461	GCGGACCCCA	ACTCGTCGAC	CGCTGGGCCC	GTGGCGAACG	GCGGATCCTC	AACGGCTACG
83521	GGCCACCGA	GGTCACCGTC	GGCGCCACCT	ACGCCGTGTG	CGAACGGACC	GGTGACGGCG
83581	CGCCCGTGCC	GATCGGCGCA	CCCTGGCCCC	ACCAGCGTGT	GTACGTCCTC	GAACACCGGC
83641	TCCGGCCCCG	ACCCGCCGGC	TGCGTCGGCG	AGATCTACGT	CGCCGGGGCC	GGACTGGCCC
83701	GCGGTATCT	GGGCCGCCCC	GGACAGACCG	CCGAACGCTT	CGTCGCCGAC	CCCTTCGGCG
83761	CCCCGGCGA	GCGCATGTAC	CGCACC GGTC	ACCTGGCCCC	CCGCCGCAGC	GACGGCCACC
83821	TGCTGTTTCA	GGGACGCGCC	GACACGCAGG	TCAAAATCCG	CGGCTTCCCG	GTCGAATCTG
83881	CCGAGATCGA	GGCGGCCCTC	GCATCGCACC	CCGGCGTCCA	GGACGCGGTG	GTCACCGTGT

83941 ACGACGACGG GCTCGGCGAC CAGCGGCTCG TCGCGTACGT CACCGGCGGC CCCGGCACAC
84001 CGTCGGCCGC CGCGCTGCGC GCCCACCTGG CGTCCCGGCT GCGCCGGCAC ATGGTGCCCG
84061 GTGACGTCTT CACCTTGGAC GCCCTGCCGC TCACCGCCAA CGGCAAGGTG GACCGCACGG
84121 CGCTGCCCGG CCCCGGCACC CAGACCGCCG CCGCCGGGCG CGCACCCAG TCGCCGACGG
84181 AACGGGTGCT GTGCGCCTTG TTCGCGGACG TGCTCGGCCG GGAGACCGTC GGCGTGGACG
84241 AGGGGTTCCT CGACCTGGGC GGTCACTCGC TGCTCGCCAC TCGCCTCGCG GCGCGGTCC
84301 GCGCGGCGCT GGGCGTGGAG ATCTCCGTGC GCACCCTGTT CGAGGCGCCG ACCCCTGCCC
84361 TGCTCGCGTC GGCGTGACG GCGGACGCCG CGGCGTACGA CCCGTTCGAG ACGGTGCTGC
84421 CGCTCGGGCG CACGGGCAGC CGGCCACCGC TGTTCTGCGT CCACGCCGGA ATGGGCTGA
84481 GCTGGGCGTA CGCCGGCCTG CTCAGCCATC TGGACGCGGA CGTGCCGGTT TACGACTGC
84541 AGGCCCCGAG GCTCACCAGC CCCGGCGGGC TGCCCCGGAG CGTCGAGGAG ATGGCTGAGG
84601 ACTACGCCGG TGAGATCCGG CGCCTGTGCC CGGATGGGCC GTACCGGCTG CTCGGCTGGT
84661 CCTTCGGCGG CACGGTCGCC CACGCCGTGC CGACCCGCTT GCAACAGCAG GGCCACACCG
84721 TCGAACTCCT CGCCGTCTCT GACGCTACC CCGTCACCGG GGCCCGGCCC GACGCCGAGG
84781 TGGACGAACA GCGCATCGTC GCCGACTACC TCGCCAGCT CGGTTCCCCC GTCGCCCCCG
84841 AGCGCCTCGA GGGCGACGCG TGGTCCCCG AGTTCCTCGA GTTCGTACGG CGCACCGACG
84901 GGCCCGCGAG GGAATTCGAC GCCGGGCGGA TCCTCGCGAT GAAGGACGTC TTCCTCAACA
84961 ACGCCCGGCT CACCCGCCGT TTCACACCCG GCGTGTTTAC CGCGACATG GTGTTCTTCG
85021 CCTCCGCACG GCCCGGTTCC GAGCAGGCCG CCGAACGCGT CGGCCGTGG CACCCCGACG
85081 TCACCGGCGA CCTCGACCTG CACTGATCCG ACTGCGCACA CGAGGAGATG ACCGATCCAG
85141 CCGCACTCAC CCGGATCGGC CCCGTGCTCG CCGCACGGCT GGGCGCCGGC ACCTGACCCC
85201 CAGGACCCCA CACGGGACAC CGGACACGGG GGGCCCCCCC TGTCCGTACA CGAAAGGAAA
85261 CATAACGCCA TGGCCAACCC CTTCGAGAAC AACGACGGCA GCTACCTCGT ACTGGTCAAC
85321 GACGAGGGCC AGTACTCCCT TTGGCCCGCG TTCGCCGATG TCCCGGCGGG CTGGACCGTC
85381 ACCTTCGGCG AGAGCAGTCG GCAGGAATGC CTCGACCACA TCAACGAGAA CTGGACCGAT
85441 ATGCGCCCCA AGAGCCTCAT CCGGCAGATG GAGAACGACC GGACGACCGC GGCCTGACCC
85501 GCAGCCGGAC AGCGGAGACG GAAGGAGGGC CGACATGAGG GCGACATCCA GGATGATCCA
85561 GGTCAACGGC GCCCGGATCG CCTGCTCCGA CAGCGGCTGC GGTGACCCGG TGCTGATGAT
85621 CGCCGGCACC GGCAGTACCG GCCGGGTGTG GGACGCCTAC CAGGTGCCTG ACCTGCACGC
85681 GGCCGGATTG CGCACCATCA CGTTCACCAA TCGCGGCGTA CCGCCGTCCG ACGAGTGCGA
85741 GCGGGGCTTC ACCCTCGCCG ACCTCGCCGC CGACACCGCC GCGCTGATCG AACAGGTGGC
85801 GGGCGGACCC TGCCGCGTCG TGGGCACGTC CCTGGGCGCC CAGGTGGCCC AGGAAGTCGC
85861 CCTGGCCCCG CCGGACCTGG TGACCCAGGC GGTGTTTCATG GCCACCCGGG GTCGCACCGA
85921 CGCGATGCGG GCGCGCCCA CAGGGCGGC CGCCGCCCTG TACGACAGCG GCGTCGAACT
85981 GCCCCCGCC TACGCGGCG CTGTCCGCGC GCTGCAGAAC CTCTCCCCC ACACCCTCCG
86041 GGACCGCCAT CAGGTGAGG ACTGGCTCCC ACTCTTCGAG TACGCCGAAC GGGACGGGCC
86101 GGGGGTCCGT GCGCAGTTGG AACTCGGCCT GCTGCCCCGAC CGCCTCGCGG ACTACCGGGA
86161 CATCACCGTC CCCTGCCTGG TCATCGCGTT CGAGGACGAC GTCGTACCC CGCCGTACCT
86221 GGGCCGCGAA GTGGCCGACG CGATCCCCGG CGCCCGCTTC GAGACCGTTC CCCGTGCGG
86281 CCACTACGGC TACCTCGAGG ATGCGAGCGC GGTCAACAAG ATTCTTCGCG ATTTCTTCCG
86341 AACGAGCTGA AAGGCACGAC GACCTTGTCC AGTACCGGCA GAGAGGGGCC CGTCGTGACC
86401 GGCGAAACCC GCACCACCAC CTACCTCCCC GGCATGACCG TGCACGACTA CCACGTGACC
86461 GTCAAGGAAC AGCACCCGGC GCTCTTCGAG CTCCTGGACC CCGCACGCCT CGTCGCGGTC
86521 ACGGACGAGC CTTGGGTAC GGAGGGAAAC GAGTTCGACG ACGACCACGC CGGCCCGGCG
86581 GTCTCCTACC GCTGTGCCCA GCAGCACGGC GAAGCCCGCC GCACCGGCAT TGAGACGATT
86641 CTCGGCATGT TCGCCGGCCC CCGCGGGCTG CGCGACATGG GCCGTGTCTT CGATGTACTC
86701 GGAGGCGAAG GCCTGCTCAG CGGCGTGTGG CGGCAACTGG CCGGCGCCGG CGACGGGGAC
86761 TCCGTGCCAC TGGTCACCGG AGACCTCAGC GGCCACATGG TGGCCGACG CCTCCGGTCC
86821 GGCCTGCCCG CCGTACGCCA GCGGCGGAC CGCATGCTGC AGCGAGACCA CTGCCTGGAC
86881 GGCGTGCTCT TCGCGTACGG CACTACCCAC GTCGACCGCT CTGTACGCCC CCGCATGCTG
86941 ACAGAGGCCT CCCGGGTCTT GGCCCTGGA GGCGCGGTCG TCCTCCACGA CTTCGCGGAG
87001 GGATCCCCCG AAGAACGCTG GTTCCGCGAA GTCGTCCACC CCCGCTCCCT CGCGGGCCAC
87061 GCGTACGACC ACTTCACCGC CCACGAGATG ACCGGCTACC TCGCCGACGC GGGCTTCACC
87121 GACATCACCG TCGGCCCCGT GTACGACCCG ATGACCCTGA CCGGGGAGAC CGACGAGAGC
87181 GCACTGGCTC GGCTCGTCTC CTACATGACC TCGATGTACG GCATCCTGCC CGACGGCGAC
87241 CGGAGCAACG AGCGGACGGA AGCCGCCCTC CGCGACATCT TCGTTTCTC GCGCGGCGAC
87301 CTCCCCGAGG ACGTCCCCCG CGACGAGGCG GTCCTGGAAC TTACCGTCCG TCCGCACGGC
87361 AATGCCTTCC GGGCCGAGCT CCCCCGGATA GCCCTCGTCG CCCACGGACG CAAACCATGA
87421 CAGCGCAGGA CACCCGGACG ACCGGGAGTG ACGGTGGCGG CCGGGGCGCC ACGTACCACG
87481 AGAGCCCGAC CTACGGGGAG CTGCTGCGCC TGGAGGACCT GCTGAACGTC GCGCACCTGC
87541 GCGACGCGGC CGCCCCGGTC CTCTTCCTTG CCACGCACCA GTCGGCGGAG ATCTGGTTCCG

87601 GCATCGTGCT GCGCCACCTG GAGGAAATCC GCGCGGCCCT CACGGACGAC GACCCGGACA
87661 CGGCACTGCA TCTGCTGCCG CACTGCCCG AGATCTTCGA ACTGCTCGTC CGCCACTTCG
87721 ACATGCTGGC CACGCTGAGT ACGGAGGAAT TCGGCAAGAT CCGCGCGGGG CTGGGCACGG
87781 CGAGCGGCTT CCAGTCGGCG CAGTACCGGG AGATCGAGTT CCTGTGCGGT CTGCGCGACC
87841 ACCGCCACAT CTCCACACCG GGCTTCACGG AAACCGAACG TCGGCGACTG CGGGAACGGG
87901 CCCGCCAGCC CTCCGTGGCG GAGGCCACG ACGCCTTCCG GACCCGATGC GCCAACGGGA
87961 AGGACGCGGA ACGGATCGGG GAAGCGCTCC TGAGGTTCTGA CGAACGGGTC ACCGTCTGGC
88021 GCGCCCGCCA CGCGGCCCTG GCGGAACGCT TCCTGGGCCC CTTGAAGGG ACGGCCGGCA
88081 CCGCCGGAGC CGACTACTTG TGGCGGGTCA CCCGGCACAG GCTCTTCCCC CCGGAGGCGT
88141 GGGGCGCCGG CTGACGGCAC CGCCCCGGCC CCGGGGACGG GACAGGCCGG TTCCCGCACC
88201 CCGGCCCCGG GGGCGGAAA CGGCCTTGCC GTGCCGTCAG AAGGCCGTCA ACCGGTCCCA
88261 CACGAGGGTC CGAGCCCTTC GTCGAGCAAG CGTCGCCACT CTGACGTTCT GTCTGTGCAC
88321 GCTCATACCG GCGGGCACCG TCACGGCCAC CGGCACCCTG GTCAGGAAGC TGGTGAAGGA
88381 GCGGGCGAG GAAGCGGCCG GTTGCATCAT GTCCACGCTG ACCGAGCCCG CAGTGCAGGC
88441 GATCGAGAAC GTCGCCGCCG ACCTGGCGGT TCAGGCCGCA GCCAACGCGG TCGGGCTGCA
88501 GAACGGGATC GACACCGGTC AGGCCGTCCA CGCCGGCAAAG GAGGGGTTCC AGGACGGAGT
88561 CGCGGGTGCG AAGGAAGGAC TGCGACTCGC CTCGGTGGAC GGCGGTCCGC CGCCGGGATC
88621 GACGGGCCGG CTGATGGGCG ACCTCAAGGC GACCAAGGGC TTTGGCGACC ATCGGGCGCC
88681 AAGACGTGCA AGAACGACCC CGTGGACGTC GCCACCGGTG AGATGCTGCT CCCGCAGACC
88741 GTCTTGGGGC TCCCCGGCGT CCTGCAGCTG GTCCTGGGGC GGACTCATCC GTGCTCGGCT
88801 GCCGACCGTC GGCAACCTCG ACGCCCGCGT GAACTCATCA GGTTCGCCAC TCGGTTTAC
88861 CTGCGACGCC GATGGACGCG TCACCTCCTG GACCGACTGC AACGACGCCA CTTTCCGTAC
88921 GTCTACGACC AGGCCGGCCG GGTGGTGC GGACGAGGCC CCGACGGCAT CCTCTCCTCG
88981 TCCTGTGCCT ATGGAGAGCC GGAACCCGAC ACCGGGCCGC GCACGACGCG CAAGGGCGGC
89041 CCGTGATCCG TCGGAGCTCA ACGCTCCGTG GCCGCACAGC GGTCTGGCAC TTCACCTGGG
89101 ACGCTCAGGA CCGGCTCGCC GAGGCCGCCG ACCACTGCTG GGACTGCGCG CCGTTCCCGC
89161 GCCTGAGAGG GGGCGCGCTG ACCTTGGTCA GAAGCCTTGC GCGATCACGA GCGTCACGTT
89221 GCGGGCGAC TTGCTCCGCC GAGTTGAAGC CGTACGAAGT GCCGACTGGA TCGATGCGGC
89281 CAGCCATCCT CAGGGTTGCC CTGACAGAGT TTGGTGGACA CGAACCGGAA CAACCCGGAC
89341 CGCCAGATGT GGTGAGTTT TCCGGGACAG TTGATCCACA CACCGTCCCC CCGCTCAGTC
89401 CCGTTCGGAC CATGAGCCTC ACCGAAGGAC CCCACCTGCA CATGAGCACC CCCGACACTC
89461 GTCCCCCGCG GTGGCCTTTA CCTCCCGCCC CGCCCGCCCA TGACCCCGTC CTCTTCGCGC
89521 GGGCGATGCG GGACATGCGC CTGACGTGGC GTGCCCGCGG GATCCTGGCC GAACTCTCCG
89581 TCGGCTACGG CCCCCGGCAG GACCCACGTA TCAGCGAGCT GGTGCGGCTC AACCGCGACG
89641 AGCGTCTGGC TGCAGAGGGC CGCGAGGCCT TCCGCACGGC GGTCCGTGAG CTGCGCGGCC
89701 TCGGCTACCT CACTCCGGAC GCCACCACGG CCTCCGGCGT CGGGGAGCGT CTGATCGTCG
89761 ATCTCGCCCC GGCCGCGGAA GCCTGGCTGA TCCCGCAGCA ACCCGGTTTC GGGTTCTACG
89821 TGGACGGGAG CTGACCGTCC GACTCTCTGC CGGGCGTGGC CGAACGCCGT CCTCGTGTGA
89881 CGGGGACGGC CCGCCTATCC TGCGGGCATG GCCCAACCCA TTGAACCTGT CATATTCGAC
89941 TGCGACGGCG TACTCGTCGA CAGCGAACGC ATCGCGGTGC GCGTGGACGC ACTCGTCCTG
90001 GCCGAGCTGG GGTGGAATCT CACCGAAGCC GAGATCGTCG ACCGGTTTCAT GGGCCTGTCTG
90061 AGCCGGTCTGA TGACGCGGCA GATCGAGGAC CACCTCGGGC GCCGTCTGCC GGCCGACTGG
90121 GAGGAAGAGT TCAAGCCCCT CTACGACGAG GCGCTCGCCG CCGAACTCAC GCCGGTCGAG
90181 GGCATCGTCG ACGCCCTCGA CGCGCTCACG CATCTCCCCA CTTGTGTGGC ATCCAGCGGG
90241 AGCCACGACA AGATGCGTTT CACGCTGGGG ATGACCGGTC TCCGCCCGCG CTTCTGAAGGC
90301 CGCATTTTCA GTGCCACCGA GGTCGAGCAC GGCAAGCCGG CCCCAGATCT GTTCTACAC
90361 GCCGCGCGGA AGATGGGGGT CGTGCCCGAG GCGTGCGCCG TGGTCGAGGA CAGTCAGTAC
90421 GGTCTTCAGG CAGCCCGGGC CGCGGGCATG CGAGCCTTCG CTTACGCCGG GGGACTGACT
90481 CCCGCGGACC GTCTCGAAGG CCCCAGCACC GTCGTCTTCG ACGACATGCG CAGACTGCCC
90541 GGCTCCTCG CGGATCACTG ACCGCCGCTT GGATCACTCC ACTCCATCGG CCACTGT

SEQ ID NO: 2

Nucleotides 36018-36407 of SEQ ID NO: 1

SEQ ID NO: 3

Nucleotides 78057-85196 of SEQ ID NO: 1

SEQ ID NO: 4
Nucleotides 85498-86350 of SEQ ID NO: 1

SEQ ID NO: 5
Nucleotides 85498-85534 of SEQ ID NO: 1

SEQ ID NO: 6
Nucleotides 85535-86350 of SEQ ID NO: 1

SEQ ID NO: 7

MTQRAMEDILPLTPLQEGLLFHSVYDEQSVDVYTVQVVVDLEGFVDPEALRAAAAALLRRHANLRAAFRYERLQRP
VQIIPREVAVPWEHTDVAKLEGAEQKAEIERLLHDQRWRRFDLTAPLLRFLLVRTGHRHRFALTFHHILMDGWS
MPVLLRELITLYRTGDETALPWVRPYRDYLAWISRRDRDEAGRAWSKALAGVDEATLVAPGADRAAEPPLWTESRL
EPDLAATLAARAREFGVTNLTLVQAAWALVLGRLTGRDDVVFVGTVSGRPPELAGVEDMVGLFINTVPLRAELLPH
ESLRDFTVRLQREQIQLLDHQYERLAVIQRLAGRTELFDTVMVFENYPVAAASSAGADGPAAEPRVADVHVVDAMH
YPLGLLVLPGPPLRLRFGHRPSALPAERVTTIRDSLVRLELMADQPD LAVGRADILGEEEKQHLLTGLNDTHRDV
PPLTVPGMIEAQAARTPGRPAVHARDGELSYAELNARANRLARHLAAAGVGPEQYVTTTTLLPLSARMVVAALAVMKT
GAAYVPVDPEYPADRIAYMLGDIGPALVLTDSRSAAAMPAGPARVLTLDLDDALDTGVRALPEHDLGTDGIAPLPDQ
PAYVIYTSGSTGRPKG VVILHRSVTGYLLRTIEEYPEAAGKAFVHSPVSFDLTVGALYAPLVSGGCLRLGSFTDDK
ILDLGEDSPTFMKATPSHLAVLDSLDPDEISPTGAITLGGEQLLSETLDPWRARHPGVTVFNVYGPTETTINCAEHR
IAPGTTLP GPVPPIGRPLWNTRLVLDGGLRVVPTGVAGELYVAGAGLARGYLGRPGLTAERFVACPF GAPGERMY
RTGDLVRWRDGTLEFVGRVDDQVKVRGFRIELGEVEATVAATPGVARAIVAVREDRPGDQRLVAYVTPADVDPTG
GLPSAVTAHAAARLPAYMVPSAVVVLHEVPLTPNGKINRAALPAPEAVSGAGFRAPGTAREEVLCGLFAEVLGLER
VGTADDF FELGGHSLLATRLVSRVRSVLGVELGVRALFDAPT PGRDLRLLGERSGAPVRAPLTARETGRDPLSYA
QQRLWFLHELEGHGATYNIPLALRLTGPLDVTALEAALTDVVARHESLRTLIARDGTGTAWQHILPTGDPRARITL
EAVPLHRDELAGRLAEAAARHPFDLTAEIPVRATVFRTERDDHTLLVVTHHIASDRWSREPFLRDL SAAYAARRAHS
APELPPLSVQYADYAAWQRDVLGTEDDGTSEMAGQLAHWRGRLAGLPQGLDLPTDRPRRPDVGRRGGRCRLEIPAA
LHRDIVTLARVTSTTVFMVVQAALAGLLSRLGAGTDIPIGTPIAGRTDEATEHLIGFFVNTLVLR TDVSGDPTFAE
LLARVRATDL DAYAHQDVPFERLVEVLN PERSLLRHPLFQILLAFQNTEDRSISDRPGTLLPDLQVTEQPLDAGTA
KFDLAFATERPPEKGEPSGITGIVEYHADLYDEGTVRQIADCFVQFLDAAVHAPGTRVDVAVGLLPEHTLHKLLTR
SRGTVTGLPPATLPELFEARVAAHPGHIAVEVAGRRPATTTYDALNRRANRLARLLTDRGVRPEQRVAIALPR SAD
LVTAWLGILKAGAVCV PVDPAYPDDRIAHMAADAAPALLIASAATRDRMLPTGIPVLDLDDPAVTAALAAAPDGNP
RGTGLLPAHPAYVIYTSGSTGTPKG VVVTHEGIPALAAATQQEALRAGPGDRV LQLVSTSF DASVWDLCSALLSGAT

LVLAPDADLFGDELAALTAHRITHVTLPPAALAAVPAGAAPPRLTVTVTGDVCGPQLVDRWAGGERRI LNGYGPT
EVTVGATYAVCERTGDGAPVPIGAPWPDQRVYVLEHRLRPVPAGCVGEIYVAGAGLARGYLGRPGQTAERFVADPF
GAPGERMYRTGDLARRRSDGHLLFEGRADTQVKIRGFRVELAEIEAALASHPGVEDAVVTYDDGLGDQRLVAYVT
GGPGTPSAAALRAHLASRLPRHMPGDVLTLDALPLTANGKVDRTALPGPGTQTAAPGRAPQSPQERVLCALFADV
LGRETGVGDEGFFDLGGHSLLATRLAARVRAALGVEISVRTLFEAPTPALLASACTADAAAYDPFETVLPRLRTGS
RPPLFCVHAGMGLSWAYAGLLSHLDADVPVYGLQARRLTAPGGLPGSVEEMAEDYAGEIRRLCPDGPYRLLGWSFG
GTVAHAVATRLQQGHTVELLAVLDAYPVTGARPD AEVDEQRIVADYLAQLGSPVAPERLEGD AWLPEFLEFVRRT
DGPARDFDAGRILAMKDVFLNNARLTRRFTPGVFTGDMVFFASARPGSEQAAERVGLWHPHVTGDLDLHLIDCAHE
EMTDPAALTRIGPVLAARLGAGT*

SEQ ID NO: 8

See Figure 4, DptH sequence.

SEQ ID NO: 9

MDMQSQRLGVTAQQSVWLAGQLADHRLYHCAAYLSLTGSIDPRTLGTAVRRTLDETEALRTRFVPQDGELLQIL
EPGAGQLLLLLEADFSGDPDPERAAHDMHAALAAPVRLDRAGTATHALLTLGPSRHLLYFGYHHIALDGYGALLHLR
RLAHVYTALSNGDDPGPCPGPLAGVLTEEAAYRDSNHRDGEFWTRSLAGADEAPGLSEREAGALAVPLRRTVE
LSGERTEKLAASAAATGARWSSLLVAATAAFVRRHAAADDTVIGLPVTARLTGPALRTPCMLANDVPLRLDARLDA
PFAALLADTTTRAVGTLARHQFRGEELHRNLGGVGRTAGLARVTNVLAYVDNIRFGDCRAVVHELSSGPVRDFHI
NSYGTPTGTPDGVQLVFSGNPALYTATDLADHQERFLRFLDAVTADPDLPTGRHRLSPGTRARLLDDSRGTERPVP
RATLPELFAEQARRTPDAPAVQHDGTVLTYRDLHRSVERAAGRLAGLGLRTEDVVALALPKSAESVAILLGIQRAG
AAVPLDPHTPAERLARVLDDTRPRYLVTTHIDGLSHPTPQLAAADLLREGGPEPAPGRPAGNAAYIIQTSGST
GRPKGVVVTHEGLATLAADQIRRYRTGPDARVLQFISPGFDVVFSELSMTLLSGGCLVIPPDGLTGRHLADFLAAE
AVTTTSLTPGALATMPATDLPRLTLIVGGEVCPPEIFDQWGRGRDIVNAYGPTETTVEATAWHRDGATHGPVPLG
RPTLNRRGYVLDPALEPVPDGTGELYLAGELARGYVAAPGPTAERFVADPFGPPGSRMYRTGDLVRRRSGGMLE
FVGRADGQVKLRGFRIELGEVQAALTALPGVRQAGVLIREDRPGDPRLVGYIVPAPGAEPDAGELRAALARTLPPH
MVPWALVPLPALPLTSNGKLDRAALPVPAARAGSGQRPVTPQEKTLCALFADV LGVTEVATDDVFFELGGHSLNG
TRLLARIRTEFGTDLTLRDLFAFP TVAGLLPLLDNGRQHTTPPLPPRPERLPLSHAQQLWFLDQVEGSPAYNI
PTAVRLEGPLDIPALAVALQDVTNRHEPLRTLLAEDSEGPHQVILPPEAARPELTHSTVAPGD LAAALAEAARRPF
DLAGEIPLKAHLFGCGPDDHTLLLLVHHTAGDGASVEVLVRDLAHAYGARRAGDAPHFEPLPLQYADHTLRRRHLL
DDPSDSTQLDHWRDALAGLPEQLELPTDHTRPAVPTRRGEAIAFTVPEHHTHHTLRAMAQAHGVTVMVMQAALAAL

LSRHGAGHDIPLGTPVAGRSDDGTEDLVGFFVNTLVLRNDVSGDPTFAELVSRVRAANLDAYAYQDVPFERLVDVL
KPERSLSWHPLFQIMIAYNPATNDTADGSRFAGLTSRVHAVHTGMSKFDLSFFLTEHADGLGIDGALEFSTDFT
RITAERLVQRYLTVLEQAAGAPDRPISSYELLGDDERALLAQWNTAHPTPPGTVLDLLESRAARTPDRPAVVEND
HVLTYADLHTRANRLARHLITAHGVGPERLVAVALPRSAELLVALLAVLKTGAAYVPLDLTHPAERTAVVLDCCR
AVILTDAGAARELPRRDIPQLRLDEPEVHAAIAEQPGGPVTDRTCTVTPVSGEHVAYVIYTSGSTGRPKGVAVEH
RSLADFVRYSVTAYPGAFTVLLHSPVTFDLTVTSFLPPLVVGGAIHVADLTEACPPSLAAAGGPTFVKATPSHLP
LLTHEATWAASAKVLLVGGEQLLGRELDKWRAGSPEAVVFNDYGPTEATVNCVDFRIDPGQPIGAGPVAIGRPLRN
TRVFVLDGGLRAVPVGVVVGELHVAGEGLARGYLQGPGLTAERFVACPFGDAGERMYRTGDLVRWRADGMLEFVGRV
DDQVKVRGFRIELGEVEAAVAACPGVDRSVVVVREDRPGDRRLVAYVTAAGDEAEGLAPLIVETAAGRPLPGYMPVS
AVVVLDEIPLTPNGKVDRAALPAPRVAPAAEFRVTGSPREEALCALFAEVLGVERVGVDDGFFDLGGDSILSIQLV
ARARRAGLEVSVRDVFEHRTVRALAGVVRESGGVAAAVDMSGVGAVERWPVVEWLAERGGGGLGGAVRAFNQSVVV
ATPAGITWDELRTVLDVAVRERHDAWRLRVVDSGDGAWSLRVDAPAPGGEFDWITRHGMASADLEEQVNAVRAAAVE
ARSRLDPLTGRMVRAVWLDRGPDRRGVLLVAHHLVVDGVSRIVLGDLGEAWTQARAGGHVRLDTVGTSLRGWAA
ALAEQGRHGARATEANLWAQMVHGSDPLVGPRAVDPSVDVFGVSVGSRASVGSRRALLTEVPSVLGVGVQEVLL
AAFGLAVTRWRGRGGSVVVDVEGHGRNEDAVPGADLSRTVGWFTSIYPVRLPLEPAWDEIRAGGPAVGRTVREIK
ECLRTLDPDQGLGYGILRYLDPENGPALAQHPTPHFGFNYLGRVSVSADAASLDEGDAHADGLGGLVGGRAAADSD
EQWADWVPVSGPFAVGAGQDPVLPVAHAVEFNAITLDTDPGPRLSVTWSWPTTLLSESRIRELARFWEALEGLVA
HARRPDAGGLTPSDLPLVALDHAELEALQADVTGGVHDILVPSPLQEGLLFHSSFAADGVDVYVGQLTFTDLTGPD
ADHLHAVVESLVTRHDVLRGTGRQAQSGEWIAVVARQVHTPWQYIHTLDTADTLTNDERWRPFDMTQGPLARFTL
ARINDTHFRFIVTYHHVILDGWSVAVLIRELFTTYRDTALGRRPEVPYSPRRDFMAWLAERDQTAAGQAWRSALA
GLAEPTVLALGTEGSGVPIEVLEEEISEELTSELVAWARGRGVTVASVVQAALVLGRLVGRDDVVFGLTVSGRP
AEVAGVEDMVGLFVNTIPLRARMIPAESLGAFVERLQREQTELLEHQHVRLAEVQRWAGHKELFDVGMVFENYPM
SLLQDSLFGHSGQLQIDGIQGADATHFALNLAVVPLPAMRFRGYPDPVFDAGRRELWGWIIVRALECVVCERDVPV
SGVDVLGAGERETLLGWGAGAEPGVRALPGAGAGAGLVGLFEERVRTDPDAVAVRGAGVEWSYAE LNARANAVA
RWLIGRGVGPVGVVMDRGPVAVMLLAVAKSGGFYLPVDPQWPTEIDWVLADAGIDLAVVGENLAAAVEAVR
DCEVVDYAQIARETRLNEQAATDAGDVTGGERVSALLSGHPLYVIYTSGSTGLPKGVVTHASVGAYLRRGRNAYR
GAADGLGHVHSSLAFTLTVTVLFTPLVSGGCVTGLDLDLTANGLGATFLKATPSHLP LLGQLDRVLAPDATLLGG
EALTAGALHHWRTHHPHTTVINAYGPTELTVNCAEYRIPPGHCLPDGPVPIGRPFTGHHLFVLDPALRLTPPDIG
ELYVAGDGLARGYLGRPDLTAERFVACPFRRSPGERMYRTGDLARWRSDGTLEFIGRADDQVKIRGFRIELGEVEAA
VAAHPHVARAIAVVREDRPGDQRLVAYVTGSDPSGLSSAVTDTVAGRLPAYMVPSAVVVLQIPLTPNGKVDRAAL

PAPGTASGTTSRAPGTAREEILCTLFADVGLDQVGVEDFFDLGGHSLLATRLTSRIRSALGIDLGVRALFKAPT
VGRILDQLLQQQTSLRAPLVARERTGCEPLSFAQQLWFLHQLEGPNAAYNIPMALRLTGRDLTALEAALTDVIA
RHESLRTVIAQDDSGGVWQNILPTDDTRTHLTLDTMPVDAHTLQNRVDEAARHPFDLTTEIPLRATVFRVTDDEHV
LLLVLHHIAGDGWSMAPLAHDLAAYTVRLEHHAPQLPALAVQYADYAAWQRDVLGTENNTSSQLSTQLDYWYSKL
EGLPAELTLPTSRVRPAVASHACDRVEFTVPHDVHQGLTALARTQGATVFMVVQAALAALLSRLGAGTDIPIGTPI
AGRTDQAMENLIGLFVNTLVLRDVS GDPTFAELLARVRTTALDAYAHQDIPFERLVEAINPERSLTRHPLFQVML
AFNNTDRRSALDALDAMPGLHARPADVLAVTSPYDLAFS FVETPGSTEMPGILDYATDLFDRSTAEAMTERLVRLL
AEIARRPELSVGDIGILSADEVKALSPEAPPAEELHTSTLPELFEEQVAARGHAVAVVCEGEELSYKELNARANR
LARVLMERGAGPERFVGVALPRGLDLIVALLAVTKTGAAYVPLDPEYPTDRLAYMVT DANPTAVVTSTDVHIPLIA
PRIELDDEAIRTELA AAPDTAPCVGSGPAHPAYVIYTS GSTGRPKGVI SHANVVR LFTACSDSFDFGPDHVWTLF
HSYAFDFS VWEIWGALLHGGRLVVVPFEVTRSPA EFLALLAEQQVTLLSQTPSAFHQLTEAARQEPARCAGLALRH
VVFGEALDPSRLRDWFDLPLGSRPTLVNMYGITETT VHVTVLPLEDRATSLSGSPIGRPLADLQVYV LDERLRPV
PPGTVGEMYVAGAGLARGYLGRPALTAE RFVADPNRSRGGRLYRTGDLAKVRPDGGLEYVGRGRDQVKIRGFRIEL
GEIEAALVTHAGVVQAVVLVRDEQTD DQRLVAHVVPALPHRAPTLAELHEHLAATLPAYMVPSAYRTLDELPLTAN
GKLDRAALAGQWQGGTRTRRLPRT PQEEILCELFADVLRLPAAGADDDFFALGGHSLLATRLLSAVRGTLGVELGI
RDLFAAPT PAGLATVLAASGTALPPVTRIDRRPERLPLSFAQRLWFLSKLEGPSATYNI PVAVRLTGALDVPALR
AALGDVTARHESLRTVFPDDGGEPRQLVLP HAEPPFLT HEVTVGEVAEQ AASATGYAFDITS DTPLRATLLRVSPE
EHVLVVVIHHIAGDGWSMGPLVRDLVTAYRARTRGDAPEYTPLPVQYADYALWQH AVAGDEDAPDGR TARRLG YWR
EMLAGLPEEHTLPADRPRPVRSSHRGGRVRFELPAGVHRSLLAVARDRRATLFMVVQAALAGLLSRLGAGDDIPIG
TPVAGRGDEALDDVVGFFVNTLVLR TNLAGDPSFADLVDRVRTADLDAFAHQDVPFERLVEALAPRRSLARHPLFQ
IWYTLTNADQDITGQALNALPGLTGDEYPLGASAAKFDLSFTFTEHRTPDGDAAGLSVLLDYSSDLYDHGTAAALG
HRLTGFFAALAADPTAPLGTVPLLTD DERDRILGDWGS GTHTPLPPRSVAEQIVRRAALDPDAVAVITAE EELSYR
ELERLSGETARLLADRGIGRESLVAVALPRTAGLVTTLLGVLRTGAAYLPLDTGYPAERLAHVLSDARPDVLVTHA
GLAGRLPAGLAPTVLVDEPQPPAAAA PAVPTSPSGDHLAYVIHTSGSTGRPKGVAIAESSLRAFLADAVRRHDLTP
HDRLLAVTTVGFDIAGLELFAPLLAGAAIVLADEDAVRDPASITSLCARHHVTVVQATPSWWRAMLDGAPADAAAR
LEHVRILVGGEPLPADLARVLTATGAAVTNVYGPTEATI WATAAPLTAGDDRTPGIGTPLDNWRVHILDAALGPVP
PGVPGEIHIAGSGLARGYLRRPDLTAERFVANPFAPGERMYRTGDLGRFRPDGTLEHLGRVDDQVKVRGFRIELGD
VEAALARHPDVGRAAAA VRPDHRGQGR LVAYVVP RPGRPDAGELRETVRELLPDYMVPSAQVTLTTLPHTPNGK
LDRAALPAPVFGTPAGRAPATREEKILAGLFADILGLPDVGADSGFFDLGGDSVLSIQLVSRARREGLHITVRDVF
EHGTVGALAAAALPAPADDADDTVP GTDVLPSISDDEFEEFELELGLEGEEEQW*

SEQ ID NO: 10

Nucleotides 38555-56047 of SEQ ID NO: 1

SEQ ID NO: 11

VNRRSKVVEEILPVSALEQGLLFHSSFAAADGVDVYAGQLAFDLVGAVDTGRLRAAVESLVARHGVLRSYRQARS
GEWVAVVARRVATPWRAVDARDGATDAAAVAREERWRPFDLGRAPLARFVLVRTDDDRFRFVITYHHVILDGWSLP
VLLRELLALYGSADPSVLPVPRPYGDFLRWAAARDDAEAETAWRDALTGLDEPSLVAPGASPDGVVPASVHAELD
KAGTENLAAWARHRGITQATAVRAAWALVLGQHTGRDDVVFVTVSGRPAELAGAEHMGVLFINTVPLRTVLDPAD
TLGTFAARLQAEQTTLEHQHVRLSDIQRWAGHKELFDTIVVFENYPIGHSGPGSIRTDDFTVTATEGSDATHYPL
TLTAVPGETLRLKLDHRPDLVDTTATALLRRVTRVLETATDDTGHTLARLDLDDDERHRLLRGWNDDTREQPPT
YYHQEFEEQARRRPHDTALVFTSTSWTYEELNDRANRLARLLVAAGAGSDDFVALAFPRSAESVVAIVLAVLKAGAA
YLPDMDQPAERLTGILADAHPTVVLTTTATPLPHPGRTLVLDSPTTARALAAAPAHNLTADRRTPLNARNAAY
I IHTSGSTGRPKGVVIEHRSANLFDHRRALIEPHAAGGSRLKAGLTASLSFDTSWEGLICLAAGHELHILDDDT
RRDAERVAELIDRQRIDVIDVTPSFAQQQLVETGILDEGRHHPAAFMLGGEGVDAKLWTRLSDVPGVTSYNYGPT
FTVDALACTVGIAPRPVIGHPLDNTAAYILDGFLRPVPEGVAGELYLAGTQLARGYAGRPGLTAERFVACPFAPG
ERMYRTGDLVRRSPGGVVEYLGRVDDQIKLRGFRIEPAEIELALAGHPAVAQNVLLHRSATGEARLVAVVPGTP
VDPRELTHGLAARLPAYMVPSAFVLLDTLPLTPNGKLDRGALPEPAFGTAPRPERPRTVVEILCGLYADVLGLPS
FGADDDFFDAGGHSLLASKLVSRIRTNLKTENLRALFEHRTVSSALATLHRAAQAGPALTAGPRPARIPLSYAQR
RLWFLNRLDRDSAAYNMPVALRLRGPLDSTAMCAALTDVAERHEALRTVFEEDRDGAHQIVLPATGLGPLLTVTGA
DGTTLRALITEFVRRPFDLAAEIPFRAALFRVGDEEHVLVVVLLHHIAGDGWSMGPLARDVAEAYRARAAGRAPDWE
PLPVQYADYALWQREVLAEDDETGELSAQLAHWRTRLAGAPAEITLPTDRPRPAVASTAGDRVEFTVPAGLHQAL
ADLARAAGATVFMVQAALAVLLSRLGAGDDIPIGTPVAGRTDEATEELIGFFVNTLVLRDVSVDGPTFAELLARV
RATDLDAYAHQDVPPERLVEVLNPERSLARHPLFQVMLTFNVPMDMGVGSALGNLGELEVSGEAIRTDQTKVDLAF
TCTEMYAADGAASGMRGVLEYRLDVFGAVQARETTERLVRVLEGGVSGGGVSVSGVDVLGVGERERLLGWGVGGP
VPVVPGGGLVGLFEERVADADAVAVRGAGVWVSYGELNARVNVVARWLVRGVGAECGVGVVMGRGVDDVVMMLLA
VAKAGGFYVPDPEWPPERVGVWLADAGVGLVVVGEGLSHVVGDFFPGGEVFEFSRVVRESCLVELVAADGVEVRNV
TDGERASRLLPGHPLVYVYTSGSTGRPKGVVVTASVGGYLRGRDVAAGAVGGVGFVHSSSLAFDLTFTVLTPLV
SGGCVVGLGELDESAQGVGASFVKVTPSHLGLLGELEGVAGNGMLLVGGEALSGGALREWRERNPGVVVNAYGPT
ELTVNCAEFLIAPGEEVPDGPVPIGRPFAGQRMFVLDAALRVVPVGVVGEVYAGVGLARGYLGRAGLTAERFVAC
PFGAPGERMYRTGDLVRWRVDGALEFVGRADDQVKVRGFRVELGEVEGAVAAHPDVVRVAVVVVREDRPGDHRLVAY
VTGVDTGGLSSAVMRAVAERLPAYMVPSAVVVLDEIPLTPNGKVDRAALPVPVGEAGAGYRAPVSPREEVLCGLFA
EVLGLERVGVDDDFGLGGHSLATRLISRVAVLGVEAGVRALFEAPTVSRLERLLRERSALGVRVPLVARERTG
REPLSFAQQRLWFLEEELEGPAAAYNIPMALRLAGVLDVEALHQALIDVIARHESLRTLIAQDAGTAWQHILPVDDP
RTRPGLPLVDIGADALQERLDEAAGRPFDLAADLPVRATVFRITDNDHILLVVAHHVAFDAMSRVPFIRNVKRAFE
ARTNGAAPDWRPLPVQYADYAAWQRDVLGTEDESSELSAQLAYWRTQLASLPAELALPTDRARPAVASYEGGKVE
FTVPAGVYDGLVALARAEGVTVFMVQAALAAALLSRLGAGDDIPIGTPVAGRTDQATEDLIGFFVNTLVLRDVS
DPTFAELLARVRATDLDAYAHQDIPFERLVEAVNPERSLARHPLFQVMLTFDNTIDREVTEGFAGLGVEGLPLGAG
AVKFDLLFGLSEVGELRGAVEYRCDLFDHPTVAQLAERLVRVLERVASDASVRTGELPVVGEAERARVLTEWNTD
GVPVGPETFLELFEAQVAARGDAPAVVYEVEVLSYRELDARANRLAGLLVGRGAGPEHFVGVALPRGLDLIVALLA
VLKSGAAYVPLDPEYPAERLVHMTDAAPVVVVTSTDVRTLRTVPRVELDDEATRATLVAAPATGPDVKMSASHPA
YVIYTSGSTGRPKGVVISHGSLANFLAWAREDLGAERLRHVVLSTSLSFVDSVVELFAPLSCGGTVEIVRNLLALV
DRPGRWSASLRSVGPVSAFAQLLEAGLDRADVGMIALAGEALSARDVRRVRVAVLPGARVANFYGPTEATVYATAWYG
DTPMDAAAPMGRPLRNTCVYVLLDDGLRVVPVGVVGEVYAGVGLARGYLGRVGLTAERFVACPFAGGERMYRTGD
LVRWRVDGTLEFVGRADDQVKVRGFRVELGEVEGAVAAHPDVVRVAVVVVREDRPGDHRLVAYVTGVDTGGLSSAVM
RAVAERLPAYMVPSAVVVLDEIPLTPNGKVDRAALPVPVSVAGFCAPSSPREEVLCGLFAEVLGVERVGVDDGFF
DLGGDSILSIQLVARARRAGLELSVRDVFEGRTVRALAAVVRGSDAGAVGVVGGAEIVLPVGVGEVERWPVVEWLAE
RGGGSLGGVVRGNQSVVLAVPAGLVWEELRVLLGAVRDRHEAWRLRVLDGALCVDGVVDDGSGWIVRCDLSGMG
VDGQVDAVRAAAVEARAWLDPSVGRVVRVAVWLERGGDRSGVLVLVAHHLVVDGVSWRVVLDGLAEGWAQVRSQGRV
ELGVVGTSLRGWAAALAEQRRGERAGEVELWSRMVRGADVLVGSRAVDGAVDVFGGVSVSDSRASVSVSRALLTE
VPSVLGVGVQEVLLAAFLAVARWRGRGGPVVDVEGHGRNEDAVRGADLSRTVGWFTSVYPVRVPVESASWDEV
AGGPVGRVREVKETLRLSLPDQGLGYGILRYLDPEHGPALARHATPQFGFNYLGRFTTGTDGDEGMTDWPVPS
GPFAVGAGQDPELPVAHAVEFNAITLDTPEGPRLGVTWSWPTTLLPESRIELARYWDEALEGLVEHARHPEAGGL
TPSDVTLVEVNQVELDRLQAGVAGGAEIIPVSALQEGLLFHSALASGGVDVYVQGVLDLVGPVVDVRLRAAVEG
LVARHGVLRSYRQRLSGEWWAVVARQVDLPWQSIDVRDGGIDGLVEEERWRRFDMGRGPLARFVLIETHDRFRF
VITYHHVILDGWSVPVLLRELLALYGSAGSDVSVLPVSGDFLRWAAARDAAAEAGAWRRALTGLLEEPSLTVAPGV
SRDGVVPAAFHGAVDGDLDSQKIVAWARGRGVTVASVVAQAAWALVLGRMLGRDDVVFVTVSGRPAEVVGVEDMVG
FVNTIPLRARLDPAESLGGFVERLQREQTELLEHQHVRLAEVQRWAGHKELFDVGMVFDNYPVSSSESPEAEFQISR
TGGYNGTHYALNLVASMHGLELELEIGYRPDVFDAGRVREVWGLVVRVLEGGVSGGGVSVSGVDVLGVGERERLL

GWGVGGPVVPGGGVLGLFEERVADADAVAVRGAGVVWSYGELNARVNVVARWLVGGRVGAECGVGVVMGRGVD
VVVMLLAVAKAGGFYVPVDPEWVVERVGVVLADAGVGLVVVGEGLSHVVGDFPGGEVFEFSSRVVRESCLVELVAAD
GVEVRNVTGDERASRLPGHPLYVVTSGSTGRPKGVVTHASVGGYLARGRDVYAGAVGGVGFVHSSLAFLDTVT
VLFTPLVSGGCVVGLGEDESQAQGVGASFVKVTPSHLGLLGELEGVAGNGMLLVGGEALSGGALREWRERNPGVVV
VNAYGPTTELTVNCAEFLIAPGEEVPDGPVPIGRPFAGQRMFVLDAALRVVPVGVVGEVLYAGVGLARGYLGRVGLT
AERFVACPFVGPGERMYRTGDLVRWRVDGALEFVGRADDQVKVRGFRVELGEVEGAVAAHPDVVRVAVVVVREDRPG
DHRLVAVYVTAGGVGGDLRSASISGLVAERLPAYMVPSAVVVLDEIPLTPNGKVDRAALPVPEVEAGTGYRAPVSPR
EEVLCGLFAEVLGVERVGVDDDFELGGHSLLATRLISRVRVLGVEAGVRALFEAPTVSRLERLLRERSGLGVVRV
PLVARERTGREPLSFAQQRLWFLEELGPGAAYNIPMALRLAGVLDVEALHQUALIDVIARHESLRTLIAQDAGTAW
QHILPVDDPRTRPGLPLVDIGADALQERLDEAAGRPFDLAADLPVRATVFRLTNDNHILLVLHHIAGDGWSMGP
ARDLSTAYSARAAGAASAWRPLSVQYADYAAWQRDVLGTEDDESSELQAQLAYWRTQLASLPAELALPTDRAPAV
ATYRGGRIEFTIPADVHRSLADLARAEGVTVFMVVQAALAALLSRLGAGDDIPIGTPIAGRTDQATEDLIGFFVNT
LVLRTDVS GDPTFAELLARVRATDL DAYAHQDIPFERLVEAVNPERSLARHPLFQVMLAFNNAETSTPLPMAEGLA
ASRQDIEPGVAKFDLALYCNESETGDHQRISVFEYRRDLWDEDTVRLADRFLHVLAAFAAAPEQRASSVDVL
RAGERDQLLHEWNTAAALPPALLPQLFEEQVRRTPHDVALVSGNIRLTAYELDARANRLAHLLLARGAAPETFVA
VALPRTEELLVALLAVQKTGAGHLPLDPGFPAERLSYMLDDARPAVVLTTEDISARIPGGSHVVLDSQVTEGELHD
HPATSPAGRGNPAGPAYVIYTSGSTGQPKGVVPSAALVNFLADMVPRGLRGGRDRLSVTTTVGFDIAALELFVPL
LSGATVVLADGETVRDPALARQTCEDHGVTMVQATPSWWHGMADAGDSLGRGVHAVVGGEALSPGLRDALTRGARS
VTNMYGPTETTIWSTSAGQAAGDSAPPSIGTPIINTRVYVLDAAALCVVPPGVAGELYIAGDGLARGYLGRAGLTAE
RFVACPFVGPGERMYRTGDLVRWRVDGALEFVGRADDQVKVRGFRVELGEVEGAVAAHPDVVRVAVVVVREDRPGDH
RLVAVYVTVGDTGGLSSAVMRAERLPAYMVPSAVVVLDEIPLTPNGKVDRAALPVPGVEAGAGYRAPVSPREEVL
CGLFAEVLGVERVGVDDDFGLGGHSLLATRLISRVRVLGVEAGVRALFEAPTVSRLERLLRERSGLGVVRVPLVA
RERTGREPLSFAQQRLWFLEELGPGAAYNIPMALRLAGVLDVEALHQUALIDVIARHESLRTLIARDSGTARQQV
LPVGDPAARPALPVVQTDADTLVAKLNEAVGRPFDLTAEMPLRATVFRVADEDHALLLVFHHIAGDGWSTGLLARD
LSTAYAAARLEGRDPQLPPLPVQYADYAAWQRDVLGTEDDESSELQAQLAYWRTQLADLPAELALPADRVPARASY
EGGRVGFVTPAGVLRDLTRLARVEGVTVFMVVQAALAALLSRLGAGDDIPIGTPIAGRTDQATEDLIGFFVNTLVL
RTDVS GDPTFAELLARVRATDL DAYAHQDIPFERLVEAVNPERSLARHPLFQVMLAFDNTADGGPVEDFPGLSAG
LPLGAGAAKFDLLFGLSEVGELRGAVEYRCDLFDHPTAARIAERLVRVLERVAADASVRLGELPVVSDAERACVL
TEWNTAVPGVTGTLTALFEARAAARGDAPAVVYEGEELSYRELNTRANRLAHVLAEHGAGPERFVGVALPRS PDL
VVALLAVVKSGAAYVPLDPEYPADRLAYMAGDAAPVAVLTRGDVELPGSVPRIGLDDTEIRATLATAPGTNPGTPV
TEAHPAYMIYTSGSTGRPKGVVSHGAI VNRLAWMQAEYRLDATDRVLQKTPAGFDVSVWEFFWPLLEGAVLVFAR
PGGHRDAAYLAGLIERERITTAHFVPSMLRVFLEEPGAALCTGLRRVICSGEALGTDLA VDFRAKLPVPLHNLGY
TEAAVDVTHHAYEPATGTATVPIGRPIWNIRTYVLDAAALRPVPPGVPGELYLAGAGLARGYHGRPALTAERFVACP
FGVGPGERMYRTGDLVRWRVDGTLEFVGRADDQVKVRGFRVELGEVEGAVAAHPDVVRVAVVVVREDRPGDHRLVAV
TVGGVGGDLRSASISGLVAERLPAYMVPSAVVVLDEIPLTPNGKVDRAALPVVPSVAGFCAPSSPREEVLCGLFA
EVLGVERVGVDDGFFDLGGDSILSIQLVARARRAGLELSVRDVFEGRTVRALAAVVRGSDAGAVGVVGGAEIVLPG
VGEVERWPVVEWLAERGGGSLGGVVRGFNQSVVLAVPAGLVWEELRVLLGAVRDRHEAWRLRVLD SGALCVDGVVP
DDGSWIVRCDLSGMGVDGQVDAVRAAAVEARAWLDPSVGRVVRVAVWLERGGDRSGVLVLVAHHLVVDGVS WRVVLG
DLAEGWAQVRSGGRVELGVGTSLRGWAAALAEQGRGERAGEVELWSRMVRGADVLVGSRAVDGAVDVFGGVSV
DSRASVSVSRALLTEVPSVLGVGVQEVLLAAFGLAVARWRGRGGPVVVDVEGHGRNEDAVRGADLSRTVGWFTSVY
PVRVPVESASWDEV RAGGPVGRVREVKETLRLSLPDQGLGYGILRYLDPEHG PALARHATPQFGFNYLGRFTTGT
DETTTADALDRAPAWSLLARSAAGQDPELPAHAFEVNAITLDTPEGPRLGVTWSWPTTLLPESRIRELARYWDEA
LEGLVEHARHPEAGGLTPSDVGLAELSFAEIELLEDWRTQG*

SEQ ID NO: 12

Nucleotides 56044-78060 of SEQ ID NO: 1

SEQ ID NO: 13

GTCTTCACCACGGGG

SEQ ID NO: 14

GTGGTGAAGAC

SEQ ID NO: 15

VSESRCAGQGLV GALRTWARTRARETAVVLVRDTGTTDDTASVDYQQLDEWARSIAVTLRQQLAPGGRALLLLPSG
PEFTAAYLGC LYAGLAAVPAPLPGGRH FERRRVA AIAADSGAGVVLTVAGETASVHDWLTETTAPATRVVAVDDRA
ALGDPAQWDDPGVAPDDVALIQYTSGSTGNPKGVVVTANLLANARNLAEACELTAATPMGGWLP MYHDMGLLGT

TPALYLGTTCVLMSSSTAFIKRPHLWLRITIDRFGLVWSSAPDFAYDMCLKRVTDEQIAGLDLSRWRWAGNGAEPIRA
ATVRAFGERFARYGLRPEALTAGYGLAEATL FVSRSQGLHTARVATAALERHEFRLAVPGEAAREIVSCGPVGHFR
ARIVEPGGHRVLP PGQV GELVLQGA AVCAGY WQAKEETEQT FGLTLDGEDGHWLRTGD LAALHEGNLHITGRCKEA
LVIRGRNLYPQDIEHELRLQHPELESVGAFTVPAAPGTPGLMVVHEVRTVPVADDHPALVSALRGTTINREFGLDA
QGIALVSRGTVLRTTSGKVRRGAMRDLCLRGELNIVHADKGWHAIAGTAGEDIAPT DHAPHHPA*

SEQ ID NO: 16

Nucleotides 36408-38201 of SEQ ID NO: 1

SEQ ID NO: 17

MNPPEAVSTPSEVTAWITGQIAEFVNETPDRIAGDAPLTDHGLDSVSGVALCAQVEDRYGIEVDPELLWSVPTLNE
FVQALMPQLADRT*

SEQ ID NO: 18

Nucleotides 38270-38539 of SEQ ID NO: 1

SEQ ID NO: 19

MIGVAPPAYDPAAPESATTLPGVTPTTVRSYVRSLLRRHRAFTVLI AVNAVAVVASITGPYLLGGGLVEDLSAGVT
DLHLERTAAIFAVALVVQVLFTRSMRLRGAMLGEEMLADLREDFLVRSVGLPPGVLERAGTGDLLSRITTDIDRLA
NAMREAVPQLAIGVVWAGLLLGALT V TAPPLALAVLIALPVLIVGCRWYFRRAPSAYRSEAAGYAAVAAMLAE TVD
AGRTVEAHLRGRRVALSDRRISQWTAWERYTLFLRSVLF PVINATYVTILGAVLLLGGWFVLEGWLTVGQLTTGA
LLAQMMVDP IGLILRWYDELQVAQVSLARLVGVRDIEPDAGDAEVGPEGRDVRADDEVRFYREGVDVLHKVSLDVA
PGTRLALVGP SGAGKSTLGRLLAGIYAPRTGEVTLGGAELSRMTAERVREHVALVNQEHV FVGS LRDNLRLAREG
AKDAELWASLA AVDADGWAKALEKGLDTEVSGSGFTLTPAQAAQ QIALARLV LADPHTLV LDEATSLLDPRAARHLE
RSLARVLEGRTVV*

SEQ ID NO: 20

Nucleotides 1637-1 of SEQ ID NO: 1

SEQ ID NO: 21

MSPPAPPEALQRPAPTAQEPVRTGSRTGLVAICVSLFAALVVSVVVAIGLGPVAVPPAETARFLWAALSGGPISAD
EVTTYQIIWQIRTPRVLLAALVGAGLSAVGVAIQALVRNALADPFVLGVSSGASVGAVGVTVMGGLAVFGIYAVSV
GAFGLVASVLVYGASSTKGALSPLRLVLTGVAMSLGFGQAVMSVIIYFAPSSEATSMVLYWTMG SFGAASWGSLP
VVTAAVLLGVLVLRHGRPLDVLALGDETAASLGISPDRHRKSLLVLSLVTGVMVAVSGSIAFVGLVMPHLVRMV
VGATHARVLAVAPLAGAVFMVWVDLVSRTL VAPRELPLGVITALVGVPVFITLMRRKSYMFGGR*

SEQ ID NO: 22

Nucleotides 3502-1634 of SEQ ID NO: 1

SEQ ID NO: 23

MNDDARPAPEPQDIPPHSGAADEVNRQDPSRRSVLWTTAGVAGAGLGLGALGAGTASAAGRSAPDAVAAAEAVAAA
PPRQGRTMAGVPFERRSTVRVGIIGLGNRGDSMIDLFLALPGVQVKAVCDTVRDKAEKAAKKVTAAGQPAPAIYAK
DEHDYENLCKRGDIDFVYVTPWELHFPMAKTAMLNKGHVGECP IAMRLEELWQLVDLSERTRRHCMQLENC CYG
KNEMRVLMAHAGLFGELQHGAGAYNHDLRELMFDPDYEGPWRRLWHTRLRGDLYPNHGFGFPVANYMDVNRGDRV
VSISVGTTPGLAAYREEHMPAGDPSWKESYIGADRTISLVQTAKGRVIRLEHDVSSPHPYSRINSLGGTKGVFE
DYPERIYLEPTNTNHQWDDFKKYAEWDHVLWKEHANPPGGHGM DYIMVFRLMQCMLRGLVPDFDVYDAAVWTAPV
PLSHLSIKAKGVPLPIPDFTRGEWKKTRSGMDSEKPAE*

SEQ ID NO: 24

Nucleotides 5144-3659 of SEQ ID NO: 1

SEQ ID NO: 25

MPLLEPDPEALRPGTAREPAPDRVTDGSAGGTPEPLRSELTALLGADKVLWKISDLVRYASDASPYRFLPRVVLVP
EDLDDVSAILSYAHGKGRSVVFRAAGTSLNGQAQGEDILVDVRRHWTGVEVLDDGARARILPGTTVMRANAALARY
GRLLGPD PASIACTLGGV VANNASGMTAGTTRNSYRTLASLT FVLPSGTVVDTAHPAADEELAHAEPELCAGLLE
LKAIEADAELTARIRAKYTIKNTNGYRLDAFLDGATPVQILRGLMVGSEGTFGFI SEVVFDTLPLDRRVSSGLLF
FPSLTAAAAVPRFNEAGAI AVELMDGNTLRASVSVPGVPADWAALPRETTALLVEFRAADEAGRAAFERAADAVV
AGLDLVRPAASVTNAFTRDAGTIAGYWKARKAFVTA VGGSRPSGTTLITEDFAVPPARLADACAALLELQSRHGF

AAVAGHAAHGNLHFLAFDAAKPADVARYDAFMQEFICALVVDVDFDGSLSKAEHATGRNIAPFLEREWGPRATELMWR
TKQVIDPAGVLAPRIVLDRDPRAHLRGLKTI PKVEAVADPCIECGFCEPTCPSEDLTTPRQRIVLREMMRQTDG
SPVESGLLDAYGYDAVDTACAGDSTCKLACPVGIDTGAMMKGFRHRRHTPREERIAALTAKNFRAVEASARLAVAAA
DTVGNRVGDAPLQAVTRLARKAVRPDLVPEWLPQIPGAAARRLPDARVVGASAVVYPACVNRIFAGPDDGDAGPAL
SLAEAVVAVSGRAGKPVWIPEDVTGTCCATIWSKGYDAGNRIMANRIVEAAWGWTAGGTLPLVVDASSCTLGIAE
EVVPYLTEDNRALHRELTVDLSLVWAAEELLPHLTVFRTAGSAVLHPTCSMEHLGDVQQLRALAEACAQEVVVPDD
AGCCAFAGDRGMLHKELTDSATAKEAAEVDRRPYDAYLSANRMCEIGMERATGHPYRSALIELEHATRPTLP*

SEQ ID NO: 26

Nucleotides 8364-5410 of SEQ ID NO: 1

SEQ ID NO: 27

MDAPDSPDSPSPESRDSRDSRDLGLAEQLLRLTRRLHRIQRRQLEPIDITPAQFRLRLTVASYDAAPRMADL
ARRLDVVPRAVTTLVDALEASGRVRRAPDPDSRRVVRIEITDEGRATLRLSLRSARRAAAEILAPLTADQREVFG
LLSALVDGMPERHC*

SEQ ID NO: 28

Nucleotides 8916-8416 of SEQ ID NO: 1

SEQ ID NO: 29

MKPDEPTWTPPPDARPAADRRPAEVRRIILRLFRPYRGR LAVVGLLVGASSLVGVASPFLLREILDTAIPQGR TGLL
TLLALGMILTAVMTSVFVGLQTLISTTVGQVRMHDLRTAVYTQLQRMPLAFFTRTRTGEVQSRIANDIGMQATVT
STATSLVSNLTAVIATVAMALDWRLTVVSLLLLPVFVAISRRVGRERKKITTQRQKQMAAMAATVTESLSVSGI
LLGRQTMGRSDSLTQGF AEESERLVDLEVRSNMAGRWRMSVIGIVMAAMPAVIYWAAGLTFASGAAAVSIGTLVAFV
TLQQGLFRPAVSLSTGVQMOTSLALFQRIFEYLDLTVDITEPEHPVRLERIRGEIAFEDVDFS YDEKNGPTLTGI
DVTVPAGDSLAVVGSTGSGKSTLSYLVPRLYDVTGGRVTLGDIDVRDLDFDTLARAVGVVSQETYL FHASVADNLR
FAKPEATDEEIEAAARAAQIHDHIASLPDGYDTMVGERGYRFSGGEKQRLAIARTILRDPPVLILDEATSALDTRT
EQAVQE AIDALSAGRTTTLTIAHRLSTVRDADQIVVLEDGRVAERGTHEELLDRDGRYAALIRRD SHPVPVPVPAP*
*

SEQ ID NO: 30

Nucleotides 9030-10853 of SEQ ID NO: 1

SEQ ID NO: 31

HRHLAERPRRC AVLALLRPAAGPAGRAGRRPGPAARS DPLHRQGGRRPHRDIGEAAGRAARPAADTQTAAAEPAQR
PGVHRQLHRAARRMQHRGEDPGGGARHDGHAGSRGDGQARPRPVLPAAPLRPRGPGRAALSHGGSRPVGRGVPGPS
AHPGPPDARHPRGGGDTVRRAAALHRTGSGERLSRPAAYTQHTAHRAHGAHSTHGGAAPVGRGATAPGGAMV
RRANPRSGRRRQAGWSGSSGLSPCTWCICGTAQ*

SEQ ID NO: 32

Nucleotides 10933-11544 of SEQ ID NO: 1

SEQ ID NO: 33

MVNESPDARPRRRLRPTRRGKIVLVGALLVVTAAVLIPLSLTGSDEPPKKQETPQSTLMIPEGRRVSQVYEAVDK
ALDLKPGSTLKAAS TVDLKLPQAEGNPEGYLFPATYPIDDTTEPAGLLRYMADTARKHFAADHVTAGAQ RNNVSV
YDTVTIASIVQAEADTPADMKGVARVVYNRL LKDMPLQMDSTINYALKRSTLDTSTADTQLDSPYNSYRIKGLPPT
PIGNPGEDALRAAVRPTPGPWLYFVTVGPGDTRFTDSYDEQQKNVEEFNRGRGSATTG*

SEQ ID NO: 34

Nucleotides 11990-12850 of SEQ ID NO: 1

SEQ ID NO: 35

MIPGARRVSRSVNISGVRELDVVVIGAGQAGLSAAAYHLRRVGLPEPDNDFVVDHAPRPGGAWQFRWPSLTYGKVHG
MHALPGMELTGADPDRPSSEVIGAYFAAYEDRFLRVHRPVEVS AVREGSGGRLLVETSEGTYAARALINATGTWD
RPFWRPYPGQETFRGRQLHTANYPGPEEFAGQRVLVVGGGASGTQHLMEIAEHAADTFWVTRSEPVFREGPFTEEW
GRAAVAMVEERVNRGLPPKSVSVTGLPLNDAVRRARERGVLDRLPMFDRITPTGVAWDDGRTVETDVILWATGFR
PAVDHLAPLKLREPGGGIRAEDTRAVRDGRVHLVGYGPSASTIGANRAGRAAVRSVMRL LKETGADGGASAVVSV
APVPGV*

SEQ ID NO: 36

Nucleotides 14038-12878 of SEQ ID NO: 1

SEQ ID NO: 37

VPGLARPTRSTPPRQLRRGHPPSLSRPPTTEPLTTTPPPPEPTQRHTSLCNTDSLAVAMSERPRHRPQKRSIACGAC
RAGSSPLAHTGVGLVRGGACTALVGSHAEVADRIEYHALGVEHFVLSGYPHLEEAYWFGEGVTPELSRRLSTV
PASPLLGVSGAESRTATAPGGAPLLLAGGR*

SEQ ID NO: 38

Nucleotides 14348-14070 of SEQ ID NO: 1

SEQ ID NO: 39

VAVVAEDLRRRFAATKVTFILVDLTGRALARLSTTTAAGSENERIPLFGGSVYEQVIRTQRPHEPAGQEQRVI
VPVTNRGDAIGLLELLLPAGRSDEEEVVLAVGEAAHALAYVVIANGRFTDFYTWGKRSPPTLAAEIQYQLLPQAL
SCEAAQFTLSGSLEPSEDLSGDTFDYALDRDTLHLSVTDPMGHDLGAALAAATVLVGALRRARRAGAPLAEQARQGD
QALTSHGQGHATGQLLRINLHTGKAELVNAGHPWPMRMAGMVETIPCQVDQPFGLAVVSPRPYRVQTLDLHPGDR
LLMLTDGMLERHGEKIDVAALLRQTRSLHPRETTMLTSAVRDAAGGRLEDDATVVCLDWHGPPQEVHRHVSSGADT
HQASAARPPNR*

SEQ ID NO: 40

Nucleotides 15697-14522 of SEQ ID NO: 1

SEQ ID NO: 41

MRVRLQVGVALCGLGVLVTQERERRRCGARSAGMVPDPLLLAVAFEAGAFQFQASRSRVRAEHGQGGALRQTARK
FANSGPATGRAVGQDDPMSQDLVTFILHARLDEEANLAGRCDGDCGEWAPHGHTVDFCQGELSGFHSITIALHVALH
DPAVLREAERARVRLARHGLSPATGDPPELPWDNRDDCRYDGATWPCDDLGLASPYADHPDYPQRP*

SEQ ID NO: 42

Nucleotides 17597-16938 of SEQ ID NO: 1

SEQ ID NO: 43

MSVRDLVGMPCHPCEPPRRAEGRRRGVGRMRWWKGVLMTVRHQGVRRWFALLALVGCVCVLCVVALSGAGHYFGL
SLWAGIALVVVGALFPLGGLGFLYWDDGRSEDSFLVKFLCFVAHSAVLGLAAVSCTGAEEAWAFEQGRWTEATVV
GYSPRVVPGDPPTKVRASCALETAEGERVPRPLPEGRGCRDGVRRHGSRLDVLYDPRGLLAPRATEPMDHGVTPV
LGGVATLSGFLGCVALAWRWETLRVRSARRTAARRGRESAAG*

SEQ ID NO: 44

Nucleotides 17870-18682 of SEQ ID NO: 1

SEQ ID NO: 45

MKFTKLAIPVAASALLLTGCGAEVESQKGSGKSTVKRCGESVEYTVPKRAVAYEGGSADKLFSLGLADHVHGYVM
PPANPPVSESPWAKDYAKVKMLSDLLNKEIVDAKSDFFVAGWNSGFSQDQGITPEILDKLGVQSFMHSESCYNY
PGHPEKLTFFKGLYTDLERLGRIFQVEEEAEKVAGLKKREAAVAEQAPKGPVPVFLYDSGTDQPFAGNQVPPN
DIIKTAGGKNIFDGLERWTQVNWEAVTQAEPEVIMIFDYGDQPAEKKIEFLKKSPTKELPAVKKNFFVLDYNE
GISSPRNIDGLEKFGKYMRAFKK*

SEQ ID NO: 46

Nucleotides 19898-18915 of SEQ ID NO: 1

SEQ ID NO: 47

MDLELDGLSVVTDGKSLVRDLSLDVGSQVVGVLVGPNGSGKSTALRCVYRALKPSSGTVKVDGQELSSLTMRSAQ
LIAAMTQDGAVDLDFTVEEVIALGRTPHQGSTPLNGHERDLCEHAMRRDLILHLARRGILTLSGGERQRVLLARA
LVQEPKILVLDEPTNHLDRHQVRLLSLLRGAGLTVLVVLHDLNLAAAACDRIGVLSEGRILITSGTPKDVLTPELV
DEVFGVRASVVPPLTGDQPQLLYSLDS*

SEQ ID NO: 48

Nucleotides 20674-19907 of SEQ ID NO: 1

SEQ ID NO: 49

MSPPAPPEALQRPAPTAQEPVRTGSRTGLVAICVSLFAALVVSVVVAIGLGPVAVPPAETARFLWAALSGGPISAD
EVTTYQIIWQIRTPRVLLAALVGAGLSAVGVAIQALVRNALADPFVLGVSSGASVGAVGVTVMGGLAVFGIYAVSV
GAFLGALVASVLVYGASSTKGALSPLRLVLTGVAMSLGFQAVMSVVIYFAPSSEATSMVLYWTMGSGAASWGSPLP

VVTAAVLLGVLVLRHGRPLDVLALGDETAASLGISPDRHRKSLVLVSLVTGVMVAVSGSIAFVGLVMPHLVRMV
VGATHARVLAVAPLAGAVFMVWVDLVSRTLVAPRELPLGVITALVGVPVFITLMRRKSYMFGGR*

SEQ ID NO: 50

Nucleotides 21782-20676 of SEQ ID NO: 1

SEQ ID NO: 51

VSAGTSRSAVAPEKSPMPGDLKMARALWPVLVASAVGLLPFTVFSTYLVPIAEETGSGVAAVGGLRGLGGLAALA
VGTALAPLIDRVPKSKAVAVGLVVLAVSSALGASGDFLLTAVFCLLVGAGTAVINPALTAAADRFGDGKSAARAA
TLVTSTTSMTAMLAAPLIALPALLWGWEGDLLAVTVVSLLLAAVFLVRGRKGEDPVVEGGPRTGYFASFALAQVR
GSPVLLAISFLRTAVFMGYLAYLAVYYDDRFHLDPALFSLVWTLSGASFFVSNLLTGRITNAEKSTVGTEQLLLVG
LLAALVTATGFWFTTWLPLALFTSLHAASHAAVAACAVSLLVRRCGSMRGSALSLNAAGQSLGVFAGAALGGAGL
GLAGYPGIAAAFGLLVAVAVVAGLLVLRSEDEIPGSA*

SEQ ID NO: 52

Nucleotides 23130-21877 of SEQ ID NO: 1

SEQ ID NO: 53

MTPPPTRRKPSDMPFPTPQSVAELTDAVLGADYGPDPKDMTIVTSAFWLYHTTRLAGGPVITYHNHYLVLRVGRSFGG
CSFEAGELTPDFCENASGHPLEKLLRHESAPVRIAALDAYLAQIQPHREAPEQEAIVPLPVGTPEVRAKARDASIAG
LLDIEEGAKVALIGVVNPLVAAIRERGGVCLPCDLNLRRTTQWGEVADDMTFVLAEAHAVVATGMTLSNGTFDLIL
EHCREQKVPLVVYAQTGSAVARAFLGSGVTALNAEPFPFSQFSADETTMYRYRAGGDL*

SEQ ID NO: 54

Nucleotides 23987-23127 of SEQ ID NO: 1

SEQ ID NO: 55

MYEHIAEAIKKPDLIALRPDLVCLRFETMKIYSALGAVRHLLSESGTVKPGDTLVDSSSGIYAQALALACHRYGMKC
HIVGSTTVDRTLKAQLEILGATLEQVRPSRNLRLDQELRVRRIAEILEENPSYHWMRQYHDSIHYYGYREVAETIA
DEVPAGPLTLVGGVGGSGASTGAIASYLREAGRDVSLVGVQPFSGVTFGSEHVSDPDMIAGIGSAIPFENVRHDL
DRIHWVSFDSALAGAVHLLRSSGIFAGLSAGAAYLTTRWERSKDDSRITYVFIADTGHRVDSAYAKHTEAPDIED
LEPREITSLDELSPWSAMTWTDDSTSDQKKAL*

SEQ ID NO: 56

Nucleotides 24966-23953 of SEQ ID NO: 1

SEQ ID NO: 57

MDTGVTAYGTFTGELLQGELPEEAGDFLVTLPVARWARASFRCDPAMGDVIVRPSHKEKARRLACLILEEAPGMTG
GVLTVNSVIPEGKGLASSADLVATARAVGRALRLDMPPSRIEGLRLIEPTDGVLYPGIVAFHHRVRLRAMLG
LPAMSVGVDEGGAVDTVDFNRIKPFPTADRREYADLLNRLSGAVRSRDLAEVGRVATRSALMNQPLRYKRLLEP
MREICRDAGGLGVAVGHSGTALGVLLDAADPAYPHRATAVARACGDLGAVAVYRTLSFPNAVSHGGRTVG*

SEQ ID NO: 58

Nucleotides 25228-26127 of SEQ ID NO: 1

SEQ ID NO: 59

MLTAQQPAPGVVPAIRIHVTDRLEAAHPLAADGAVVLTGVEPSGDGLVLAAGVLRERLQQVFPHRLRASDGSNFBH
LHADSFDVFNVGGEVHRRRDPDEDYVLIQCVRQSDSGGDSFVADAYRFVDHCATADPELWDFLTRGDVDLYGAW
GLRGPATPFVGRHVVEYTRAGRRIVRRGDGVTPLHRDPGADHTRMLARLEEAVHALEETLPRFRLDKGEILVLDN
YRCWHGREAHTGDRAVRILTVRSSDAR*

SEQ ID NO: 60

Nucleotides 26445-27212 of SEQ ID NO: 1

SEQ ID NO: 61

MTTMFNNNPPFPFATELRNERVRFQRLSAGYPGRPVHLQLSAAIPPLAMTALVGPNNGSGKSTLLGVLAGVITATSG
QLRYAEGSPPAFVPQRGAVGDTLPLTARQTVEMGRWGQRLWRRRLTRTDRTAVDSAMERLGVADLGRQLGELSGG
QRQRLIAQGLAQQSDLLLLDEPTTGLDPEARERITALTDLVADGTTVVQATHDLDAARSADACLLADGRIGQ
GSPPEVLTPEALARIWQPA*

SEQ ID NO: 62

Nucleotides 28124-27381 of SEQ ID NO: 1

SEQ ID NO: 63

MEWLTAPFEVAFVQRALWAGILVSAICALAGTWVVLRGMAFLGDAMSHGLLPGVAVASLLGGNLLVGAVVSAAVMA
AGVTALGRTPRLSQDTGIGLLFVGMLSLGVIIIVSRSSQSFVAVDLTGFLFGDVLAVRGSDLLLLGVALLLALAVSVLG
YRAFLALAFDERKARTLGLRPRLAHAVLLGLLALAIVASFHIVGTLLVLGLLIAPPAAAMPWARSVQAVMVLALL
GAAATFGGLLLSWHLRTAAGATVSALAVALFFLSHLASGLRHRRRRARRGGLAEPVAVAPGRDLLHVLTERNLRRSPC
SSEKTSRHLRLRLRP*

SEQ ID NO: 64

Nucleotides 28139-29098 of SEQ ID NO: 1

SEQ ID NO: 65

VILLTAGCGGGDEAKSGSGPASSPTPHGYVEGATEAAEQQSRLLLGDPGSGETRVLDLITGKVYDIARSPGATAL
TTDGRFGYFHGPDGIRVLDSGAWMVDHGDHVVHYRAKIKEVGELPGGTGTSIRGDAGVTVASSADGKASVYRRADL
EKGALGTPSPLPGTFAGAVVPYAEHLVTLTAESGAPAKVAVLDRSGKRVAAPAECEEPQGDVTRRGVVLGCADG
ALLVHEDDGAFTAEEKIPYGEDVPKTERAVEFRHRPGSSTLTAPAGKDAVWVLDAGEGAWTRVKTGPVVAANTAGEG
SPLVVLETDGALHGDIPTGKETGVTDPLLKELPGTGAGGGAAPVIEVDRSRAYLNDPEGKRVYEIDYNDLRLVAR
TFDVDVVRPSLMVETGR*

SEQ ID NO: 66

Nucleotides 29095-30285 of SEQ ID NO: 1

SEQ ID NO: 67

MSARVGAPMRALLVSLAGFFVAGAATGCAGGGDERPRVVVTNILGDITREIVGDEAGVSVLMKPNADPHSFGL
SAVQAAELENADLVVYNGLGLEENVLRHVEAARESGLVAAFAAGEAADPLTFHAGQDGGPEEDAGKPDPHFWTDPDR
VREAAGLIADQVAEHVEGVDEKKVRENAERYDGLADLTGWMEKSFAAIPEDRRALVTNHHVFGYLADRFGLRVIG
AVIPSGTTLASPSDDLRLSLTQAMEKAKVRTVFADSSQPTRLAEVLRQEMGGDQDVVSLYSESLTEKGKGAGTYLE
MMRANTSAMAEGLTGD*

SEQ ID NO: 68

Nucleotides 30282-31244 of SEQ ID NO: 1

SEQ ID NO: 69

MNKPTRARVFTGTALVVAASMALTACGGNGNDAPSGKEPKEQKSSEAAVGNPIVASYDGGLYVLDGETLKLAKT
IALPGFNVRNPGADNEHVVVSTDSGFRVFDATRQEFTDAEFKGSKPGHVVRHGGKTVLFTDGTGEVNVFDPADLSD
GKKPDGRITYTSKPHHGVAIELAGGELVTTLTGTEEKRTGALVLDKDNKEIARAENCPGVHGEAAAQGEVAGFGCED
GVLLYKDGKFTKVDAPGDYARTGNQAGSDASPIILLGDKYKTDPAELERPTRISLIDTRTAKMKLVLDGTSYSFRSL
ARGPHGEALVLGTNGTLHVIDPETGKVEKKIDAVGDWTEPLDWQQPRPTLFRVDHTAYVSEPGKRLHSIDLESGK
KLASVTLPKGTNELSGTVAGH*

SEQ ID NO: 70

Nucleotides 31332-32537 of SEQ ID NO: 1

SEQ ID NO: 71

VSWMNDVLTAVSDMNPVTRFALASVFAPAESGLGAGMAVPGEVAVLALSAGTEGTRPLLALFLVVTLSSSAGDHIG
YFLGIRYGQRMRETRLVRRIGQHWDRAQELCHRYGARAVFLTRLLPVVRTLTPTAGVGSVRYLRFLPASLAGAA
MWSALYVSAGTLVSTSLREAESVLSTILWALLGVAAAFTLAIVWVRRRHRRRSS*

SEQ ID NO: 72

Nucleotides 32816-33427 of SEQ ID NO: 1

SEQ ID NO: 73

MELCALHSRDRDATVKTCAAGRPKRKPSYGFGLGRPTAAEELAAVTSCGGGACAATTRSRA*

SEQ ID NO: 74

Nucleotides 32686-32868 of SEQ ID NO: 1

SEQ ID NO: 75

MGGSAIRTRQLTKHFGAVQALVGVDLEVPAGSVLGLLGHNGAGKTTLIQILSTVLPPSGGSAEVAGFDIVRDARRV
RACIGVTGQFAALDEHLSGLANLVLLISRLLGARPREARRRAELVEQFGLTEAADRPMRTYSGGMRRRIDLAASLV

ARPSVLFLDEPTTGLDPVSR TALWETVEGLVAEGTTVLLTTQYLDEADRLADRIAVLSSGHVVTVGTAAELKAAGT
RSVRLTFGSAADLESAEGALRLEGLGLTTDPVSRTVSLPLAATAELAGIFRILGAAGVELAELALKEPTLDDVYLS
LAESWETTSGGTVRC*

SEQ ID NO: 76

Nucleotides 34195-35154 of SEQ ID NO: 1

SEQ ID NO: 77

LTTTRRTGPGTSPVADGPGWRGGGAGIGTQFRVLTGRQFRIIYGDRRIALFSL LQPIIMLM LFSQVLGRMANPEIFP
PGVRYLDYLV PALLLT TGIGSAQGGGLGLVRDMESGMMVRLRVMPVRLPLVLVARSLADLARVALQLVALLACAMG
PLGYRPAGGVSGIVGATLLALLVAWSLIWVFLALAAWLRSIEVLSSIGFLVTFPLMFASSAFVPLDILPGWLRVIA
TVNPLTYAVEASRD LALDHSALGAALAAVGTSLALLAVTGLLAVRGLRRPPGAGGPHRTP*

SEQ ID NO: 78

Nucleotides 35148-36017 of SEQ ID NO: 1

SEQ ID NO: 79

MANPFENNDGSYLVLVNDEGQYSLWPAFADVPAGWTVTFGESSRQECLDHINENWTDMPKSLIRQMENDRTTAA*

SEQ ID NO: 80

Nucleotides 85270-85497 of SEQ ID NO: 1

SEQ ID NO: 81

MTVHDYHVTVKEQHPALFELLDPARLVAVTDEPWTEGNEFDDDHAGRGVS YRCAQQHGEARRTGIETILGMFAGP
GGLRDMGRVLDVLGGEGLLSRVWRQLAGAGDGSVPLVTGDL SGHMVAAALRSGLP AVRQPADRMLQRDHCLDGVL
FAYGTHHVD RSVRPRMLTEASRV LAPGGRVVLHDFAEGSPEERWFREV VHP RSLAGHAYDHFTAHEMTGYLADAGF
TDITVGPVYDPMTLTGETDESALARLVSYMTSMFGILPDGDRSNERTEAALRDI FRFSAGDLPEDVPRDEAVLELT
VRPHGNAFRAELPRIALVAHGRKP*

SEQ ID NO: 82

Nucleotides 86434-87420 of SEQ ID NO: 1

SEQ ID NO: 83

MTAQDTRTTGSDGGGRGATYHESPTYGELLRLLEDLLNVAHLRDAAAPVLFLATHQSAE IWF GIVLRHLEEIRAALT
DDDPDTALHLLPRLPEIFELLVRHFDMLATLSTEEFGKIRAGLGTASGFQSAQYREIEFLCGLRXHRHISTPGFTE
TERRDCGNGPASPPWRRRLRRLPDPMRQREGRRERIGEALLRFDERVTVWRARHAALAE RFLGPLEGTAGTAGADYLW
RVTRHRLFPPEAWGAG*

SEQ ID NO: 84

Nucleotides 87417-88154 of SEQ ID NO: 1

SEQ ID NO: 85

MDREAEAPLRAAPHATPAERAALGKAARREAPRSGHAEFSPSPRRPDPLTVLEAQ SADRVP ELVPIRYARMTESPF
RFYRGAAALMAADLAGTPVSGIRAQLCGDAHLLNFRLLASPERNLLFDINDFDETLPGPWEWDVKRLAASLVIAGR
ANSFTLRERAGVVRATVRSYREAMARFAGMRNLDVWYARTDAERLRTVATEQLGGRGRNRNVD RALGKARSRDSLQA
FGKLAEVVDGRLRIAADPPMVVPLTDLTPGVDRDAVFRQFGSMLAGYARSLPSDRRSLLDFALVDVARKVVGVS
VGTRCWIVLLLGRDGGD

SEQ ID NO: 86

Nucleotides 965-1 of SEQ ID NO: 103

SEQ ID NO: 87

MIHIRAVSPDLDTEVVGLLSADPCVLNLIVQRDAARRPDGDAIACDVL TGAANDVLHRLRAAHLDRGSLVIEPV
DMAFSGAATEGGQRELGPLSRAPVWEQVEARIRSGGRYPPSFYLYLV IAGLIGSVGIVTNSQILIVGAMVVGPEYG
AIVSVALGIDRRHRSMVRSGLAALGVGLLLTIVVTFLFALLIRGFGLESEAFDRGLRPVSHLINTPNFFSVAVATL
AGIVGIVSLTEARTSALLGVFISVTTIPAAADIAVSTAYTSWSDVRGSAIQLVVN ILVLIVVGAFALKAQRAIWQR
VRLRRDRERRIAEQA*

SEQ ID NO: 88

Nucleotides 989-1948 of SEQ ID NO: 103

SEQ ID NO: 89

VTRPGWDHEGVDTPTDTPDAFPEPLPGADEAVREERATDDGTPEGRRLLVRCRLCGRPLTGADSRAGLGPSCDAKLH
PAPPDIRRRHEVDQDPLPGT*

SEQ ID NO: 90

Nucleotides 2099-2392 of SEQ ID NO: 103

SEQ ID NO: 91

MTNPAERLVDLLDLERIEVNIFRGRSPEESLQRVFGGQVAGQALVAAGRTTDGERPVHSLHAYFLRPGRPGVPIVY
QVERVRDGRSFTTTRVTAVQEGRTIFNLTA SFHRPEEAGFEHQ LPPARIVPDPEELPTVAEEVREHLGALPEALER
MARRQPFDIRYVDRLRWTKDEIQDADPRSAVWMRAVGPLGDDPLVHTCALTYASDMTLLDAVRIPVEPLWGPRGYD
LASLDHAMWFHRPFRADEWFLYDQESPIATGGRGLARGRIYDRSGQLLVSVVQEGLFRRLEQ*

SEQ ID NO: 92

Nucleotides 3277-2405 of SEQ ID NO: 103

SEQ ID NO: 93

VIFVPSAGSLIRAEDRQDGGVTLIDQLPQTADPDALFEAFSSWTESQGITMYPAQEEALIEVSGANVILSTPTGS
GKSLVAAGAHFTALA QDKVTFYTAPIKALVSEKFFDLCKLFGTENVGMLTGDA SVNADAPVICCTAEVLASIALRD
GKYADIGQVVMDEFHFYAEPDRGWAWQIPLLELPQAQFVLMSATLGDVSMFEKDLTRRTGRPTSVVRSATRPVPLS
YEYRFTPITETLTLELDDTRQSPVYIVHFTQAAVERAQSLSINMCTKEEKEKIADLIGSFRFTTKFGQNL SRYVR
HGIGVHHAGMLPKYRRLVEKLAQAGLLKVICGTDTLG VGVNVP IRTVLFTALT KYDGNRVRTL RAREFHQIAGRAG
RAGFDTAGFVVAQAP EHV IENEKALKKAGDDPKKKRKVVRKKAPEGFVAWSESTFDKLIQSEPEPLTSRFRVTHM
LLAVIARPGNAFEAMRHLLDNHEPRRAQLRHIRRAIAIYRSLLDGGVVEQLDTPDAEGRIVRLTVDLQQDFALNQ
PLSTFALAAFDLLDAESPSYALDMVSVVESTLDDPRQILAAQONKARGEAVGQMKADGVEYEERMERLQEVTPKP
LSELLWHAYDVYRTSHPWVNDHPVSPKSVIRDMYERAMTFTFTSHYELARTEGIVLRYLASAYKALEHTIPDDVK
SEDLQDLISWL GEMVRQVDSSLLDEWEQLANPEVETAEQAQEKADDEVKPV TANARAFRVLVRNAMFRRVELAALDR
AGALGELDGESGWDED AWGEALDAYWDAHEEIGTGPDARGPKLLKIEEDPAHGLWRVWQAFADPAGDHDWG IKA EV
DLAASDEEGRAVVRVTEVGQL*

SEQ ID NO: 94

Nucleotides 5885-3312 of SEQ ID NO: 103

SEQ ID NO: 95

MMGPAHSLSGAAAWLGVGAAAAAAGHTMPWPVLVVGALICAGAALAPDLDHKSATISRAFGPVSKALCEIVDKLSY
AVYKATKSAGDPRRTGGHRTLTHTWLWAVLIGGGCSVAAITGGRWAVLVILFVHLVLAVEGLLWRAARVSSDVLVW
LLGATS AWILAGVLDKPGYGADWLF DAPGQ EYMWLGLPIVLGALVHDIGDALTVSGCPILWPIPIGRKRWYPIGPP
KAMRFRAGSWVEMKVLMPAFMVLGGVGGAAALNYI*

SEQ ID NO: 96

Nucleotides 5963-6754 of SEQ ID NO: 103

SEQ ID NO: 97

MLLAELAQVSLEVAATSARSKKVALLAGLFRDAGPEDVPVVI PYLAGRLPQGRIGVGWRS LGDPVEPAAEPTLTVT
GVDARLTALA AVSGPGSQARRKEHLRALFAAATEDEQRFLRALLTG EVRQ GALDALAADALARAADAPPADVRRV
MLAGSLQEVAGVLLADGPEALAAFR LTVGRPVQ PMLAHTAASVGEALDKLGACAVEEKL DGIRVQVHRDGDRI RAY
TRTLDDITDRLPELTAXVAALPAGRFI

SEQ ID NO: 98

Nucleotides 6850-8403 of SEQ ID NO: 103

SEQ ID NO: 99

VNHPVNGAGERTTQAREGTQTVAPPRILVVGAGFAGVECVRRRLERRLAPGEAQITLVTPFSYQLYLPLLPQVASG
VLTPQSVAVSLRRSRHRTRIVPGGAIGVDTQAKVCVIRKITDEIVNEPYDYLVL AAGSVTRTFDIPGLLDNARGM
KTLAEAAVVRDHVIAQLDLADASHDEAERASRLQFVVVGGGYAGTETAACLQRLTTNAV KHYPRLDPRLIKWHLID
IAPKLMPELGDKLGQAAL EVL RKRNI E VSLGVSIAEAGPEEVTFTDGRVLP CRTLIW TAGVAASPLVATLGAETVR
GRLAVTPQMRLPGADGVFSLGDAAAVPDLAKGDGAVCPPTAQHAMRQGRVLADNLIASLRHEPLKDYVHKDLGLVV
DLGGTDAVSKPLGIELRGLPAQAVARGYHWSALRTNVAKTRVMTNWL LNAVAGDDFVRTGFGSRK PATLRDFEYTD
VYLTP EQIKEHTAATVIKH*

SEQ ID NO: 100

Nucleotides 9860-8433 of SEQ ID NO: 103

SEQ ID NO: 101

VTGRDLTWDTTSTVDRGRFPDAVTPWEDPAWRAEALAWTEGLAAHGLTETGPRAVRLRPWSVLVRLAVAGPAPV
WFKAVPPAAAFEAGLTEALARWVPARVLAPLAVEAERGWILVPDGGPVLSEVLDGRPGAPDPGYWEEPLRQYAAMQ
RELTPYAEAEALGVPAARPRDLPALFDRLVAGNAALPREDRVALEVLPRVADWCEELASSGVADSLDHADLHEK
QLFAPVSGRYAFFDWGDALVGHFPFCSLLVPARAARERCGPEVLPRLRLDAYLEPWTGGGVTAAGLRRRAVSLAWRLAA
LGRAASWGRMFPVPPGGPGVAGDAEGAHWLRELAAAPPL*

SEQ ID NO: 102

Nucleotides 10784-9921 of SEQ ID NO: 103

SEQ ID NO: 103

1	GGATCCCCGC	CGTCCCGGCC	GAGCAGCAGG	ACGATCCAGC	ACCGGGTGCC	GAACTGCCCC
61	ACACCGACGA	CCTTCCGGGC	CACGTCCACC	AGCGCGAAGT	CCTCCAGCAG	ACTGCGCCGA
121	TCCGATGGCA	GGCTGCGTGC	GTACCCCGCC	AGCATGGAGC	CGAACTGCCG	GAACACCGCG
181	TCCCGGTCCA	CCCCCGGCGT	CAGATCGGTC	AGCGGGACGA	CCATCGGCGG	ATCCGCGCGG
241	ATCCGCAGCC	GCCCGTCGAC	CACCTCGGCG	AGCTTCCCGA	ACGCCGTAAG	GCTGTCCCGG
301	GACCGGGCCT	TCCCCAACGC	CCGGTCGACA	TTCCTGCGCC	CCCGCCCGCC	CAACTGTTCC
361	GTGGCCACCG	TGCGCAGCCG	CTCGGCATCC	GTCCGCGCGT	ACCAGACGTC	CAGATTGCGC
421	ATGCCCGCGA	ACCGGGCCAT	CGCCTCCCGG	TACGAGCGGA	CCGTGGCCCG	GACGACCCCG
481	GCCCGCTCCC	GGAGCGTGAA	GCTGTTTCGCC	CGCCCCGCGA	TGACGAGGCT	CGCCGCGAGC
541	CGCTTGACGT	CCCACTCCCA	GGGACCCGGC	AGCGTCTCGT	CGAAGTCGTT	GATGTCGAAC
601	AGGAGATTCC	GCTCGGGGGA	GGCCAGCAGC	CGGAAGTTCA	GCAGATGGGC	GTCACCGCAC
661	AACTGCGCCC	TGATTCCCGA	CACCGGGGTG	CCGGCCAGGT	CGGCGGCCAT	CAGCGCGGCC
721	GCTCCCCGGT	AGAAGCGGAA	CGGGGACTCC	GTCATCCGGG	CATAGCGGAT	CGGGACCAGC
781	TCGGGAACCC	GGTCCGCCGA	CTGGGCTTCG	AGGACGGTCA	GCGGATCGGG	GCGGCGCGGC
841	GACGGGGAGA	ACTCCGCATG	GCCCCACCGG	GGCGCCTCAC	GGCGGGCCCG	CTTGCCCACT
901	GCCGCCCGTT	CGGCCGCTGT	CGCGTTCGGT	GCGGCGCGCA	GCGGCGCCTC	GGCTTCCCGG
961	TCCATGACGT	GGCTCCTTCC	GGTCTTCCTC	AGGCCTGTTT	GCGGATCCCG	CGCTCACGGT
1021	GCGGCGCGGAG	GCGGACCCGC	TGCCAGATCG	CCCGCTGGGC	CTTGAGCGCG	AACCGGCCCCA
1081	CCACGATCAG	CACGAGGATG	TTGACGACGA	GCTGTATGGC	CGAGCCCCGT	ACGTCCGACC
1141	AGCTGGTGTA	CGCCGTGGAG	ACGGCGATGT	CCGCGGCGGC	CGGGATCGTC	GTCACGGAGA
1201	TGAACACCCC	GAGCAGAGCA	CTGGTTCTGG	CCTCGGTGAG	CGACACGATC	CCGACGATTC
1261	CGGCCAGGGT	GGCGACGGCG	ACGGAGAAGA	AGTTCGGCGT	GTTGATGAGA	TGGGAGACGG
1321	GCCGCAGCCC	CCGGTCGAAC	GCCTCCGACT	CCAGCCCGAA	ACCCCGGATG	AGGAGGGCGA
1381	AGAGGAAGGT	GACCACGATG	GTCAGGAGAA	GGCCGACGCC	CAGGGCGGCC	AGCCCGCTGC
1441	GCACCATGGA	CCGGTGGCGC	CGGTGATATC	CCAGCGCCAC	GCTGACGATG	GCGCCGTACT
1501	CCGGGCCGAC	GACCATCGCC	CCGACGATCA	GGATCTGCGA	GTTGGTGACG	ATGCCGACCG
1561	ACCCGATCAG	ACCGGCGATG	ACCAGGTAGA	GGTAGAAGCT	CGGCGGATAC	CGGCCCCCGG
1621	ACCTGATGCG	GGCCTCGACC	TGTTCCCGAG	CCGGCGCCCG	GCTCAGCGGC	CCCAGCTCGC
1681	GCTGCCCCGC	CTCGGTGGCG	GCGCCGGAGA	AGGCCATGTC	GACGGGTTCG	ATGACGAGGG
1741	AGCCCCGCGG	GTCGAGGTGG	GCGGCGCGCA	CCCGGTGCAG	TACGTGTTTG	GCCGCCCCCG
1801	TCAGTACGTC	GCAGGCGATG	GCGTCGCCGT	CGGGGCGGCG	CGCGGCGTGC	CGCTGGACGA
1861	TCAGATTGAG	CACGCACGGG	TCGGCCGAGA	GCAGGCCGAC	GACCTCGTCG	GTCAGGTCCG
1921	GCGGGCTCAC	CGCGCGGATG	TGGATCATGT	CCATCCCGGC	ACCTCCGCGG	CTCCCTGCCC
1981	CGTCACACGG	AGCTGTGCCC	GGCAGGCGGC	CCGGGGCTCA	CTCCAGTAAC	GCGGCACCGG
2041	CAACGTTCGG	CAAACCGGCG	GTCGCCCCGA	CGGGCCGGGG	CACCGGGGCC	GCGGGCGGGT
2101	GACCCGCCCC	GGCTGGGATC	ATGAAGGGGT	GGACACCCCC	GACACACCCG	ATGCCCTCCC
2161	CGAACCGCTG	CCCCGGGCGG	ACGAAGCGGT	CCGGGAGGAG	AGGGCCACCG	ACGACGGGAC
2221	GCCGGAGGGC	CGCCGCCTCG	TCCGCTGCCG	TCTCTGCGGC	CGGCCCCCTG	CCGGGGCCGA
2281	CTCGCGGCGG	GCCGGCCTCG	GCCCGTCCTG	CGACGCCAAG	CTGCACCCGG	CGCCGCCGGA
2341	CATCCGCACC	CGCCGCCACG	AGGTGACACCA	GGACCCGCTG	CCGGGCACCT	GAGCCGGAAC
2401	GGGGCTACTG	CTCCAGCCGC	CGGAACAGCC	CCTCCTGCAC	CACCGACACC	AGCAGCTGCC
2461	CCGAACGGTC	GATGATCCGC	CCCCGCGCCA	GGCCCCGCCC	GCCCGTGGCG	ATCGGCGACT
2521	CCTGGTTCGT	CAGGAACAC	TCGTCCGCCC	GGAAACGGCC	GTGGAACAC	ATGGCGTGGT
2581	CCAGGGACGC	AAGGTCATAT	CCGCGCGGGC	CCACAGCGCG	CTCCACCGCG	ATACGGACCG
2641	CGTCCAGCAG	CGTCATGTCG	CTCGCGTACG	TCAGCGCGCA	CGTGTGCACC	ACGGGTTCGT
2701	CGCCCAGCGG	GCCCACCGCC	CGCATCCACA	CCGCGCTGCG	CGGATCGGCG	TCCTGGATCT

2761 CGTCCTTCGT CCAGCGCAGC CGGTGACGT AACGGATGTC GAAGGGCTGG CGGCGGGCCA
2821 TCCGCTCCAG CGCCTCCGGC AGCGCGCCCA GATGCTCGCG CACCTCCTCG GCGACCGTCG
2881 GCAGCTCCTC CGGGTCCGGG ACGATCCGGG CGGGCGGCAG CTGGTGTTCG AAGCCCGCCT
2941 CCTCGGGGCG GTGGAAGGAC GCCGTCAGGT TGAAGATCGT CCGGCCCTCC TGGACCGCCG
3001 TCACCCGACG GGTGGTGAAG GACCGGCCGT CCCGCACCCG CTCCACCTGG TAGACGATCG
3061 GCACACCGGG ACGCCCCGGC CGCAGGAAAT AGGCGTGCAG CGAGTGCACC GGCCGCTCCC
3121 CGTCGGTGGT CCGGCCCGCC GCCACCAGCG CCTGGCCCGC GACCTGCCCG CCGAAGACCC
3181 GTTGACAGGA CTCCTCCGGG CTGCGCCCCC GGAAGATGTT GACCTCGATC CGCTCCAGGT
3241 CGAGCAGGTC GACCAGACGC TCGGCCGGAT TCGTCATGCC GCACCTCTCC CGTCACACGT
3301 CAGGGTCCGC TTCACAGCTG GCCGACCTCG GTGACCCGGA CGACCGCCCG GCCCTCCTCG
3361 TCGGACGCCG CGAGGTCCAC CTCGGCCTTG ATGCCCCAGT CGTGATCGCC CGCCGGATCG
3421 GCGAACGCCT GCCAGACCCG CCACAGCCCG TCGCGCCGGT CCTCCTCGAT CTTACAGCAGC
3481 TTCGGGCCCC GCGCGTCCGG ACCGCTCCCG ATCTCCTCGT GCGCGTCCCA GTACGCGTCC
3541 AGCGCCTCGC CCCACGCGTC CTCGTCCAC CCGGACTCGC CGTCCAGCTC GCCAGCGCGC
3601 CCGGCCCGGT CCAGCGCGGC CAGCTCCACC CGGCGGAACA TCGCGTTGCG CACCAGCACC
3661 CGGAAGGCGC GCGCGTTCGC CGTGACCGGC TTGACCTCGT CCGCCTTCTC CTGAGCCTGC
3721 TCCGCGGTCT CCACCTCGGG GTTGGCCAGC TGCTCCCACT CGTCCAGCAG ACTGGAGTCC
3781 ACCTGACGCA CCATCTCGCC CAGCCAGGAG ATCAGGTCCT GGAGGTCCTC CGACTTCACG
3841 TCGTCCGGGA TCGTGTGCTC CAGCGCCTTG TACGCGCTCG CCAGATACCG CAGCACGATG
3901 CCCTCGGTCC GGGCCAGCTC GTAGTGCGAA GTGAAGTCCG TGAACGTCAT GGGCCGCTCG
3961 TACATGTCCC GGATCACCGA CTTGCGCGAC ACCGGATGGT CGTTCACCCA CGGGTGGCTC
4021 GTGCGGTACA CGTCGTACGC GTGCCACAGC AGCTCGCTCA GCGGCTTGGG GTACGTGACC
4081 TCCTGGAGCC GCTCCATCCG CTCCTCGTAC TCGACCCCGT CCGCCTTCAT CTGCCCCACG
4141 GCCTCGCCGC GCGCCTTGTT CTGCTGGGCG GCCAGGATCT GCGGGGATC GTCCAGCGTC
4201 GACTCGACGA CCGAGACCAT GTCCAGCGCA TACGACGGCG ATTCGGCGTC CAGCAGGTCG
4261 AACGCGGCCA GCGCGAACGT GGACAGCGGC TGGTTCAGCG CGAAGTCCTG CTGGAGGTCG
4321 ACCGTCAGCC GCACGATCCG GCCCTCGGCG TCCGGGGTGT CCAACTGCTC CACCACCCCG
4381 CCGTCCAGCA GCGAGCGGTA GATGGCGATG GCCCGCCGGA TGTGCCGAG CTGCGCCCCG
4441 CGCGGCTCGT GGTGTCCCTC CAGCAGATGC CGCATCGCCT CGAAGGCGTT GCCCGGGCGG
4501 GCGATGACCG CGAGCAGCAT CGTGTGGGTG ACCCGGAAAC GGGAGGTCAG CGGCTCCGGC
4561 TCGGACTGGA TCAGCTTGTC GAACGTGCAC TCCGACCAGG CGACGAAGCC CTCGGGGGCC
4621 TTCTTGCGGA CCACCTTGCG CTTCTTCTT GGGTCGTCGC CCGCCTTCTT CAGCGCCTT
4681 TCGTTCTCGA TGACATGCTC GGGGGCCTGT GCCACGACGA ACCCGGCCGT CTCGAACCCG
4741 GCGCCCGCCG CCGGCCCGC GATCTGGTGG AACTCCCGCG CGCGCAGCT CCGCACCAGG
4801 TTCCCGTCGT ACTTGGTGAG CGCCGTGAAC AGCACCGTAC GGATGGGGAC GTTGACGCCG
4861 ACGCCGAGCG GTGCCGTCCC GCAGATCACC TTCAGCAGCC CCGCCTGGGC CAGCTTCTCC
4921 ACCAGGCGGC GGTACTTCGG CAGCATCCCC GCGTGGTGCA CCGCATGCC GTGGCGTACG
4981 TAACGGGAGA GGTCTTGCC GAACCTGGTG GTGAAGCGGA AGCTGCCGAT CAGATCGGCG
5041 ATCTTCTCCT TCTCCTCCTT CGTGACATG TTGATGCTCA TCAGCGACTG CGCCCGCTCC
5101 ACGCCGCCCG CCTGCGTGAA GTGCACGATG TAGACCGGCG ACTGCCGGGT GTCCAGCAGC
5161 TCGGTGAGCG TCTCGTGAT CCGCGTGAAG CCGTACTCGT AGCTCAGCGG CACCGGGCGG
5221 GTCGCCGAGC GCACCACCGA GGTCGGGCGG CCGGTACGGC GGGTCAGGTC CTTCTCGAAC
5281 ATCGAGACGT CGCCGAGCGT CGCCGACATC AGCACGAACT GCGCCTGCGG CAGCTCCAGC
5341 AGCGGAATCT GCCAGGCCCA GCCCGGTCC GGCTCGGCGT AGAAGTGGA CTCGTCCATC
5401 ACGACCTGGC CGATGTCGGC GTACTTGCCG TCGCGCAGCG CGATGGAGGC CAGCACCTCG
5461 GCCGTACAGC AGATCACCGG GCGTCCGCG TCGACGAGG CGTCGCCGGT GAGCATGCCG
5521 ACGTTCTCGG TGCCGAAGAG CTTGACAGG TCGAAGAACT TCTCCGACAT CAGCGCCTTG
5581 ATCGGAGCCG TGTAGAAGGT GACCTTGTCC TGGGCCAGCG CCGTGAAGTG CGCGCCCGCC
5641 GCCACCAGGC TCTTGCCCGA GCCGGTCGGG GTGGACAGGA TCACGTTGCG CCCGGAGACC
5701 ACCTCGATCA GCGCCTCCTC CTGAGCCGGG TACATCGTGA TGCCCTGGCT CTCGGTCCAT
5761 GAGGAGAAGG CCTCGAAGAG GGCGTCCGGG TCGGCGGTCT GGGGAAGCTG GTCGATGAGG
5821 GTCACGCCCC CATCTTGCCG GTCTTCCGCC CGGATGAGGG AACC GGCGGA CCGCACGAAG
5881 ATCACGACG GTACGCTGCG GACTCAACCT GCCCGCGCCG CACCGGTGAT GGGCGCACGA
5941 ACCACTGGGG GCGGGACAGA CCATGATGGG ACCGGCACAC TCTCTGTGAG GGGCAGCGGC
6001 CTGGCTGGGG GTGGGCGCGG CCGCCGCCGC CGCGGGCCAC ACGATGCCCT GGCCCGTCCCT
6061 CGTCGTCCGG GCGCTGATCT GCGCGGGAGC CGCACTCGCC CCCGACCTCG ACCACAAGTC
6121 CGCGACCATC TCGCGCGCCT TCGGCCCGGT CTCCAAAGCC CTCTGCGAGA TCGTCGACAA
6181 GCTCTCTAC GCCGTCTACA AGGCCACCAA GAGCGCCGGG GACCCCGCA GGACCGGCGG
6241 GCACCGCACC CTCACCCACA CCTGGCTGTG GGCGCTCCTC ATCGGCGGCG GCTGCTCCGT
6301 GCGGGCGATC ACCGGCGGCC GCTGGGCCGT CCTCGTGATC CTCTTCGTCC ACCTCGTGCT
6361 CGCCGTGAG GGCCTGCTGT GCGGGGCCG CCGCGTCTCC AGCGACGTTG TGGTGTGGCT

6421 GCTCGGCGCG ACCAGCGCGT GGATCCTGGC CGGCGTCCTG GACAAGCCCC GCTACGGGGC
6481 CGACTGGCTC TTCGACGCCC CCGGCCAGGA GTACATGTGG CTCGGCCTGC CCATCGTGCT
6541 CGGCGCCCTC GTCCACGACA TCGGCGACGC CCTCACGGTC TCGGGCTGCC CGATCCTGTG
6601 GCCCATCCCG ATCGGCCGCA AGCGCTGGTA CCCGATCGGC CCGCCGAAGG CCATGCGCTT
6661 CCGGGCCGGC AGCTGGGTGG AGATGAAGGT GCTGATGCCC GCCTTCATGG TCCTCGGGGG
6721 AGTGGGCGGG GCCGCCGCCC TCAACTACAT ATGACGCACC GCCGGTCGCG CCCGGTGCCT
6781 CTCCGGCGGG CGGCGCGTCC GGTGTCTTC CGGCGGGCGG CGCGCTCCGC CCCGTAGCAC
6841 CATGGGCGCA TGCTGCTCGC CGAGCTCGCC CAGGTGTCCC TGGAGGTCCG GCCACCTCC
6901 GCCCAGTCCA AGAAGGTGGC GCTCCTCGCC GGA CTCTTCC GGGACGCCGG ACCCGAGGAC
6961 GTCCCCGTGC TCATCCCGTA CCTCGCCGGA CGGCTGCCCC AGGGCCGGAT CGGCGTGGGG
7021 TGGCGCTCCC TCGGCGACCC GGTGGAGCCC GCGGCGGAAC CCACCCTCAC CGTCACCGGC
7081 GTCGACGCCC GGCTGACCGC CCTCGCCGCC GTCTCGGGCC CCGGCTCCCA GGCCCGGCGC
7141 AAGGAGCACC TCGCGCCCTT CTTCGCCGCC GCCACCGAGG ACGAACAGCG CTTCTGCGG
7201 GCCCTGCTCA CCGGCGAGGT ACGCCAGGGG GCCCTGGACG CCCTCGCCGC CGACGCCCTG
7261 GCCCGCGCCG CCGACGCCCC GCCCGCCGAC GTCCGGCGCG CCGTGATGCT CGCCGGATCG
7321 CTCCAGGAAG TCGCCGGGGT CCTCCTCGCG GACGGGCCCC AGGCGCTCGC CGCCTTCCGG
7381 CTCACCGTCG GACGGCCCCG CCAGCCGATG CTGGCGCACA CCGCCGCCTC GGTGCGCGAG
7441 GCCCTCGACA AACTGGGCGC GTGCGCGGTC GAGGAGAAGC TCGACGGCAT TCGGGTGACG
7501 GTCCACCGCG ACGGCGACCG GATCCGCGCC TACACCCGGA CCCTCGACGA CATACCGAC
7561 CGGCTGCCCC AGCTACCGC CGCCGTGCGC GCCCTCCCCG CCGGCCGCTT CATCTGGAC
7621 GCGGAGGTGA TCGCCCTGGG GGAGGACCGG AGGCCCCGGC CTTTCCAGGA GACCCCTCC
7681 CGGGTGGGCT CGCGGCGGGA CGTGGCGGAG CGGCGGCGC ACGTGCCCGT CGCCCGGTC
7741 TTCTTCGACG CGCTCCTCGT CGACGACGAG GACCTGCTCG ACCTGCCCTT CACCGACCGC
7801 CACGCCGCCC TGGCCCGGCT CCTCCCCGAG CACCTGCGCG TCCGCCGCAC CCTCGTTCCC
7861 GACGCGGAGG ACCCGAAAGC CCGCGCGGCG GCCGACGCGT TCCTACCGA CACCTGGAA
7921 CGCGGCCACG AGGGAGTCGT CGTCAAGGAC CTCGCCGCCG CCTACAGCGC GGGCCGCCGG
7981 GCGCGTCTCT GGCTGAAGGT GAAGCCCGTG CACACCCTGG ACCTGGTGGT GCTGGCCGTC
8041 GAGTGGGGCA GCGGCCGGCG CACCGGCAAG CTCTCCAACC TGCACCTGGG CGCCCGCCGC
8101 CCCGACGGTA CGTTCGCGAT GCTCGGCAAG ACCTTCAAGG GGCTCACGGA CGCCCTGCTC
8161 GACTGGCAGA CCCAGCGCCT GGGCGAGCTG GCCACCGACG ACGACGGCCA CGTCGTACC
8221 GTACGCCCGG AACTCGTCGT GGAGATCGCC TACGACGGAC TCCAGCGCTC CACCCGCTAC
8281 CCCGCCGGGG TCACCTCCG CTTCGCCCGC GTCCTGCGCT ACCGCGACGA CAAGACCGCC
8341 CAGGAGGCGG ACACCGTGGA GACGGTCCTG TCCCGGCGGC GGTGAGCGCG CCCCCTCTCT
8401 GAAGGGGCGC GCTCGTACAG GGCCCGGCGA CTCAGTGCTT GATGACCGTC GCCCGCTGT
8461 GCTCCTTGAT CTGCTCGGGC GTCAGGTAGA GTCCCGTGTA CTCGAAGTCC CTGCACTCG
8521 CCGGCTTGCG GGA CTGGAAC CCGGTCCGTA CGAAGTCGTC ACCGGGACC GCGTTCAGCA
8581 GCCAGTTCGT CATGACGCGG GTCTTCGCCA CGTTCGTCCG CAGCGCCGAC CAGTGGTAGC
8641 CCCGGGCCAC CGCTGCGCG GGCAGCCCGC GCAGCTCGAT GCCAGCGGC TTGGACACGG
8701 CGTCCGTGCC GCCGAGGTCC ACGACGAGCC CCAGATCCTT GTGCACGTAG TCCTTGAGCG
8761 GCTCGTGGCG CAGCGAGGCG ATCAGGTTGT CCGCCAGCAC CCGGCCCTGA CGCATCGCGT
8821 GCTGTGCGGT GGGCGGGCAG ACCGCCCGCT CGCCCTTCGC CAGATCGGGC ACGGCGGCCG
8881 CGTCGCCGAG CGAGAACACC CCGTCCGCGC CCGGCAGTCT CATCTGCGGG GTCACGGCGA
8941 GCCGGCCGCG TACCGTCTCC GCGCCGAGCG TGGCGACCAG CGGACTCGCG GCCACGCCGG
9001 CGGTCCAGAT CAGCGTCCGG CAGGGCAGCA CCCGGCCGTC GGTGAACGTG ACCTCCTCCG
9061 GCCCCGCCCT GCGGATCGAC ACCCCGAGCG ACACCTCGAT GTTCCGCTTG CGGAGCACCT
9121 CCAGCGCGGC CTGCCCGAGC TTGTGCGCGA GCTCCGGCAT CAGCTTCGGC GCGATGTCGA
9181 TCAGATGCCA TTTGATCAGG CGCGGGTCAA GACGCGGATA GTGCTTACC GCGTTGGTGG
9241 TCAGACGCTG GAGACAGGCG GCCGTCTCCG TGCCCGCGTA CCGCGCCCG ACCACCAGCA
9301 ACTGGAGCCG GGAGGCCCCG TCGGCCTCGT CGTGACTGGC GTCCCGCAGG TCCAGCTGGG
9361 CGATGACGTG ATCCCGTACG TACGCGGCCT CGGCCAGCGT CTTTATCCCC CGCGCTTGT
9421 CCAGCAGCCC GGGGATGTGC AAGGTGCGGG TGACGCTGCC CGCCGCCAGC ACGAGGTAGT
9481 CGTACGGCTC GTTCACGATC TCGTCCGTGA TCTTCCGGAT CACACAGACC TTCGCTGCG
9541 TGTCACGCC GATCGCCCCG CCCGGCACGA TCCTGGTCCG GTGACGGCGG CTGCGGCGCA
9601 GCGACACCGC CACGGA CTGC GCGGTGAGCA CCGCGAGGC CACCTGGGGG AGCAGCGGCA
9661 GATACAGCTG GTAGGAGAAC GGTGTGACGA GCGTGATCTG GGCCTCGCCC GGAGCGAGCC
9721 TCGCTCCAG ACGGCGTACG CACTCGACGC CTGCGAAGCC GCGCCGACG ACGAGAATCC
9781 TGGGTGGTGC CACGGTCTGC GTCCCTTCTC GGGCTTGCGT GGTTCGCGC TCGCCTGCCC
9841 CGTTTACCGG GTGATTACCC CCTCATCCTC ACCGGAGGCT CCGGCATCCG CCTCCTGGCA
9901 GGGGTGAAAC GGGGCCCGGT CACAGGGGCG GGGCGGCCGC CAGCTCCCGC AGCCAGTGGG
9961 CACCTCGGC GTCCCCGGCC ACGCCCGGAC CACCCGGCGG TACGGGGAAC ATCCGCCCCC
10021 ACGAGGCGGC CCGGCCAGC GCCGCCAGCC GCCACGCCAG GTCACCGCA CGGCGCAGCC

10081 CGGCCGCCGT GACGCCCCCG CCGGTCCACG GCTCCAGATA GGCGTCCCGC AGCCGGGGCA
10141 GCACCTCGGG ACCACAGCGC TCACGGGCGG CACGGGCGGG TACCAGCAGG CTGCAGAACG
10201 GATGGCCGAC GAGGGCGTCC CCCAGTCGA AGAAGGCGTA CCGCCCGGAC ACGGGCGCGA
10261 ACAGCTGCTT CTCGTGCAGA TCGGCGTGGT CCAGCGAGTC CGCCACCCCC GACGACGCCA
10321 GCTCCTCGCA CCAGTCGGCC ACCCGGGGCG GCAGCACCTC CAGCGCCACC CGGTCTCTCC
10381 GGGGCAGCGC GGCCTTCCCC GCGACCAGCC GGTCTGAACAG CGCGGGAAGG TCGCGCGGCC
10441 GGGCCGCCGG AACCCCCAGG GCCTCGATCG CCTCCGCGTA CGGGGTACAG TCCCGCTGCA
10501 TCGCGGCGTA CTGGCGCAGC GGCTCCTCCC AGTAGCCGGG GTCAGGGGCG CCGGGACGCC
10561 CGTCGAGGAC CTCCGACAGC ACCGGGCGGC CGTCCGGGAC GAGTATCCAG CCGCGTTCCG
10621 CCTCGACGGC GAGCGGGGCC AGCACC CGGGACCCA GCGCGCCAGC GCCTCGGTGA
10681 GCCCCGCCCTC GAAGGCCGCG GCGGGCGGAA CGGCCTTGAA CCAGACGGGC GCGGGCCCGG
10741 CGACGGCCAG CCGCACCAGC ACCGACCAGG GACGACGCCG CACCGCCCGG GGGCCCGTCT
10801 CCGTCAGCCC GTGAGCGGCG AGGCCCTCGG TCACCCAGGC GAGGGCCTCC GCCCGCCAGG
10861 CCGGGTCTCT CCAGGGCGTC ACGGCGTCCG GGAAGCGCCC CCGTCCACG GTCGAGGTCT
10921 TGTCGGTCCA GGTACGGTCT CTTCCGGTCA CGGCGGTCTG GGTCTGCGG GGGCCGTCTA
10981 CGGCGGTCTG GGTACGGCA TCCGGGGCGG CGTCCGGCAT GGGCATCTCG TCTCCGCGCA
11041 TCCGATCATG GGATCACCGG CCCC GGCGCG TGCGCACCGC AATTTCCGGG AACACCCGTT
11101 CCCGTCTTGC CCGGATCGGC TGTCCTCCCC CCTCCGGCCC CTGGAACGGC GGGAGTTCGG
11161 CCGCCCGCCC CGTGCAGGA TGCTGTGGTG ACCACCTCGC CTCCTCGCC CGTGGCCGAC
11221 GACTCTTCCG TGCTTCCGT GGACGACGCC CCGCCCCGCG ACCAGGGGCT GAGTCTCCCG
11281 GCCGCGGCGG TCCTCGTCTT CCGGTCTCTC GCCGCGGTCC TCGTGGTCTGA GATCGTCGCC
11341 CTGCGGCTGC TCGCCCCGTA CCTCGGCCTC ACCCTGGAGA CCAGCACGCT GGTGATCGGC
11401 ATCGCGCTGA CCGCCATCGC CCTGGGTTCC TGGCTGGGCG GGCGCATCGC GGACCAGGTC
11461 GATCCGCACC GGCTCATCGC CCCC GGCGCTC GGGGTGTCTG GCGTGGGCGT CGCGCTCACC
11521 CCGTCTCTGC TCCGTACCAC CGCGGAGTGG TCTCCCGCGC TGCTCTGTCT GGTCTGCTTCG
11581 GCGACCTTCC TGGTGCCGGG CGCGCTGCTC TCCGCGGTGA CCCC GTTCTGT GACGAAGTTG
11641 CGGTTCACCA GCCTCGCCGA GACCGGGACG GTCGTCTGGG GGCTGTCTGG CGTCTGGCACC
11701 TTCGGAGCCA TCGTCTGGC GGTGCTCACC GGATTCTGTC TGGTCTACCG GCTGCCCCGT
11761 AGTCCATCC TGATCGGCCT CGGCACGCTG CTGGTGTCTG GGGCGGCCCT CGTCTGGATG
11821 CAGGCCCCGG GGTGGCGGCG CGCCACGGCC GTGGCCCTCG CCACCGTCTG CGCGGGCACT
11881 CTCGCCACCG GGTTCGCTCC CCGCGGCTGC GACGCGGAGA CCGCTACCA CTGCGCCCCG
11941 GTCGTCTGCG ACCCCGACCG GGACAGCGGC CGCACCTCTG TCCTGGACGG CCTGCGCCAC
12001 TCCTACGTGC ATGTCGAGGA CCCCAGATAC CTGAAGTTCT CGTACGTACG CGCCTTCGCC
12061 TCCGTGGTCG ACACGGCCTT CCCCAGGGC GAGCCGCTGA CCGCCACCA CATCGGGGGC
12121 GCGGCGCTCA CCTTCCCCCG CTACCTCGCG GCCACCCGCC CCGGAACCCG CAGCCTCGTC
12181 TCCGAGATCG ACCCCGGGGT CGTCCGCATC GACCGCGACC GGCTCGGCCT CGGCACCCCT
12241 GCCGCGACCG GCATCGACGT ACGCGTCGAG GACGGGCGTC TCGGCCTGCG GCGGCTGGAC
12301 GCGGGCAGCC ACGACCTGGT CGTCTGGCGC GCCTTCGGAG GCGTCAGCGC GCCCTGGCAC
12361 CTCACGACGT CCCAGGCACT CAAGGACGTA CGCCGGGCGC TCGACCGGGA CGGCCTGTAC
12421 GTCACCAACC TCATCGACCA CGGCCGGCTC GCCTTCGCCC GCGCCGAGGT CGCCACCCTC
12481 GCCGCGACCT TCCCGCATGT CGCGCTGCTC GGGCAGCCCG CGGACATCGG CCTTGACCCC
12541 ACGGCTTCGA GCATCGGCGG CAACATGGTG GTCGTCTGCT CCGCCCGGCC GCTCGACGCC
12601 CCCGCCATCC AGGAAGCCAT GGACGCCC GGAGTCTGGT GGAGGATCGC CACCGGCGAC
12661 ACCCTACCA CCTGGACGGG GAACGCCC GTGCTACCG ACGACCACGC GCCGTCGAC
12721 CAACTCTCTC AGCCCCACCG CGTCCCAGCG GCCCGGTGAG GCCCGAACGG GCCCGATGAT
12781 CCCGCCGAA CGCCCCGTG ACGACGAAC AGCCGTGTA TCCCGGCCG ATCGCGGGG
12841 GATCGCCGGG CGGTTCGGCC AAGACGCCTC ACCCGTGCCA GGACCGCCAC AGCGACCGT
12901 ACGCGCCGCC CGCCGCCACC AGCTCGTCAT GGCTGCCAG TTCCTGATC CGGCCGTCTT
12961 CCACGACCGC GATCACATCC GCGTCTGCG CGGTGTGAC CCGGTGCGCG ATC

SEQ ID NO: 104

MAQPIELVIFDCDGVLDVSDSERIAVRVDALVLAELGWNLTAEIIVDRFMGLSSRSMTTRQIEDHLGRRLPADWEEEFK
PLYDEALAAELTPVEGIVDALDALTHLPTCVASSGSHDKMRFTLGMTGLRPRFEGRIFSATEVEHGKPPADLFLHA
ARKMGVVPEACAVVEDSQYGLQAARAAGMRAFYAGGLTPADRLEGPGTVVFDMMRRLPGLLADH

SEQ ID NO: 105

Nucleotides 89908-90561 of SEQ ID NO: 1

SEQ ID NO: 106

1 CGCCGTAGAG CCACGCAGGA GAGGCGGTGCG GCCCATCGGG GTCGACCGCN TCTCCCGTTT
61 CCCTCAGGTC AGGTACAGCAC GCGCCCGCCG GCGGTCCCCG CTTACAGACG TTCGTGTAAC
121 CCGTCGTGCC GTCCAGCCAG ACGACGCGCT TGCCCTCGTC GGCCGAGGCG TAGACCTGGT
181 CGCCGCGGTT GCAGGACACG CGCTGGGCGC CGCCCTTGCC GTCGGGTGCC ACCTGGAACA
241 GCTTCGGCAG GGTCTCGTTG GCCCATGAGG TGGCCGGCGG CAGCGAGGTG ACGTTGACCG
301 CGTCGTGCGA GGCCGTGAGC TGGTAGGCGT GCAGGGAGCC TTGGCCGGTT TCGGGGACGA
361 CCGTGGTGCG GTTCGTGCCG TCGGCGTCCG CGCGCACGAC GGCGGCCCTG GCGGGGTCCG
421 TGCCGCGGCC GTCCCTCGATC CAGAAGACAC CGTCGTGCGT GAAGGCCGTC TGGCCGATGC
481 CCTGCGCGCC CTCGGCTCCG GCCGGCATCA TCCGGGTGGT GCCGGTGGCG GTGTCCAGGA
541 CCCCAGTGCC CAGTTGGACG TCGCTGCCGT CGATCCACAT CGTGGCGTAG GCGATCTTCC
601 CGTCCCGGAT CGAGGGCATG GCCGTGACCT CGATGCCCTT GGCGAGGATC ATGCCCGCCT
661 CCCGGGTCTT GATGTTGATG TAGCCGATCT GCTGCCGGCC GGTCTGCGGG TTCGTACCCG
721 TGTAGGCGAG GAAGTCGCCT TCGAGGACGA GCTGCGAGAC ATTGGTCTCCG ACGCCGCCGA
781 CGACCCTGGC CGGGCCGCCG CCCACCGGAC GGACATGGAA GAGGACCGAC CCCGGGGAGA
841 ACTCGGCCCA GGCGACGTTT TTGCCGTCCG TCGTCGCGTA GACGTTGTAG TTGCCGTTGT
901 TGCCGCTCAC CAGCTGCGGC TTGCCCTTGC CGTCGGTGTT GCCGAGCCAC ACCGAGTACG
961 GTTCGCTGCC TTCTCGTCG GAGCGGGAGA CGGCGTACTT GCCACCGCCG GCGCCGACGC
1021 CCGTGCCCGC CACGTTGACC TTGGACACGA GGGTGTCCGT GTCGACACCG AGCGCCTTCT
1081 CCCAGCCGGG GACGATGCCG TCGCGCTCGA ACGAACGCTT CACGATCTTC GTCTGCGCCT
1141 TGGTGGCGCC CAGCTCCTGC GCTGCGCGCA TGACCGCCGC GCGGCCCTCG GTGAAGCCGT
1201 CGAGCGCGCT CATGTACGCG GACAGCGCGC GGTAGACGAT CTGGTCGGCC GGTCTCGGCG
1261 CGAGTCTCTC GCGCATGTCC CACAGGGCGC CCGAGAAGAT CGTGGAGTTG AGGTGTACGC
1321 CACCGTTGTC GCCGTTGTGG CTGACGCCGA TGAAGTCCTT CGACGTGGTG GCACCGTCGT
1381 TGAGGTCCCG CAGCGCGCAG TCGCGCGGGC CGAGCGTGGT GCACAGGTCC TCGCCGAGCA
1441 GACCGGCGTC CGGGTCGTCC ATCGACCGGC CGTCCGCCCTC CAGGTCGATG GCGTTGCCGA
1501 AGTAGTCGGC TATCGCCTCG TTCACGGCGC CGGACTGGCC CGCGTACACC AGGTTCCGCG
1561 TGTGCTCGAT GACACCGTGC GTCATCTCGT GGCCGACCAC GTCGGTGTCC GCCGAGAAGG
1621 TGCGGTAGTC GCCGCCGCCG TGGCCGTACA CCATCTTGGT GCCGTCCCAG AAGGCGTTGT
1681 TGAACGGCTT CCCGCCGTTT ACCACACCCA CGAGCGAGTA CATGAACCCG TCCTTGCCGT
1741 CCAGGCTCTT ACGGCCGAAG TGGTTCCTTG AGTAGTCGTA GACCTTCCCG GCCGCCAGT
1801 GGGCGTCCAC CGCCCCGGAG TCGGTGAACT CCTTGCCCAG GTCGGGGGTG TCCGAGCGGA
1861 ACGTCTGGAT GCCGCTGGGC CACGTGCCGG ACGCCGTGGC GGCTTCCCG CCACGCGCGT
1921 CCCAGGTCGA CAGCAGCGGA CCGTCGTACG GGCTGTGCGC GCGCCGCTTG CTGTAGTCGT
1981 ACATCTGGTA GCGCGCGTCG GCGCCCTGGT ACAGGTTGAT CTCGCCCTCC TCGCCGTTGT
2041 AGCGGGTGCC CGAGCCCTTG ACGACAGCT GCTGCGTGGC CTCCGCGGTG TCGCTCGGCCG
2101 TGGGCGCGTC GCCGGTGCGC CCGCTGCTCG TACCGGCGCC CGAACCCGGT GCGCCGAACG
2161 ACTTGATGCC GCTGTACTGC ATCACGGGGA AGCCGGAGTG GGCGTCGATG TACACCTCCT
2221 GCTTACCCGG CGTGCCGTCG GCCGGGCTGA CGCCGGTGAC GGTGATGTGC CGGGTACGCA
2281 CACCGGATCC CTGGGGCAGG ACGGTACGT CACCGGCCGT ACCCGACAGC TGAACCCCGC
2341 CGGGCTTCTC GCCCTTGGCC GGGGCCCTGA GCAGGCCCCC GCGCAGCTGC GCGGAGACGG
2401 CGCCGACCGC GCGCTCGATC GCGGTCTTCT CGGAGACCTC CGGCGTCACC GTGTCGAGCT
2461 TCAGCTCGGT GAAGTAGCTG CCGGACGTGC CGGTGACGGT GCGCTTGCCG TCCTTCTTCG
2521 TCATCCGGAC CAGGTACTGG CCGCCGAGGA CCGTGACGCC GCGGTACTTC TGCTGGAGAC
2581 GGACCGTCTC CTGCGCCGCC CTCCCGGTG TGGTCCCGGC CGCGGCCAGG TCCTTCGCCG
2641 AGGTGTGCGC GATGCGGTAG CGGCTCTTCT TCGCCTTGAG GTGGCCACGG GCGGCGTCCG
2701 CCGCGGTGCC GGTGGCGTCG GCCTCCTCGT TGAGTCCCGT GACGAGCGAG GGGGTCCCGG
2761 TCCCGGCACC CGGCGTCACC TCGGACGCGG CGGACGTGGT GGCGGTGTCC GCGTGAGCGG
2821 CGGGCGTGCG CGTCATCAAC AGGCCCGCGG CGGCCAGTAC CGCCACCAGC CCTTGGCCGA
2881 CGAACGTGGT GCCGGACCCC TGACCTGGTC CGGCCGATCT GCTGCGCGAT GAACTCCGCA
2941 CAGTTGCCCC TCCCGAATGC GATTGCTGCA TGAACGGATC TCGGATGCCG ACATGGAAGT
3001 ACCGGGGCGG GGGTGGGAAC AATGCGTTCC GCGCGTGAC TGTGAGACGT ACACACTTGT
3061 GTGGCCTGCA TTGACAGGTG ACAATTCCGC CGGTGGGGCG GACTGGGTGG GACACGCAAT
3121 GGCAGGCGCT CGGGCTCGGA GCGGACGAAC TCCGGGTGTA CGAAGCGCTC CTCGACGCAC
3181 CCGTACCCGC CTCCCGTACC GCCCTCGCCC AGCACCTGGG GCTGACCGTG CGCCGGGTCA
3241 CGAACGCCCT GGATCACCTG GCCGAGCGCC GCTTCACCCA CCCGGCCCGG GCGCGCGGGC
3301 TGCCCGTGCC CATGGCCCCC GCCACCGCCC TGCGCAATCT GATCCATCTG CACCAGGCGG
3361 AACTCCTGCA CCGATCGGCG GAGTTGGAGG AGCTCAGCGG ATCCGTGCGC CCGATCGCGG
3421 CCCAACTGCT GAGCTCCGTG CACACCCCCC GCGCCACCGG CATCGAGACG GTCCGCGGCG
3481 GAGCGGCGAT CGCCGCCCGG GTGGCCTCCC TGCTGGTGTC GGCGAGCGAG GAGATCGCCC
3541 TCCTGGACCG GCCGCCGTAC GCGTCGAGCG AGCCGGACCG GATGCCCGTC CCCCTCGACA

3601 TGGCCGAACC GGTACGTCGG GGGGTACGCG TCCGGGTCGT GGTGGACCGC GAGGGGCTGA
3661 GCTTCCACGG GCGCGCCCCG GGCCTGGGCG ATCTGGCCGT GCAGGGGGTG CAGATCCGGG
3721 TGGGCACGGA TCTGCCGACG AAGCTGATCA CGGTCGACCG GCGGGTGACC CTCCTGCCGC
3781 CGACGGACCG GCGGGACCCG ACGGCCTCCG CACTGGTGGT GAGCGACCGC CTGCTGAGCA
3841 ACGCGCTGGT GCCGCTGTTC GAGTCGGTCT GGAACCGGGC GCTCCCCATC GGCTCCGTGA
3901 CGCATGATCA GATCACGGAC GCGCACCGGG AGTTGCTGAC GATGCTCGCC TCCGGCCTGA
3961 AGGACGAGGC GATGGCCCCG CGCCTCGACA TCCATGTGCA CACGGTCCGG CGCCGCATCA
4021 CCCGTCTGAT GCAGGTCCCTC AACGCGGAGA CCCGTTTCCA GGCCGGGGTA CAGGCCGCGC
4081 TCCGGGGCTG GCTCACCGGG TGAGCCAGCC CCGGCCCCCT CAGCCCGCCT TCACCGTCCC
4141 GCCGAGACGG GCGCGGTGCC GCGCAGGAC GCACGGATGA TCAGGTGGG CCACAACTCG
4201 ACTCGGTGCA CCGGGCGTTC GACGGGATCG ACGGCGCGTG CCAGGGCCAT TTCGGCGGCC
4261 AGGACGCCGA GCCGGTGCTT CTGCGGGCGC ACGGCGGTCA GGGGAGGGTC GCGGCGGGCT
4321 GCCACCTCGT CGTCGTAGGT GATGACGGCG AGGTCCTCGG GGACCGCGAG TCCGTGTTTCG
4381 TGGGCGCGCT CCACCAGGCC GACGGCCTCA CCGTCGGAGT GGACGAACAG GCGGGTGGCG
4441 CCGGCCTCGC GGCAGCTGCG GAGGGCCGCG TCGTAGGCGT CCGCCAGACC GGGGGAGCCG
4501 TACGACGGTA CGTCGTGTGC GAGGCCGTCG CCGGAGGGCA GGCCGAGAGA GCGGGTGGTC
4561 TCCTGCCACC CCTCGCGCAG GGCCTTGTTG GTGGGGCTGA AGCGCGCGGT GACGAAGGCG
4621 ATACGGTGAT GGCCAGGGT GACCAGATGC GTACGGCCA GCGCCGCGCC GCGCTGTGG
4681 GCGGTGGTGA CCGCGTCGAG CCGCAGGGTG GGCAGAGCGG GCGGCGGCAG CCGTTTCGACG
4741 AGGATGACGG GGACGTTTCA CGAGCCCAGC CAGCGCAGCA GGTCTTGCC GGCCTCGCCG
4801 GTGTTGGTGG GGGCGACGAG CAGCGTGCGC ACACCGCGGT CGAGGAGGCC GGTGACCTGG
4861 CCGCGGTCTT CCGCCGCGTC GTAGGCGGAA GCGCGCAGGA CCAGCCGTCC CCCGGCGGCG
4921 GCCACCGCCG ACTGGGCGCC CTGGATGACC GCCGGCCAGT AGTACTCGAC GGTGGGGACG
4981 ACCATGCCGA CGAACGCCTG GTCGGGCAGG CCGCCGTGGG TGAACACGCC CCGGGTGGCC
5041 CGCTCGTCGG CCGTCGGCCC GCGATAGGGG AGGGTGACGC CTCCGTGGAC GCGCGACACC
5101 AGCCCCCGGT CCGCCAGGAT CGTGACGTCC CCGCGGACCG TCACGCGCGT GACGCCCAGC
5161 CGCTCGACCA GATCGGCCAG GCGCACGGTG CCGCGTTCCC GCACGGTGGC GAGGATGAAC
5221 TCCTGCCGTT CCGATGCGAG TGCCATCCCG GCCTCTCCTT TTCCTTCTGA TCGAACTGTG
5281 TCGTTCGACT TCTCTTCGAC TTCTCGTGGA TGTCGTCTGT TCTATCGCAG GCGCCCCCGA
5341 TCCAGGAAGG CGCTACCCCG TTTGCCGGGA ATGCGAGCTA GCGGGAGGGC TCTCGGATCG
5401 GGAGCGGGGC CTCTCCGGAG CTCGTTGAGC TTTGCTTTCC TATCGATCCT GATCGATTCA
5461 TTTTCGCTACC GTTGACCGAA ATCTCGGCAG GGGCTGAAGT GTGCAACAC CTCTCTGCAC
5521 GAAAGGAAAC CCCTGGTGCT CCGCTTCTCC TCCGACGCCG GTGACAGCGC ACACCCCCCT
5581 CGGACGCACC GATCGCGGAG GTCCCTGCTG CCGACCTCCG CCGGCCTGGG GGCAGCGCTT
5641 CTCGCCGGCG GCACACTCAC CCGCTGTTCG GGCCGGACCT CCGGAACCGT CCGGGTCCGT
5701 ATGTGGTCGT GGCTGACCGG CATGGACCGG TACGTCGCCG CGTTCAACTC CTCCCAGCGC
5761 GACGTCCGGG TCGACCTGAG CGTGATCGCC GCCGGTACGT CCGGGGGCTA CGCCCAGCAG
5821 ACCAACCGGA TCCGCGCCCA CAACGCCCCG GACATCCTGC ACGTCGAGTA CCAGGCCCTG
5881 CCGCAGATCG CGACCACCGG CGGGCTGCGC GATCTGAGCG CGGACGTGGA CGATCTCGCC
5941 GACGGCTTCC TGCCCGCCGC GTGGCAGAGC GTCCGCCCCG GCGGGAAGAC CTGGGCGGTG
6001 CCCATGGACT TCTGCCCAGT GCGGTTCTTC TACCGCAAGG ACCTGTTCTGA CCGGGCCGGC
6061 ATCGGAGTGC CCCGTACCTG GGAGGAGTTC CGCCGGGCCG CCGCGGCCGT CCGCACGGCC
6121 GACCGGGATG CGCGGATCAC CACCTTCCC GTCACGACG GCGCCTTCTT CGCCGGGATG
6181 GCCTGCCAGG CCGGCGACCC CTGGTGGGAC ATCGCCGGGA ACGCGTGGCG GGTGCGGATC
6241 GACGGCGCGG GCACGATCGC CGTGCCGAG TACTGGCAGG ACATGGTGTC CTCCGGTCTT
6301 GCGGCGCGCG ACGCCACCGG CACCCAGAGC TGGATCTCCG CGATGCACCA CCGCCGGCTC
6361 TGGGGGATGC TCGGGGCCAC CTGGGGCGTG GGCATGCTGA AGAAGTCCCT GCCCGCCAC
6421 CARGGCCGCT GGGCGGTGCG ACCCTGCCC ACCTGGGGCG ACGACCCGTC CACCGGGGTC
6481 TGGGGAGGGT CGGCGTTCCG GGTCTCCGCC GAGAGCGAGG TGCCCGAGGC CGCCCTGACG
6541 TTCCTGCGCT GGCTCAGCAC CGACCCCGCG GTTCCCCGGA TCGGCTCGGM GTTCACTTTC
6601 CCCTCGCCCC CTTTCTTCGC CAACCGGAAG GTGGCCCGTA CCGCCTACGA GGACGGGTAC
6661 TTCACGGCG ACCCCGTCTT CGACGTGCTG GACGAGTCCG CGACCCGGGT CCCCGGGTGG
6721 ACGTGGGGCC CGACCTCGCT GAGCACCTTC GCCACGATCG CCGAGGCGCT CGGCCCGGTC
6781 TCCACCGGAG CCACCACCAT CCCCCGGGCG CTGCACCGGG TTCACACCTC GACCGTCGCC
6841 GCGATGCGGG CCCGCGGCCT GGCCGTCACC GAAGGGAGCC GGACATGACA GCCGTCCCCA
6901 CACCGGCGAC GACGCCCCCT GGCCCGGCCT CCGCGCGGTC ACGCCGGGCA GCCCGCCGGC
6961 GCGGCTTCAC CCGGGCCGCA CTGGCGGCAC CGTTCGCACT GCTGCTGCTC TGTACCGTGC
7021 TGATCCCGAT CCGGTACGCC TTCTACCTCA GCCTCTTCGC CGACCGCTC TCCGGGCTCG
7081 GCTTCGACGG GTACCGCCAG GTCTTCACCG GCTTCGACAA CTACACCGAC GTCTTCGCGC
7141 ACCCGGCCGC CCGCGACAGC GTGGCCAACG TCGCCCGCTA CCGCCCGGATC CACATCCCCG
7201 TCATGTTTCG CGGCGCCCTG CTCATCGCGC TGCTGCTCAA CTCCGCCACC GTACGGCTGC

7261 GCCAACTCTG GTCCCTGGCG GTCTTCCTGC CGTACGCGGT GCCCGGTGTG ATCGCCGGTC
7321 TGATCTGGGG CTACCTCTAC AGCCCCGGCA TCGGCCCCGT CGCCGACGTC CTGCCGTTCCG
7381 ACCCGTTTCGG CGAGTCCGGA GTGCTGCCCT CGATCGTCAA CATGGCCACC TGGCAGTGGA
7441 CGGGCTACAA CGTGATCATC TTCTACACGG CCCTGCAGAC CGTGCCCCGC GACGTGCTGG
7501 AGGCCGCGCG GGCCGACGGA GCCGGCGCCC TGGCGACCGC GCTGTCCATC AAGATCCCCA
7561 TGATCCGCCC GACGATCTTC GTCGCCCTCG TCTTCACCAC CATCGGCTCC CTCCAGCTGT
7621 TCACCGAACC CATGGTCCTG CGCACCTTCA CCGGCGCGGT CACCAGCACC TGGACCCCCA
7681 GCCTCTACAT CTACGACGCG GCCTTCATCA GGAACGACTA CGGGCGCGCC GCCGCCGCGT
7741 CCATTCTGCT CGCCCTCGCC TCGGCCCTGC TCTCGGCGCT GGTGATGCGG CTGTCTCTCC
7801 GGAGGTCACG ATGAGTGCAG CCGTCACCAC CTCACCGCCC CTGGCGACGC GCGGAACATC
7861 CGGCCGCGCG CGCGCCGCCC CCGACCTACC GCCGTACGGT CCCC GCGGCT CCCGCTGGAC
7921 CTCACGCGCC GCGGTCCACG GCGTGCTCCT GCTCGCCGCC CTCTACAGCG TCCTCCCGCT
7981 GGTGTGGCTG GTCACCTCGG CCACCAAATC CGTCGGCGAC TTCTCCGCCA CCAGCGCCTT
8041 CGAACTCGCC GACTTCAACC TCTGGGGCAA TCTGCGCGAC CTGTTCCGCC GGGAGGACGG
8101 CGTCTTCCTC CACTGGATAC GCAACTCGCT GCTGTACGCG GGCGTCGGAG CCGTTCCTGGG
8161 CGCGCTGATC TGCACCGCCT GCGGATACGC CGTCGCCAAA CTCGACTTCC CCGGACGGCG
8221 CGCCCTGTTC GCCGTGACCC TGGCCGGGGT CCTCGTACCG CCCACGCGC TCGCTCTGCC
8281 CCTGTACCTG CTCGCCTCGA ACCTGCACCT CGTGAACACC TTCTGGGCGG TGTTCCTCCC
8341 GCTGCTCAGC AACCCCTTCG GCGTCTACCT CGCAGCACC TTGCGGACG CAGCCGTCCC
8401 CGACGAGGTC CTGGAGGCGG CACGACTCGA CGGGGCGGGT GAGTTCCGCA TCTTCTTGAC
8461 CGTCGCCCTG CCGATGATGA TGCCCGGCTT CGTCACCATC GTGCTGTTCC AGTTCGTCAG
8521 CATCTGGAAC AACTTCTTCC TGCCGTGGT GATGCTCACC GACCCGAAC TGTTCCTCAT
8581 GACCCTCGGC ATCTACCAGT GGAACACCCG GGCCCCGAG TTCCCCGACT ACAACCCCT
8641 GGTCACTACG GGATCGCTGC TCGCCGTCGT ACCGCTGATC ATCGCCTTCG TCGTCTGCA
8701 GCGCCAGTGG CGCAGCGGCC TGGCCGAGG GAGCGTCAAG TGACCGCCAC CGCGACCCG
8761 TCCACCGGAT CCGGGGCGGC CACGGTCCC CGCAGGCCG AGACCCTGCT GCTGATGAAC
8821 GACCGGACCC GGCACACCCA GTTCGGGCC GAGGAGTTGG ACCGGCTGCA CGCCGTCGCC
8881 TCGCTCGGAC AGCCGCTCGG CGCCGCGGAC CTGGACTCCG GCGCCGTCCG CGCCCGCCTC
8941 GCCGAGGTGG AGGTGCTGAT CACCTCCTGG GGCTGCCCCA CCCTCGACCA CCGCGTCTTG
9001 CGAGCCGCCC CGCGCCTGCG TGCCGTGCTG CACGCGGCGG GATCGGTACG GGGCCATGTC
9061 CCCGACGCCG TGTTCGACCG GGGTCTGCTC GTCACCACGG CCGCCGACAG CAACGCCGAA
9121 CCCGTCGCCC AGTACACGCT CGGCGCGATC CTCTGGGCGT TCAAGAAGGT GCCGTTCTC
9181 GCCATGGACG CCCGGCGCTC CCGCGAGGAC TGGAGTACC TGCAACGAG GGGAGGCTC
9241 AGCGGCCGTG ACCGCACCGT CGTCTCTCGT GGCTTCTCCC GCATCGGCCG CCGCGTCGTC
9301 GACCTCCTGC GCACCCTGGA CCTGGCCCCG GTCTCTGTCG TCGACCCGGT GCGGACCCC
9361 GCCGAGGTG TGGCCGCGGG AGCCCGGCCG GCCAGCTCG CGGAGGCGCT CCCGAGGCC
9421 GACGTACTCA GCCTGCACGC ACCGGCACTG CCCGGGACCC GGCACATGAT CGGAGCGGCA
9481 CAGCTGGCCG TGCTGCCCCC CGGCGCCGTC CTCGTCAACA CCGCCGCGG CTCCCTCGTC
9541 GACACGCGCG CCCTGGAGGC GGCCTGCCGC GCCGACTGC ACGCCGTCTT GGACGTACC
9601 GACCCCGAAC CCCTGCCCCG CGACAGCCCG TTGTACGACC TCCACAACGT CGTCTCACT
9661 CCGCACATCG CCGGTCCCT CGGCTCCGAG ACCCGCACGA TGGCCGGCGC CGCCCTCACC
9721 GAACTGGAGC GCTACGCCGC CGGGCTGCCC CCACTGGCCC CGGTACCCCG GCAGTCCCTG
9781 ACCACGAGG CCTGACCTGC TCCGGTCCG ACCGGACCCT CCACCCACCC CAGAGGAGCG
9841 TCATGAACCA TCAGATCCC GGCCGAAGGA CCGTACTGCG GGCGGGCGGC CTGGCAGGCC
9901 TCGCGGCCGC CGCCGGCCTC GCCGCGAACC CGGCGGCCG AGCCCCCGC GCCGCCCGG
9961 TCCCGCGCAT CGACGATCTG GGCCCCGCTG TGAGCCAGTT CTCCCTCATG AGCGCCGTAC
10021 TGACCGGCGA CACCGTCCAC ATCGGCTCCC GCAACATCGA ACCCGTCCAG ATCATCTCCG
10081 TACACCTGCC GACCCGAAAG GTGGTGGCCC GCACCGAGCT GTCCAACGGC CACTCCATCC
10141 AGGCGCTCGC CGCGACGAC AGCGGCAGGT ACCTGTACGC CGGAATCCTC CAGAAGGCCT
10201 CCGGCGACCG GCCCAACCTC TTCCGCTGGG ACCTGAGCAC CCCCAGCAG CCGGCCGAAC
10261 CCCTCGGCGC CATCGGCGAC CGCGACGTCC GCGCCCTCGC CGTGGCACCC GACGGTGTGCG
10321 TCTACGCGGT CGGCGGCGGC AGCAGCGGCG CCCC GGCCCT GTGGGAGTAC GACCCCGAGA
10381 CCGGAGGAT CAGCAACGTG GGAACCCCGG ATCCCAGGGC CACCCTCGCA CCGGCCGTGCG
10441 CGGCGACCCG CACCCACGCC TTCTTCGGCG CGGGCACCAC CTTCAACGGC GCGGGGACA
10501 CCAGCAGGGC CGCCCTCTAC GCCTACGACC GGTCCGCCG GACCTTCACC TCGATCGCGC
10561 CCCC GGAGAT GCTGCGCGAC CCGAGCATCC GCGACCTCGC GGTCTCGGG GACCGGCTCG
10621 CCGTCAGCAC CGCCGGTTCC ACGGAGGACG CGAAGGTCGC CCTCATGGAC CTGAAGGACC
10681 TGTCTCTGTA CAAGATCGCG ACCTCGATCG GCAAGGTCGC CAAGTCGTT ACCATGACCG
10741 GCGACACCGT CCACTACGCC AACGAATCCG GTCTGCTCG CTACTGCGG TCCGGAACA
10801 CCGTGACCCC CGTCGCGTAC GACGGCCCCG GACTCGGCGA GATCTGGGG GTGGACAGCC
10861 GCGCGGGCAA GGCCGTGCTC ACCTCGGCCT TCGGCTTCGT GGCCGAGATC GACCCGGCCA

10921	GGAGGACCGC	CGCCGTCATC	GATCTCGGCA	CCGCCGGCGC	CCCCATGGCC	CCGCAGACGG
10981	TCATGGGCAT	CGCCGCCCGC	GGCGGCCACG	TCTACGTCGG	CGGAAACCGC	ACGATCGCCC
11041	GCCACACCCC	CGGCCGCGAC	ACGGTCGTCA	AGTCCGCGC	ACCCGCGCAG	GCCAAAGGACG
11101	CCGTCTCGT	CAACGGCATC	CTGTACACCG	GCCAGTACAA	CTCCGAGGGC	ATCTGGTGCC
11161	ACGACCCCCG	CACCGGAAAG	ACCGGCCAGG	TGGCCGCCTT	CCCGTCCGAG	CAGAACCGCC
11221	CGCTGGACGT	CACCTTCGAC	GAGGACACCG	GTCTCGTGCT	CGTCGGAGTG	CAGTCCGACA
11281	CCGAGGGCGG	CGGCGCCCTC	TGGACGTATC	ACCCCAGGAC	CGGCCGGCAA	CGGAGCTTCG
11341	CCGACCCGGT	CGACGAAGCG	CAGTGCCTCC	GGGGCATCGC	GAGCCGGAAC	GGCGTCGCCT
11401	TCCTCGGCGG	CGACAACACC	GCCCCACCG	GACCGCGCAG	CACCGTCGTC	GCGTTCGACC
11461	CGGTGGCCGG	CCGGGAGCTG	TGGCGGATCG	ACCCGCAACA	GACGGCGGGC	ACCGCCGCCC
11521	TCGCGGTACG	CGGACGGCAC	CTCTACGGAC	TGTCCCGCAA	GGGCGGCCTC	TTCGTCATCG
11581	ACATCCCGCG	GCGCGAAGTG	ATCCACCGCG	CGGACGTCAC	CTCGGTGAGC	AACGGGTTCG
11641	CCGCCCTGGT	GACCAACCGC	GGTGTCTGCT	ACGGCGTCTC	CGACACCACC	GTCTTCCGCT
11701	TCCACCCCAG	GACCTTCGCC	GTCTCCGTCG	TCGTCCCCGA	CATCGACGGC	GGCTGGTACA
11761	GCGGCGCGCA	CATCACCAC	GACGAGCAG	GCCGGCTCTA	CACCATGCGG	GGCCGCAACC
11821	TGGTCCGGAT	CACCGACCGC	CCCGGCCACT	GACGTCGGGG	TGTCTCCCCC	CACCCCGGGC
11881	ACCCCGCGCC	ATCCCCCAC	GCCCTCCAG	GGAGCTTCCA	TGAAGAACGC	GCACAAGAGG
11941	ATCTGCATCG	CCGTCTGCCT	GGGCGCCGTC	GCCGGCGCCG	GTTACGCCAC	GACCGCCACC
12001	GCGACGCCGG	AGAAGTCGGC	CCGGCCGACG	CAGGCGTCCG	ACCTGCTCGA	CGTCGACTAC
12061	GAGTCCGGGG	CCCTCGACTC	CGGCATCCCC	GGGCTGACCA	CGACGCACGC	CAAGGCCTCG
12121	GACGCCTCCA	CGGTCGAGAG	CGGCGGCGAC	GGGAGCGCAC	ACTCGGTCTC	GCACAAGGTG
12181	ACGCTGAACG	ACCCGGATTA	CGTCTCCGAC	GGCGCACCCC	GCAGCGAGAG	CGCGAACAAC
12241	GAGATCGAGA	AGTCGCTCAT	CCACGTGGGC	GACGTGCACC	GCTACGAGTT	GACGCTCATG
12301	CTCAAGGACT	GGGAGACCAG	CAGCAGCAAG	GCGGGCGACA	TCATCTTCCA	GGGCAAGCAC
12361	GCGGGCGGGA	ACAAGCCGTC	GTTCTACCTC	ATGGCACAGC	GCAACGCGAT	CGCGTTCCGC
12421	TCCCCGCTCC	TCGACCTGCA	GGCGCCGGTG	GTCGAGGACT	TCCGCCCTTA	TGTCGACCAG
12481	TGGATGCGGT	TCCGCATCGA	CGTCCGCTGG	GCCGAGAGCA	GGACCGGCTA	CTACAAGGTG
12541	TCCGTCCGGC	TCCCGGGAGA	GAAGGACTTC	ACCCTCAAGA	AGACGTACAC	CGACGTCAAC
12601	ACCTTCCACC	CGGAGAACCC	CACGGCCTTC	GGCTACATCA	AGTGGGGCCT	GTACCGCCCC
12661	GCCTCCTCGA	CGGAAGCGGG	CGATCCCGCG	ACCCGCGTGG	TCCGGCACGA	CGACATCCGG
12721	GTGATCGACC	TCGGGTGAGC	TCGGCGCGGG	GTGATCGCCT	GCCGCAGGCG	GTCACCCCGC
12781	GGTCACGTCG	ACCAGACCAC	CTTGCCCTGG	TGCACGATGA	CCACCTTGGC	GTCGGGCCTC
12841	AGCACCATCT	GGGCGCCGCC	AAAGTCATGC	GTCTTGGAAG	CCCAGATGGC	CCGGTCGTCG
12901	CCGTTGTGGA	TGACCAGGTT	GCCGTCGGGC	TGGAAGATCG	CCCGGTGGTT	CTCGCCGAAG
12961	GTCATGGCGG	CCCAGATCGG	CTTGTTCCGC	TCGTTGTAGA	CGACGAGGTT	GCCGTCTGTC
13021	TGCATGGCTA	TGCGGATGCG	GTTGGTGCGC	CAGGACTGGC	CGACCCCGAT	CGCACTGCTC
13081	GCGGAGACCG	TCTGGGTGCC	CCAGTCCGGC	TTCGGGGCCG	CCTTGGGCTT	CGGCTTCGGT
13141	GAGGCCTTCG	GCTTCTCGGC	GGGGGCGCTC	GTCCGGGGTG	CGGGCGAAGG	GGCCACGGGT
13201	GCGGGAGCGG	CTTTCGGCGG	GGTGGTCTTC	TTCTGTTCCG	TCGGGGACGG	CGTGGGCTTC
13261	ACCGCCACGT	AGTCGTCCAG	GGCCGCCGGG	GCCGACTCGG	GGTTGAGGAC	CGTGTGCGCG
13321	CTCCCGGCCA	CCGGTGCCGT	ACCGCTGTTA	CGGGGCTTCT	CGTCGTTCGC	GCTGCCCATC
13381	AGCAGCACGG	GTATCGCGAC	CAGTGCGGCG	CCGACGAACG	CCGCCCCCGC	CAGGACCGGT
13441	GCGCGGGGGC	GGCCCCGTGC	GGCAGCGGTG	AGGGTCTTGG	TGTCCGGCTT	CGCGGCGACG
13501	GCCGCTGCCA	GGGCGATGGC	CTCCGTGCGG	CCTTCGGGTG	CTTCCGTGGT	TGTGGTTGTG
13561	GCCGGGGCCT	GGGTCTGGGC	TTTTTTCGGTA	CCGGTCGTTG	CGGCGGGGCTC	GGTCGGGGGC
13621	TCCCCGCTGC	CGGCCGTCTG	GGCGGCCCTC	CTCCGGGGCT	TCCGCGCCTC	GGCCGGGGCG
13681	GCGGACGGCG	GGGCGGCGGG	GCCGGACTCC	TCCGTGGTGG	GCGGCAGGGA	AGCGGTGCCG
13741	GACTTCTCCG	TGGCTGCCGC	CGCGGACTCG	GTGCCGGGGG	CGGGGGGCTC	GGGGGGCATG
13801	CGCTGACTCC	TCTTCGGGGG	CGAGCCCAAC	GTAGCCCGTC	GGACGAGCGG	GAGAAGCCAC
13861	TCGCTTCGAT	ACGCCCCGTG	CGCTTTGTCA	TCCCTCGCCG	AGCCTGTCCC	GGAAAGTTCC
13921	TCACTCGGGC	AGAGAACGGG	CAGGTAACG	GCCTATACGG	GCCTGCGGCC	GGACCCCGTG
13981	AACGAGGCTC	GTATGACATC	CTGTTGGGGC	CGTTCGGCAC	TCACATATTC	GTACGGAGCC
14041	AAGAGTTTCA	GTGGATGCCT	CCCTCGTGCC	GGAAGCCGTG	AAGGACGCCG	CACGCCGTGC
14101	GCCCCGCCAC	TGGGTGCGCA	TCGTCGATCC	GGAGTGGACG	CAGGCGCGGA	CGCCCCGGA
14161	GTGGGCCGTG	CTGGGGGAGT	GGCAGTCGGA	CGAGAGCGGG	AGCGTGGGCG	AGTACCGCGC
14221	CAACCCCGGC	TACCGTCCCT	CCGCCCGGGT	CCTCGGCTGG	CCGGAGCCCA	CCGATCCGGT
14281	GGACGCGGCC	GCCCAGCGGG	CCGCCACGGG	GTACGGCACG	GTGGACGAGG	CGCTGGCCGC
14341	ACTCGCGGAG	GCGGACGTCA	CCGTCTGACG	CGGGCCCGAC	GGCGGGCCAC	TCTTGGCCGC
14401	GGGGCGGGAC	GGGGCGCTGG	TGGCGCTCCT	GTTACCTTCG	CCCACCCACG	CGTTCATGTC
14461	CGCGGCACTT	CACCATGACA	CCGTGTCCGC	CCGGGAGCTG	GCCGGGTGCG	TGGCGCAGAC
14521	CGGCGTCTTG	CTGTCCGTCA	ACGCCGTGTC	CGCGGCTCCG	CTGCTCGTTC	CCACCGACAG

14581	CCTTGCCGGT	GCCGTGAGTT	CCGGCCCCGA	GACCGCGGCC	GACGGCCCCG	CGCCGACCGA
14641	ACCCTGGCCC	CACACCACAG	GGAGAACCCC	GTGACGCGTT	CGTCGTCCCA	GCAGCCCCCC
14701	GCCTCCGCTT	CCGCTCCCCG	ACAGGCGAAG	ACCGAAGCCG	CGCCGGCCGA	GGCCACCGAG
14761	GCGGAGAAGG	TCAAGGAGGT	CGAGGAGGGG	GAGAAGGCCG	AGCAGGCCGA	GCGCTCCGAG
14821	GCCCAGTCCT	CCGAGCGCCA	GTCCCCCCAG	GTCCAGGCCC	AAGCCCCCGG	CGCCGGGGCG
14881	GAAAACCCCA	AGGCGCCAC	CGCCTCCGAG	ACCGACCCGG	ACCCCGGCGC	CTCCGCCGCC
14941	GCCCAGAGTC	GGCTCCCCGC	CCTCGTACGG	ACCATGACCG	CCACCGCCAT	CGGGAAGCCG
15001	CAGCAGGAGG	CGGGGCCCCG	GGGGCGGCCG	GGCAAGGCCG	CGCTCGCGGG	GGCGGCGGTC
15061	GTGGGCGCGC	TGCTCGTGTC	GGTGCCCTTC	CTCGTGCTCG	CCGGGAACAA	CGACGACGGG
15121	CCCGAGCGGA	CGAAGGCCGC	CGGGGCGGGG	ACCGTCCTCG	ACGGGAGCGG	ACCGGAGGCG
15181	CCGGGCGAGT	TCGCCGTCAC	CGCGCCCCGAG	GCCGGCGCGC	CCGCCGAGGA	CAGGGAGAAG
15241	AGGGAGAAGG	ACGAGGCGAA	GCCGGAGAAG	GGCGGCAATC	CCGTTCCGGC	CATCCCTCCG
15301	GGGGGCGGCG	GCAAAGCGGA	GACGAAGGAG	GACAACGCCC	CCAACGACCC	GCCGAAGAAAG
15361	GCCGATTTCGT	CCAAGGGGAA	GTCCAAGGAG	CAGCCCCGAGG	CGGGCAGCGG	CGGCGGCAGG
15421	AACCAGCCGG	CCAAGGCCCA	GCCCGCCGTC	ACGTTTCAGCG	GGCCCGTCTC	CTTCCGCAGC
15481	CACCTCTCCG	GCCGCTGCCT	CGACGTCCCC	GGGCACAAC	TCAACGACGG	CCAGCCGCTG
15541	TTCATGTGGG	ACTGCAACGG	CGCCGACGCC	CAGAAAGTGGC	GCTTCGGCTC	GGACGGCAGC
15601	ATCCGGGCCA	GGGACAAGTG	CCTGGACGTG	GCCAACGCCG	ACTTCGCCAA	CGGCTCCCGG
15661	ATCCAGTCTG	CCTGGTGCAA	CGGCTCCGCC	CGCCAGAAGT	TCACCCTGAA	CGGCGCCAC
15721	GACCTAGTCA	ACACCGTTCG	CGGCAAGTGC	GTCGACATCC	CGAACCACAG	CAAGGGCCCG
15781	GGCCCGGAGA	CCTACCTGAT	CCTGTGGGAG	TGCACGGGGC	TCGACAACCA	GAAGTGGAGC
15841	ACCTGACCCA	GGCGCGGGCA	GCGAGACATG	ATGAGATCGC	CGGAGCGCCG	AACCTGTGCC
15901	CCGGCGAAGA	GGGGCGTCCG	GTCAGGACGC	CGCCGTAGCA	AGGAGGTTGG	TGCGTGTCCG
15961	GCCAGGTACT	GACGGTCGGC	CCCGCGGACC	GCTTTTCCAC	GATCGGTGAG	GCGCTCGCCG
16021	CCGCCCCTAC	CGGTGCGCTC	ATCAGCGTCC	GGCCCGGGAC	GTACGCGGAG	AACCTGGTGA
16081	TCCACACCCG	GGTGACCCCT	ACCGCAGCCG	AGGGGCGGGG	CACGGTGGAG	ATCCGGCCCG
16141	GTTCGGGCAG	CGTCGTCCGC	CTGCGCGCCG	ACGCCGTTCAT	GCTCTCCGAA	CTCACCTGTC
16201	GCGGCGGGGA	CTCCGAGCTG	CCCGCCGTCC	ACGTACGGCG	CGGGCAGGCC	GCGTTCGACG
16261	GGTGCGAGAT	CGTCGGCGCG	GCCTGGACCG	CGATGCTGGC	CGGCGGGACC	GGGTCCCTCG
16321	CGCTGCGGGA	CTGCCGGGTG	AGCAACCCGC	GGGGCGCGGG	CATCGTGGTC	ACCTCCGCCA
16381	CCCCACCAC	CGTCGAGTCC	TGCACGCTCG	AACACCTCGG	CACCAGCCGG	ATCGTCCTCG
16441	CCGAGCAGGG	CGAGGCGCGC	GTCCGCGACT	GCACGGTCCG	GGGCGCCCGG	GGCAACGGGC
16501	TGCTGGCCAA	CGGCGAGACC	CGGGGCACCG	TCGAGGACTG	CGACATCTCC	TCCACCGACA
16561	AGCCCTCCAT	CGCCCTGGAG	GGGGACTCCG	CCGTCTCCGT	CGTCCGGACC	GTGGTGCATG
16621	ACACCAGCAC	CGGCGTGATC	CTGAGCAGCG	CGGGCCGCAC	CACCCTGGAG	GACGTCCGCG
16681	TCACCGGCGC	CTCCGGCAAC	GGGATCGTCC	TCGCCGTGGG	CACCGACCCC	GTACTGCGCC
16741	GCTGCCGTGT	CTCGCGCGCA	CGGGGGCACG	GGCTGTTTCG	CACCGACCGG	GCGCGCGGCA
16801	CCTTCGAGGA	CTGCTGGGTG	GACGGGGCAC	AGGGCGCCGC	GCTGCGGGTC	GCCGGGGCCT
16861	CCTCCCCGGC	GCTGACCGGA	CTGACCGTCC	GCGACTGCGA	AGGGGCCGGG	CTGCTTCTGG
16921	AGGAGGACGC	GGCGCCCCGAG	CTGGACCGCC	TGGAGGTCAT	CGGCTCCTCG	CCCGCCGTCG
16981	CTCTGCAGGG	CGGCGCCAAC	CCGCTGCTGC	GCCGGGCCCCG	GCTGGTTCGAA	CCCGCCGGTG
17041	ACGGCATCGC	GGCCACCAAG	GACGCCCCGG	GCCGCGCAGA	GGACTGCGAG	ATCGTCCGCC
17101	CGAAGGGCGC	CGGTGTACGG	GTGGCCTCCG	GCTCCACCC	CTACCTGGCC	GGTGGCGGGG
17161	TCTCCGACAC	GGCCACCAAC	GGCCTGGTCG	TGGAGGACGG	CGGCAATGTC	ACCGTCCGTG
17221	ACTTCCGCGT	CGAAATATCG	GGTGAGGAGG	GTGTGGTTCG	CGCCGCCGGT	GGTGAACCTGA
17281	CTGCCAACCG	CACCACCGTC	CACGCCCCCA	AGGGACACGG	TTTCCTCCTC	CGCGAGGGCG
17341	CGTCTCGCCT	GCTCAGCGGC	TGCGAGGCCA	ACGGGGGCGC	CCAGGACGGC	TTCCGGGTTCG
17401	AGTCCACCGC	CCCGGTCTCG	CTCGTCAACT	GCACCGCCCCG	CGAGAACGAG	GGCGGCGGCC
17461	TGGTGCAGAC	CGCGCCCGGC	GAACGGGCTCG	CCGTGGACGG	CCTGAACAGC	GTCGGGAACG
17521	GCAAGCGCGA	CGCTGGGGG	AGCGGCACTG	CCGAGAACAC	CGACCCGGCC	GGGTCCGGCG
17581	CCGCCGACGC	GCCCCCGCCG	GACCGTGCGG	ACGGCCCCGCT	CGGCGCCCTC	AACGCGCTGA
17641	TCGGCCTGGA	GAACGTCAAG	CAGCAGGTCC	GCACCTCTCG	CAACCTCACC	CAGCTCGCCC
17701	AGCGACGCGA	GCAACTCGGC	ATGCCCGCAC	CGCCGATGAG	CAGACACCTG	ATCTTCGCCG
17761	GACCGCCCCG	CACCGGTAAG	ACCACCGTCG	CCCCCTCTA	CGGGGCGATC	CTGGCCGAGC
17821	TGGGCTCGCT	GCGCAGTGGA	CACCTGGTCG	AGGTCTCCCG	CGCGGACCTG	GTCGCCCAAG
17881	TCGTGCGCGG	CACTGCCATC	AAGACCTCCG	AGACCTTCCA	ACGAGCGCTG	GGAGGGGTCC
17941	TGTTTCATCGA	CGAGGCGTAC	ACCTCACC	CCGACAGCGG	CAACGGCGGA	GCCGACTTCG
18001	GCCGCGAGGC	CGTGGACACC	CTGCTCAAGC	TCATGGAGGA	CCACCGCGAC	GACGTGGTCG
18061	TGGTCGCGCG	CGGATACTCC	CGCGAGATGG	AATCCTTCCT	CAGCTCCAAC	CCGGGCTCTG
18121	CGTCCCGCTT	CTCGCGGACC	GTCGAGTTCG	AGAACTACTC	CGTGGACGAC	CTGGTCCGGA
18181	TCATGGAGAG	CATGTGCACC	CAGCACCAGT	ACGAGCTGGG	GGAGGGCACG	GCCCAGGCGC

18241	TCGCCGCCCA	CTTCGGGGCC	ATGGACCGGG	ACGCCGGATT	CGGCAACGGC	CGTGCCGCAC
18301	GCGGGGTGTT	CGAGGAGATG	GTGGACCGGC	AGGCGATCCG	GCTCTCCGCC	CAGGCCCAGG
18361	TCAGCGAGAA	CGATCTGCGG	CTCCTGCTCC	CCGAGGACGT	CTCCGCCACC	GCCGCCGCTT
18421	CGGCCGCCGA	GACCGCCGTA	CCGGCCGACG	ACCCGCTCAC	CCGCCTGGGC	GACATGATCG
18481	GACTGGCCCG	GGTGAAGCGT	GAGGTCGCCG	ATCTGGTCAA	CCTCATCACC	ACCGCCCCGC
18541	ACCGCGCCGC	CGCCGGGCTG	CCCGTCCCCA	CCCTCAGCAA	CCACCTGGTC	TTCACCGGCC
18601	CGCCCGGCAC	CGGCAAGACC	ACCGTCGCCC	GGCTCTACGG	CGAGGTGCTC	ACCCAGCTCG
18661	GCGTGCTGGA	GCGCGGCCAG	CTGATCGAGG	CCGCGCGCGC	CGACCTCGTC	GGCCGCTACA
18721	TAGGCCACAC	CGCCAGTTG	ACCCGTGAGG	TCTTCGAAAA	GGCCCGCGGC	GGCGTGCTGT
18781	TCATCGACGA	GGCCTACACC	CTCACCCCGC	GCGGCGGCGG	CGCCGACTTC	GGGCAGGAGG
18841	CGGTGGACAC	CCTGCTGAAG	CTGATGGAGG	ACCACCGCGA	CGAGGTGCTC	GTCATCGTCG
18901	CCGGATACAC	CGACGAGATG	GAACGCTTCC	TCGCTTCCAA	CCCCGGCCTC	TCCTCCCGCT
18961	TCCCACGCCG	GATCGCCTTC	GCCGACTACT	CCTCCGAGGA	ACTGGTCACC	ATCGTGCGCG
19021	CCCAGGCCGC	CGCCATGGGG	TACGAGTGCG	GCCCCGGCAC	CGGACCGCTG	CTCAAGGAGC
19081	ACTTCGACGC	GGTACCCCGC	GACCGGTCTT	TCGGCAACGC	CCGGCTGGCC	CGCCAGGTCT
19141	TCGAGTCGAT	GGTCACCCGC	CAGGCGGGCA	GGATCAGTTC	GCTGGCCGCG	CCCACCCCTG
19201	ACGACCTGCG	CATCCTCCTC	CCGGCGGACG	TCACGGCCGC	GGCCCCGAAG	GCGGTGCAGC
19261	AGTGAGGCCG	CCCGCGCGAC	GGGCGACCGG	GAGGCCCCGA	GCAGGCCGGC	TGTTGTACGG
19321	GGTCTCGCTC	GCCGCCGCCC	TGGTGGTGCC	GGTGGCCGTA	CCCGCCAGG	CGGCACCTGT
19381	CACCCTTACA	CCGGCCGCTG	ACGGGGCCAA	GGGACAGGAG	CTGGCCGGA	TGCCACCCGC
19441	GCTCGACGCG	CGGGCCGAGG	CCGTGCTCTG	CACCCCGGCC	TCCAAGGAGC	AGGCGAAGAA
19501	GCAGGACTGG	TCGCGCCAGC	GCCTCGACCT	GGACCGGCTG	CACCGGCACA	CCACCGGGTC
19561	GGGGGTGACC	GTGCGGCTGA	TCTCCACCGG	CGTCGACCCC	GGGGCCGAGG	GACTCGACGG
19621	CCGCGTGACC	GCCCAGGGCC	AGGCCGCCGA	CGACTGCGTC	GGACAGGGGA	CGTTCTTGCG
19681	CGGGCTGATC	GCCGGGACCG	GCGGCCCCAC	CCCGCGCCTC	GCCGGGGTCG	CCCCGGGCGC
19741	GAAGATCCTC	GCCCTGCGCG	GCACCGACCG	GCGCGGGCAG	CCGGACCCCG	CGCTCGTCAC
19801	GGCGGCGGTA	CGGGCCGCCA	CCACGGCCGA	GGCCGACGTC	ATCGCGGTCT	CGGTGGCCCT
19861	TCCCCGTAAG	GACACCGCGC	TCACCCGGGC	CGTCGCCGAG	GCCCCGAAGG	CGGGGGCGGT
19921	GGTGGTCGCG	TCGGCCACCC	CGGAGCCGCC	GTGCGCGCGC	TCGGCCGACG	AGATCCCCCT
19981	GCGTACCTAC	TGGCCCCCGG	GTGAGCCCGG	CGTCCTCGCG	GTGCGCGACA	TGCTGCCCGC
20041	CGGGGCCCGG	CCGGACGCCG	CCCTCCCCAC	CGAGGGCATC	GATCTGGCCG	CCCCGGGCGC
20101	CGGAGTGGTC	TCCGGCGGGC	CGCGCGGCAA	GGGGCACTAC	CTCGGCGGGC	GGGTGCGGGT
20161	GGCCACCGCG	TACGCGGCCG	GGGCCGTAGC	CGCTGTCCGC	GCCACCCACC	CCGAGGCGAC
20221	GGCCGACGCG	GTGGCCCGTC	GGCTGACCGC	CACCGCGTAC	CCGGCCGACA	TCCCCAGCT
20281	GGACGTCTAC	GCCGCCGTCA	CCACGGTCTT	CGGCGACGCG	GGCGCGCCGG	TCCGGCGTGC
20341	GGAGCGGGCC	GCCGAACCCG	TGGCCGTACG	CGACACCTCC	GACACCGACC	GCGCCACCGG
20401	CCGGGCCGTC	CTCTTCGTCC	TCCTCGGCTC	GGCGGGCGTC	CTCTCCCTCC	TCTGGGCGCG
20461	TTTCGCCGTC	ACCCGGGCCC	GCGCCCGGGG	CTGGCGACCG	GCGGGGGCGG	GCACCCCGGA
20521	CCAGGGCTGA	ACGTGGAAGG	GGCCCCCTCC	GGCCGGAGGG	GGMMCTTCC	ACGTTGCTC
20581	GGTCTTATTC	CACCAGGGCC	GTCTGCACCA	GCCGCGGCTT	GCGGCGCTGG	ATGTGCAGGG
20641	CCCGGCCCGG	CGGCAGGTTC	AGCGGCTTCG	CGTTGCCGAA	CAGGCGGCCC	TCCGTGGGCG
20701	GGCAGGAGAG	CAGGACCGCC	GGGTTGTTGG	CCTCGTCCAG	GCGGCGGATC	AGGCCGTCTC
20761	TCAGACCGCG	CCCCGCTCCC	ATCGCGCTGC	GGGCGACGAC	CAGGTGCAGG	CCCATCTCGT
20821	AGCCGAGGGT	CAGATGCTCG	AAGAGCGGCT	CGAAGGGGCT	CTGGAACGAG	TTGCCCGACA
20881	CCATGTCGTA	GTCGTGACCC	AGGACGAACA	GGCGCGGGCC	CGTCCACCAG	TCGCACCGGC
20941	GCATCCGGGC	GGGTGAGATG	TCGGCCCCGG	GCACCCGGGT	CTTCATCGCG	CGGGCCGCGC
21001	CCTCGATGGT	CTCGTTGAGG	TTGTGAGAGG	AGATGACATG	CCCGATCGCG	TACTCTCTCG
21061	GGATGGCGTC	GACCAGGGTG	CGGCGGTAGT	CCACCGCGAT	GATCTTCGCC	TCCTCGGGGG
21121	CGTACCGGGC	GGTGATGCCC	TGCGTGATCA	GGCGCAGCAG	ATTGGTCTTG	CCGCTCTCGG
21181	TGTCGCCGAC	CGCGATCAGG	TGCGGCGTCC	GGCTGAAGTC	GTGCCAGACC	GGCTCCAGCG
21241	CGTCTGCTC	GAGACCGAGC	ACGAGGCGCA	TGCCGCGGCC	CTCCGTGGCC	TCGGGGGAGG
21301	GGAGTTCGGA	GACCGGCAGC	CGGTGCGGGA	GCATCCGCAC	CTGCGGGGCG	GCGGGACCGG
21361	ACCAGTGCCG	GGAGATCTCG	CCGACGAGGT	GGGCCACGCC	CTCGCCAGAG	TCCTCCAGGG
21421	AGCCGCTGCC	GTCCAGCCGG	GGCAGGCCCC	CGAGGAAGTG	CATCTTGCTG	TCGGCGGTGA
21481	TGCCCCGGCC	GCCGGTGCGC	GGCACCGAAC	GGGCCTTGCG	GGTGTGATC	TCGGAGTCCA
21541	TCGGATCACC	CATCCGAGT	TCGAGGCGGG	TGGCGGCTTG	GTCGCGGACC	TGCGCGGACA
21601	GCTCCACCCA	GCGGGTCTGT	GTGATGAGCA	GATGGATGCC	GTAATTGAGG	CCCCGGGCGG
21661	CGAGTTCTGT	GAACCTTCGGG	ATCAGGCCGT	CGTAGTCTTG	GCGGACCGTG	GACCAGCCGT
21721	CGACCACCAT	GAAGACATCG	CCGAACGGCT	CGTCGGGGAA	CTCCCGGGCG	GCGCGGCGCC
21781	GCCGGTAGGA	CTGCATGGAG	TCCAGCGTGT	GGTCTACGAA	GAACGTGCTCA	CGGCGGGCCA
21841	GCAGCGTCAT	CACCTCGGCC	ACCGTCCGGT	GCACCCGCTC	CGGGTTGAGC	CGCGCCGCGA

21901 CTCCGCCGAC GTGCGGGAGC GCGGCGAGCT GGGAGAGGCC GCCGCCGCCG AAGTCCAGGC
21961 AGTAGAACTG CACTTCGGCC GGGGTGTGGG TGAGCGCGAG GGCGGCGATG AGGGTCCGGG
22021 CGACCGTGGA CTTGCCGCTC TGCGAACCGC CCGCGACGGC GACATGGCCG CCCGCCCCGG
22081 ACAGGTCCAC GACCAGCGGA TCGCGCCGCT GCTCGAACGG CTTGTCCACC AGGCCGACCG
22141 GCACCCGCAG CTTCCCCGTG CCCGCCCAGC TTGCCGCCAT CAGCCCCCGC TCCTTGTCCG
22201 GCGTGATCCC GGGCAGCAGC GCGTCCAGCG GGGACGGCTC GTCCAGCGGC GGCAGCCACA
22261 CCTGGTGGGC GTCCGGCCCC GAGCCGCGCA GCCGGTCCAG CGCCACATCG AGCAGCGCCT
22321 CCTCGTCCTC GGCCTCCTCC GTCTCCGGCT CCGGGTCGGG CGGTGTCTCC AGGGTGC CGG
22381 GCACCACCCA TCCGCTGGTC CACGCGACCA CCTGGCTCGC CACCCGGGCC TGCACCACCG
22441 CACCGGTCCT GCGCCGGTAC GTCCCCGAGG AGTACGCGGC ACGGAACCGG GTCAGGGCCT
22501 CCACCCCGGA CTTCAGATAG CCGCTGCCGG GCGCGGCGGG CAACTCGTAG GCATCGGGAA
22561 CGCCGAGCAC GCCACGGCTC TCCATCGCGG AGAAGGTCCG CAGCCCGATG CGGTACGACA
22621 GGTGGCTCTC CAGCTGGTGC ATACGCCCCCT CGTCCAGGCG CTGTGAAGCG AGCAGCAGAT
22681 GGACGCCCAG GGACCGGCCG AGGCGGCCGA TCATCACGAA CAGCTCCATG AACTCCCGGT
22741 GGGTGGAGAG CAGTTCGCTG AACTCGTCGA CCACCACGAA GAGGCTGGGC AGCGGTGCCA
22801 GGTGCGCGCC GCGGCGCGCG GCCCGCTCGT ACTCCAGCGC CGAGGTGTAG TTGCCCGCCG
22861 CGCGCAACAG CTCCTGGCGG CGGATGAGTT CGCCGTGCAG GCGCTCCTGC ATGCGTTCCA
22921 CCAGGGCGAC TTCGTGCGCC AGGTTGGTGA TCACGGCGGA GGTGTGCGGG AGTTCCTCCA
22981 GGCCGAGGAA GGTGGCGCCG CCCTTGAAGT CGACCAGGAC GAAGTTGAGG GTCTCGGAGG
23041 AGTTGGTCAG GCGGAGACCG AGGACGAGGG TGCGCAGCAG CTCGCTCTTG CCGGAGCCGG
23101 TGGCCCCGAT CAGCATGCCG TGCGGGCCCA TGCCGCCCTG CGCGGACTCC TTGATGTCGA
23161 GCTCGACCGG GCGGCCGTCC ACCCGGACCG CGATCGGCAC CCGCAGCCGG CCGGAGCCGG
23221 AGTCCCGGGC GAACAGGGTC CCCGGGTCGT GCCGGTGCAG ATCGGGGATG CCGAGCAGGG
23281 TGGTCAGCTC GACATCGGTG TCCAGCGGCT GCGCGATATC CGTGCCGAGG CTCATCCGGC
23341 GCGGGGTGAG CAGCCGGGCC AGCGCCTCCG CGCCGAGCGG GCCGAGCCGG TCGGGGCGGC
23401 CGAGCGGCAC CGTGCGCTCC TTGCGGCTGC GGTGGTGC GACCAGGTT ACCTGGCCGG
23461 GCCCCACGGT GAGGCGCAGC GTGTTGCGGC CGGGCCGCCA GCGCAGTGCC CCGGACACGT
23521 CCAGGATCAG GCGGTTGCGG TAGCCGTGGC CCTCCAGCG GTGGCCCTCC GGGACGCTGA
23581 CGCCGTCCAG GACGACGACG GTGTACGGCT CGTCGCGGCC GGGCCGGGCG TCCGGGTCCA
23641 ACCCGGGGCG CTCGGCGAAC TCGGCGCCGA GCAGATCGTC CAGCTCGGTG AGGTCTGCGG
23701 TGATCCGGCG GGCCTGCCCC GCCCGTCTCT CCTCGTGC GG ATCCAGCACG TGCGGCAGCC
23761 ACTTGACCCA CTCCCAGTCG GCCCGCCGTT CGTCGCTGAC GCAGAACCG ATCCACAGCT
23821 CCTCGGGCGC GTGGAACACC GCCAACTGCC CGAGCATCGC CCGCACCAGG GCCCGTACGG
23881 CGTCGTGCTC CGGGGGATGA GCGCCGGGCG TGGCACCGTC CGGCGCCTCC TCCGGACGCA
23941 GCAGGATCC

SEQ ID NO: 107

Nucleotides 2941-74 of SEQ ID NO: 106

SEQ ID NO: 108

VRSSRSRSAGPGQSGTTFVQGQVAVLAAAGLLMTATPAAHADTATTSAASEVTPGAGTGTPSLVTGLNEEADA
TGTAADAARGHLKAKKSRYRIAHTSAKDLAAAGTTTGRGGRET VRLQQKYRGVTVLGGQYLVRMTKKDGKRTVTGT
SGSYFTELKLDVTPEVSEKTAIERAVGAVSAQLGGGLLRAPAKGEKPGGVQLSGTAGDVT VLPQSGSVLTRHITV
TGVSPADGTPVKQEVYIDAHSGFPMQYSGIKSFGAPGSGAGTSSGATGDAPTADATAEATQQLVVKSGSTRYNGE
EAEINLYQGADGAYQMYDYSKRGADSPYDGPLLSTWDARGREAATASGTWPSGIQTFRSDTPDLGKEFTDSGAVDA
HWAAGKVYDYYKNHFGRKSLDGKDFMYSLVGVVNGGKPFNNAFWDGTKMVYQGGGDYRTFSADTDVVGHEMTHG
VIEHTANLVYAGQSGAVNEAIADYFGNAIDLEADGRSMDDPDAGLLGEDLCTTLGPRDCALRDLNDGATTSKDFIG
VSHNGDNNGVHLNSTIFSGALWDMREDLGAEPADQIVYRALSAYMTPLDGFTEGRAAVIAAAQELGATKAQTKIVK
RSFDAHGIVPGWEKALGVDTDTLVSKVNVAGTGVGAGGGKYAVSRSDDEEGSEPYSVWLGN TDGKGKPKQLVSGNNGN
YNVYATTDGKNVAWAEFSPGSVLFHVRPVGGGPARVVGGVGANVSQVLVLEGDFLAYTVTNPQTGRQQIGYINIRTR
EAGMILAKGIEVTAMPSIRDGKIAYATMWIDGSDVQLGIGVLDTATGTTRMMPAGAEGAQIGQTAFTDDGVFWIE
DGAGTDPKAAVVRADADGTNRRTTVVPETGQGS LHAYQLTASDDAVNVTS LPPATSWANETLPKLFQVAPDGKGGGA
QRVSCNRGDQVYASADEGKR VVWLDGTTGYTNVVKRDRPAGAC*

SEQ ID NO: 109

Nucleotides 3078-4103 of SEQ ID NO: 106

SEQ ID NO: 110

VTIPPVGRTGWDTQWQALGLGADELRVYEALLDAPVPASRTALAQHLGLTVRRVTNALDHLAERRFTHPARGAGLP
VAMAPATALRNLIHLHQAEELLHRSAAEELSGSVDRIAAQLSSVHTPRATGIETVRGGAAIAARVASLLVSASEE
IALLDPRPPYASSEPDGMPVPLDMAEPVRRGVRVRVVVDREGLSFHGRARGLDLAVQGVQIRVGTDLPTKLITVDR
RVTLLPPTDAADPTASALVVS DALLSNALVPLFESVWERALPIGSVTHDQITDGDRELLTMLASGLKDEAMARRLD
IHVHTVRRRITRLMQVLNAETRFQAGVQAALRGWLTG*

SEQ ID NO: 111

Nucleotides 5246-4131 of SEQ ID NO: 106

SEQ ID NO: 112

MALASERQEFILATVRERGTVRLADLVERLGVTAVTVRRDVTILADRGLVSRVHGGVTLPYRGPTADERATRGVFT
HGRLPDQAFVGMVVPVTEYYWPAVIQGAQSAVAAAGGRLVLRASAYDAAEDRRQVTGLLDRGVRTLLVAPTNTGEA
GQDLLRWLGSLNVPVILVERLPPPALPTLPLDAVTTAHS LGAGLAVRHLVTLGHHRIAFTARFSPPTTKALREGWQ
ETTASLGLPSGDGLAHDVPSYGPSGWADAYDAALRSCREAGATALFVHSDGEAVGLVERAHEHGLAVPEDLAVITY
DDEVAAAADPPLTAVRPQKHLRGVLAEMALARAVDPVERPVHRVELWPNLIIRASCGGTAPVSAGR*

SEQ ID NO: 113

Nucleotides 5536-6888 of SEQ ID NO: 106

SEQ ID NO: 114

VLGFSSDAGDSAHAPRTHRSRRSLLRTSAGLGAALLAGGTLTGCSGRTSGTVRVRMWSWLTGM DRYVAAFNSSQRD
VRVDLSVIAAGTSGGYAQQTNAIRAHNAPDILHVEYQALPQIATTGGLRDLSADVDDLADGFLPAWQSVRPGGKT
WAVPMDFCPMAFFYRKDLFDRAGIGVPRTWEEFRRAASAVRTADR DARITTFPLNDGAFFAGMACQAGDPWWDIAG
NAWRVGIDGAGTMRVAEYWQDMVSSGLAARDATGTQSWISAMHHGRLWGMLGATWGVGMLKKSLPAHQGRWAVAPL
PTWGDDPSTGVWGGSAGFVSAESEVPEAALTFLRWLSTDPVPRIGSEFTFPSPAFLANRKVARTAYEDGYFTGDP
VFDVLDESATRVPGWTWGPSTLSLSTFATIAEALGPVSTGATTIPRALHRVHTSTVAAMRARGLAVTEGSRT

SEQ ID NO: 115

Nucleotides 7017-7814 of SEQ ID NO: 106

SEQ ID NO: 116

VLIPIGYAFYLSLFADRLSGLGFDGYRQVLTGFDNYTDVLADPAARDSVANVARYAAIHIPVMFGGALLIALLLNS
ATVRLRQLWSLAVFLPYAVPGVIAGLIWGYLYSPGIGPLADVLPFDPFGESGVLP SIVNMATWQWTGYNV IIFYTA
LQTVPRDVLEAARADGAGALRTALS IKIPMIRPTIFVALVFTTIGSLQLFTEPMVLRFTTGAVTSTWTPSLYIYDA
AFIRNDYGRAAAASILLALASALLSALVMRLSSRRSR

SEQ ID NO: 117

Nucleotides 7943-8743 of SEQ ID NO: 106

SEQ ID NO: 118

VLLLAALYSVLPLVWLVT SATKSVGDFSATS AFELADFN LWGNLRDLFAREDGVFLHWIRNSLLYAGVGAVLGALI
CTACGYAVAKLDFPGRRALFAVTLAGVLVPPTALALPLYLLASNHLVNTFWAVFLPLL SNPFGVYLARTFADAAV
PDEVLEAARLDGAGEFRIFLTVALPMMMPGFVTIVLFQFVSIWNNFFLPLVMLTDPELFPMTLGIYQWNTRAPQFP
DYNPLVITGSLLAVVPLIIAFVVLQRQWRSGLAAGSVK*

SEQ ID NO: 119

Nucleotides 8815-9795 of SEQ ID NO: 106

SEQ ID NO: 120

MNDRTRHTQFGPEELDR LHAVASLGQPLGAADLD SGAVRARLAEVEVLITSWGCPTLDHRVLRAAPRLRAVLHAAG
SVRGHVPDAVFDRLVLTAAADSNAEPVAQYTLGAILWAFKKVPFLAMDARRSREDWSYLHRRGELSGRDRTVVLV
GFSRIGRRVVDLLRTL DLARVLVVD PVADPAEVVAAGARPAQLAEALPQADVLSLHAPALPGTRHMIGAAQLAVLP
PGAVLVNTARGSLVDTAALAEACRAGLHAVLDVTDPEPLPADSPLYDLHNVLTPHIAGSLGSETRTMAGAALTEL
ERYAAGLPPLAPVTRQSLTTQA*

SEQ ID NO: 121

Nucleotides 9843-11852 of SEQ ID NO: 106

SEQ ID NO: 122

MNHQHPGRRTVLRAGGLAGLAAAAGLAANPAAAAPAAARLPRIDDLGPAVSQFSLMSAVLTGDTVHIGSRNIEPVQ
IISVHLPTRKVVARTELSNGHSIQALAADDSGRYLYAGILQKASGDRPNLFRWDLSTPDEPAEPLGAIGDRDVRAL
AVAPDGVVYAVGGSSGAPALWEYDPETGRISNVGTPDPRATLARAVAATRTHAFFGAGTTFNGGGDTSRAALYAY
DRSAGTFTSIAPPEMLRDPISIRDLAVLGDR LAVSTAGSTEDAKVALMDLKDLSYKIATSIGKVAKSFTMTGDTVH
YANESGLLAYLPSANTVTPVAYDGPGLGEIWGVDSRAGKAVVTSAFGFVAEIDPARRTAAVIDLGTAGAPMAPQTV
MGIAAGGGHVYVGGNATIARHTPGRDTPVVKLRAPGEAKDAVVVNGILYTQYNSSEGIWCHDPRTGKTQVAAFPSE
QNRPLDVTDFEDTGLVLVGVQSDTEGGGALWYHPRTGRQSFADPVDEAQCVRGSIASRNGVAFLGGDNTAPTGP
STVVAFPDVAGRELWRIDPQQTAGTAALAVRGRHLYGLSRKGGLFVIDIPRELIHRADVTSVSNNGFAALVTNRGV
VYGVSDTTVFRFHPRTFAVSVVVPDIDGGWYSGAHITNDEHGRLYTMRGRNLVRITDRPGH*

SEQ ID NO: 123

Nucleotides 11860-12738 of SEQ ID NO: 106

SEQ ID NO: 124

VSPPTPGTTPRHPTPSQGASMKNAHKRICIAVCLGAVAGAGYATTATATPEKSARPPQASDLLDVYESGALDSGI
PGLTTTHAKASDASTVESGGDGAHSVSHKVTLNPDYVSDGAPRSESANNEIEKSLIHVGDVHRYEFSVMLKDWE
TSSSKAGDII FQKGHAGGNKPSFYLMQRNAIAFRSPLDLQAPVVEDFRPYVDQWMRFRI DVRWAESRTGYKVS
VRLPGEKDFTLKKTYTDVNTFHPENPTAFGYIKWGLYRPASSTEAGDPATRVVRHDDIRVIDLG*

SEQ ID NO: 125

Nucleotides 13799-12783 of SEQ ID NO: 106

SEQ ID NO: 126

MPPEPPAPGTESAAAATEKSGTASLPPTTEESGPAAPPSAAPAEARKPRSEAATTAGTGEPPTPEAATTGTEKAQT
QAPATTTTTEAPEGRTEAIALAAVAAPKPTKTLTAADRGRPRAPVLGAAAFVGAALVAIPVLLMGSADEKPRNS
GTAPVAGSADTVLNPESAPAALDDYVAVKPTPSPTEQKKTTPPKAAPVAPSPAPRTSAPAEKPKASPKPKPKAA
PKPDWGTQTVSATSSIGVGQSWATNRIRMTMQQDGNLVVYNERNKP IWAAMTFGENHRAIFQPDGNLVIHNGDDRA
IWASKTHDFGGAQMVLRLPDAKVIVHQQKVVWST

SEQ ID NO: 127

Nucleotides 14051-14674 of SEQ ID NO: 106

SEQ ID NO: 128

VDASLVPEAVKDAARRAPGHWWGIVDPEWTQARTPPEWAVLGEWQSDSGSVGEYRANPGYRPSARVLGWPEPTDP
VDAAAQRAATGYGTVD EALAEADVTVVRGPDGGPLLAAGRDGALVALLFTSPTHAFMSAALHHD TVSARELAG
SLAQTGVLVSVNAGAAAPLLVPTDSLAVSSGPETAADGPAPTEPWPHTTGRTP*

SEQ ID NO: 129

Nucleotides 14671-15846 of SEQ ID NO: 106

SEQ ID NO: 130

VTRSSSQPPASASAPAQAKTEAAPAEATEAEKVKEVEEGEKAEOGERSEAQSSERQSPQVQAQAPGAGAENPKAP
TASETDPDPGASAAAQSRPALVRTMTATAIGKPQQEAGPVGRPGKAALAGAAVVGALLVSVFPFLVLVLAGNNDGPE
RTKAAGAGTVLDGSGPEAPGEFAVTAPEAGAPAEDREKREKDEAKPEKGGNPVPAIPPGGGGKAETKEDNAPNDPP
KKADSSKGKSKEQPEAGSGGGRNQPAKAQPAVTFSGPVSFRSHLSGRCLDVPGHNFNDGQPLFMWDCNGADAQKWR
FGSDGTIRARDKCLDVANADFANGSRIQLAWCNGSAAQKFTLNGAHDLVNTVVGKCVDIPNHSKGRGPETYLILWE
CTGLDNQKWST*

SEQ ID NO: 131

Nucleotides 15954-19265 of SEQ ID NO: 106

SEQ ID NO: 132

VSROVLTVGPADRFSTIGEALAAARTGALISVRPGTYAENLVIHTRVTLTAAEGRGTVEIRPRSGSVVALRADAVM
LSELTLRGGDSELPADVRRGQAAFDGCEIVGAAWTAMLAGGTGSLALRDCRVSNPRGAGIVVTSATPTTVESCTL
EHLGTSGIVLAEQGEARVDRCTVRGARGNGLLANGETRGTVEDCDISSTDKPSIALEGDSAVSVVVRTVVHDTSTGV
HLSSAGRTTLEDVRVTGASNGIVLAVGTDVLRRCRVSRARGHGLFVTD RARGTFEDCWVDGAQGAALRVAGASS
PALTGLTVRDCEGAGLLLEEDAAPELDRLEVIGSSPAVALQGGANPLLRARLVEPAGDGIATKDARGRAEDCEI
VRPKGAGVRVASGSTLYLAGGGVSDTATNGLVVEDGGNVTVRDFRVEISGEEGVVVAAGGELTANRTTVHAPKGHG
FLLREGALASLSGCEANGGAQDGFVSTAPVSLVNCTARENEGGLVQTAPGERLAVDGLNSVGNKGRDAWGS
AENTDPAGSGAADAPPPDRADGPLGALNALIGLENVKQVRTLNLTLQAQRREQLGMPAPPMSRHIFAGPPGTG
KTTVARLYGAILAELGSLRSGHLVEVSRADLVAQVVGTAIKTSETFQALGGVLFIDEAYTLTADSGNGGADFG
EAVDTLLKLMDHRDDVVVVAAGYSREMESFLSSNPGLASRFSRTVEFENYSVDDLVAIMESMCTQHQYELGEGTA
QALAAHFAMDRDAGFGNGRAARGVFEEMVDRQAIRLSAQQAQVSENDLRLLLPEDVSATAAASAAETAVPADDPLT
RLGDMIGLAGVKREVADLVNLITTAHRAAGLPVPTLSNHLVFTGPPGTGKTTVARLYGEVLTQLGVLERGQLIE
AARADLVGRYIGHTAQLTREVFEEKARGGVLFIDEAYTLTPRGGGADFGQEAVDTLKLMDHRDEVVVIVAGYTDE
MERFLASNPLSSRFPRIAFADYSSEELVTIVRAQAAMGYECGPGTGPLLKEHFDVPRDRSFGNARLARQVVE
SMVTRQAGRISSLAAPTLDLRLILLPADVTAAAPKAVQQ*

SEQ ID NO: 133

Nucleotides 19262-20530 of SEQ ID NO: 106

SEQ ID NO: 134

VRPPARRATGRPRAGRLLYGVSLAAALVVPVAVPAQAAPVTLTPAADGAKGQELPGMPTALDARAEAVACTPASKE
QAKKQDWSRQRLDLRLHRHTTGGSGVTVALISTGVDPGAEGLDGRVTAQQAADDCVGGTFLAGLIAGTGGPTPR
LAGVAPGAKILALRGTDRRGQPDPAALVTAAVRAATTAEADVIAVAVALPRKDTALTRAFAEARKAGAVVVASATPE
PPSRGSADEIPSRTYWPAGEPGVLAVADMLPAGARPDAALPTEGIDLAAPGAGVVGSGPRGKGHYLGGGVAVATAY
AAGAVA AVRATHPEATADAVARRLTATAYPADIPQLDVYAAVTTVLGDAGAPVGGAEAAEPVAVRDTSDTDRATG
RAVLVLLGSAGVLSLLWAGFAVTRARARGWRPAGAGTPDQG*

SEQ ID NO: 135

Nucleotides 23947-20585 of SEQ ID NO: 106

SEQ ID NO: 136

ILLRPEEAPDGATPGAHPPEHDAVRALVRAMLGQLAVFHAPEELWIAFCVSDERRADWEWVKWLPHVLDPHEEDGA
GQARRITADLTELDLLGAFAERPFGFDPDARPGRDEPYTVVLDGVSVPGEHRWEGHGYRNALILDVSGALRWRP
GRNTLRLTVGPGQVNLVVRTDRSRKERTVPLGRPDRLGPLGAELARLLTPRRMSLGTDIAQPLDTDVELTTLGIP
DLHRHDPGTLFARHSGSGRLRVPIAVGVDGRPVELDIKESAQGGMGPHGMLIGATGSGKSELLRTLVLGLALTNS
ETLNFVLVDFKGGATFLGLEELPHTSAVITNLADEVALVERMQDALHGELIRRQELLRAAGNYTSALEYERARAAG
ADLAPLPSLFVVVDEFSELLSTHREFMELFVMIGRLGRSLGVHLLASQRLDEGRMHQLESHLSYRIGLRTFSAME
SRGVLGVPDAYELPAAPGSGYLKSGVEALTRFRAAYSSGTYYRRRTGAVVQARVASQVVPWTSGWVVPRTLETTPDP
EPETEEAEDEEALLDVALDRLRGSGPDHQAQVWLPPLDEPSPLDALLPGITPDKERGLMAASWAGTGKLRVPVGLVD
KPFQRRDPLVVDLSGAGGHVAVAGGSQSGKSTVARTLIAALALHTPAEVQFYCLDFGGGGLSQAALPHVGGVA
ARLNPERVHRTVAEVMTLARREQFFVDHTLDSMQSYRRRRRAAGEFPDEPFGDVFMVVDGWSTVRQDYDGLIPKFN
ELAARGNLNYGIHLLITTRWVELSAQVRDQAATRLRLMGPMDSEIDTRKARSVPRTGGRGITADSKMHFLAGLP
RLDGSGLDLEGEVAHLVGEISRHWGPAAPQVRMLPHRLPVSELPSPEATEGGMRLVLGLDQDALEPVWHDFS
RTPHLIAVGDTESGKTNLLRLITQGITYAPEEAKIIAVDYRRTLVDALPEEYRIGHVISLDNLNETIEGAARAM
KTRVPGADISPARMRCDWWTGPRLFVLVDDYDMVSGNSFQSPFEPLFEHLTLGYEMGLHLVARSAMGAGRGLSD
GLIRRLDEANNPVLLSCPPTGRLFGNAKPLNLPGRALHIQRRKPRLVQTALVE*

SEQ ID NO: 137

Nucleotides 41305-41325 of SEQ ID NO: 1

GCTCGTCCCCCTCCCCGCACT

SEQ ID NO: 138

CGAACAGGTGGGCTTTGAGTGG

SEQ ID NO: 139

CTTCGTGAACACCCTCGTCC

SEQ ID NO: 140
GTTTCGTCGAGGTCCAGTACG

SEQ ID NO: 141
GCACCAGCGTGTGCGGATCG

SEQ ID NO: 142
CACGTACGTGACGATCCTCG

SEQ ID NO: 143
CGTACTGGACCTCGACGACC

SEQ ID NO: 144
CGACCAGCGTGTGTACGTCC

SEQ ID NO: 145
AGTCCTCAGCCATCTCCTCG

SEQ ID NO: 146
GAGACCGTCGGCGTGGACG

SEQ ID NO: 147
AGGGCCACACCGTCGAACTCC

SEQ ID NO: 148
ATCGTCGCCGACTACCTCGC

SEQ ID NO: 149
GGCAGCTACCTCGTACTGG

SEQ ID NO: 150
TGTACGACAGCGGCGTCGAAC

SEQ ID NO: 151
CGATTCTCGGCATGTTCGCC

SEQ ID NO: 152
TCGTCTCCTACATGACCTCG

SEQ ID NO: 153
TTCACGGAAACCGAACGTCG

SEQ ID NO: 154
GGTTCAGGCCGCGAGCCAACG

SEQ ID NO: 155
CGCTGACCTTGGTCAGAAGCC

SEQ ID NO: 156
GCGAAGCTTCTGGTGGCGCATCACCTGG

SEQ ID NO: 157
GCTCTAGATGGAAGTATGTCCTCCATCGC

SEQ ID NO: 158
CGGATCCCGCCGGCACCTGACCC

SEQ ID NO: 159
CCGAATTCCGCCTCCGAGTACATCGAGG

SEQ ID NO: 160

ACSSWSGGSGTSSCCTTCATGAA

SEQ ID NO: 161

ATGGTGTTTCGAGAACTAYCC

SEQ ID NO: 162

SCSCTSCAGGAGGGSHTSSTSTTCC

SEQ ID NO: 163

CCGAASACSACGTCGTCSCGSCC

SEQ ID NO: 164

CDA III protein (see Figure 3)

SEQ ID NO: 165

Probable hydrolase associated with CDA NRPS (see Figure 4)