

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
6 February 2003 (06.02.2003)

PCT

(10) International Publication Number
WO 03/010193 A2(51) International Patent Classification⁷: **C07K 14/195**Saint-Laurent, Québec H4L 5M5 (CA). **YANG, Xianshu**
[CA/US]; 39 Trowbridge Street, Belmont, MA 02478
(US).

(21) International Application Number: PCT/CA02/01177

(22) International Filing Date: 26 July 2002 (26.07.2002)

(74) Agent: **LOOPER, Ywe, J**; 7290 Frederick-Banting,
Saint-Laurent, Québec H4S 2A1 (CA).

(25) Filing Language: English

(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, PH, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU,
ZA, ZW.

(26) Publication Language: English

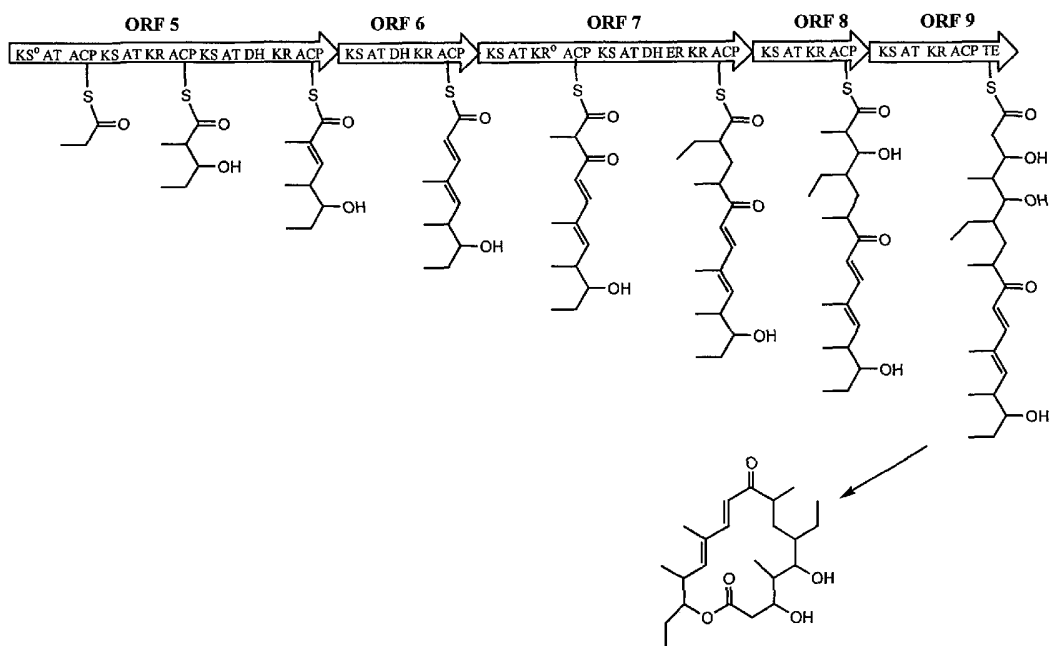
(30) Priority Data:
60/307,629 26 July 2001 (26.07.2001) US(71) Applicant (*for all designated States except US*): **ECOPIA
BIOSCIENCES INC.** [CA/CA]; 7290 Frederick-Banting,
Saint-Laurent, Québec H4S 2A1 (CA).(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, 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).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **FARNET, Chris,
M.** [US/CA]; 772 Bloomfield, Outremont, Québec H2V
3S3 (CA). **STAFFA, Alfredo** [CA/CA]; 395 Ouimet,

[Continued on next page]

(54) Title: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN



(57) Abstract: Genes and proteins involved in the biosynthesis of macrolides by microorganisms, in particular the nucleic acids forming the biosynthetic locus for the 16-member macrolide rosaramicin from *Micromonospora carbonacea*. These nucleic acids can be used to make expression constructs and transformed host cells for the production of rosaramicin. The genes and proteins allow direct manipulation of macrolides and related chemical structures via chemical engineering of the proteins involved in the biosynthesis of rosaramicin.

**Declarations under Rule 4.17:**

- as to the identity of the inventor (Rule 4.17(i)) for the following designations 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, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, 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)
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations 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, PH, PL, PT, RO, RU, SD, SE,

SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, 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)

- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations
- of inventorship (Rule 4.17(iv)) for US only

Published:

- without international search report and to be republished upon receipt of that report
- with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description

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.

TITLE OF INVENTION: Genes and proteins for the biosynthesis of rosaramicin

CROSS-REFERENCING TO RELATED APPLICATION:

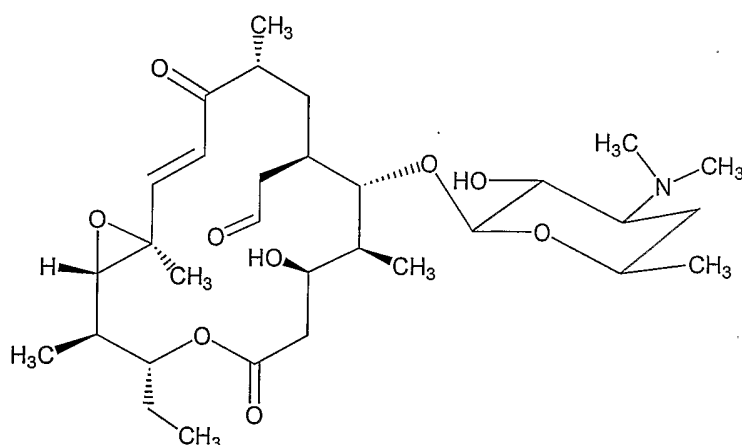
This application claims benefit under 35 USC §119 of provisional application USSN 60/307,629 filed on July 26, 2001 which is hereby incorporated by reference in its entirety for all purposes.

FIELD OF INVENTION:

The present invention relates to nucleic acid molecules that encode proteins
10 that direct the synthesis of macrolides, in particular the 16-member macrolide
rosaramicin. The present invention also is directed to the use of nucleic acids and
proteins to produce compounds exhibiting antibiotic activity based on the rosaramicin
structure.

BACKGROUND:

Rosaramicin is a 16-member macrolide antibiotic. Macrolides constitute a group
of antibiotics mainly active against Gram-positive bacteria. They have clinical
applications in the treatment of bacterial infections. Macrolides compounds are
structurally characterized by a macrolide lactone ring to which one or several deoxy-
20 sugars moieties are attached.



The carbohydrate ligands and macrolide lactone ring serve as molecular
recognition elements critical for biological activity. Variations in the sugar composition

of a macrolide or in the structure of the macrolide lactone ring may vary the biological activity of the molecule. Elucidation of gene clusters involved in the biosynthesis of rosaramicin expands the repertoire of genes and proteins useful to macrolides via combinatorial biosynthesis.

The increasing number of microbial strains that have acquired resistance to the currently available antibiotic compounds is recognized as a dangerous threat to public health. The genes and proteins involved in the biosynthesis of rosaramicin may be used to generate new unnatural compounds having desirable biological activity. The genes and proteins from the rosaramicin locus may also be used as probes to identify
10 new rosaramicin-like natural products.

The genome of many microorganisms contains multiple natural product biosynthetic loci that are not normally expressed in nature or under conventional experimental conditions. For example, twenty-five secondary metabolic gene clusters in the genome of the actinomycete *Streptomyces avermitilis* were identified by whole genome shotgun sequencing of the genome despite the fact that the organism was known to produce only two antimicrobial natural products (Osura *et al.* PNAS, vol. 98, no. 21 12215-12220). An important new source of antimicrobial compounds lies in the products of cryptic biosynthetic loci. It is desirable to discover and characterize a biosynthetic locus producing an antimicrobial product and present in the genome of
20 organisms not known to product the antimicrobial product of the locus.

SUMMARY OF THE INVENTION:

Micromonospora carbonacea is known to produce the antimicrobial orthosomycin natural product everninomicin. *Micromonospora carbonacea* was not previously reported to produce other natural products. We have surprisingly discovered, in the *Micromonospora carbonacea* genome, a type I polyketide biosynthetic gene cluster directed to the production of a rosaramicin-type polyketide.

The invention provides polynucleotides and polypeptides useful in the production and engineering of macrolides. In one embodiment, the polynucleotide
30 molecules are selected from the contiguous DNA sequence SEQ ID NO: 1. Other embodiments of the polynucleotides and polypeptides are provided in the accompanying sequence listing. SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 provide nucleic acids responsible for biosynthesis of the

16-member macrolide rosaramicin. SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 provide amino acid sequences for proteins responsible for biosynthesis of the 16-member macrolide rosaramicin. Certain embodiments of the invention specifically exclude one or more of open reading frames of the rosaramicin biosynthetic locus, most notably any one or more of ORFs 3, 11, 13, 16, 17 and 18 (SEQ ID NOS: 7, 23, 27, 33, 35 and 37) and the corresponding gene products (SEQ ID NOS: 6, 22, 26, 32, 34 and 36) deduced therefrom, although other ORFs and polypeptides listed in the sequence listing can be excluded from certain embodiments without departing from the scope of the invention.

10 The polynucleotides and polypeptides of the invention provide the machinery for producing novel compounds based on the structure of rosaramicin. The invention allows direct manipulation of rosaramicin and related chemical structures via chemical engineering of the enzymes involved in the biosynthesis of rosaramicin, modifications which may not be presently possible by chemical methodology because of the complexity of the structures. The invention can also be used to introduce "chemical handles" into normally inert positions that permit subsequent chemical modifications. Several general approaches to achieve the development of novel macrolides are facilitated by the methods and compositions of the present invention. For example, tylosin is structurally related to rosaramicin but, unlike rosaramicin, it does not contain
20 an epoxide. Accordingly, genes and proteins disclosed herein may be used to enzymatically create a tylosin derivative that contains an epoxide modification.

 Various macrolide structures can be generated by genetic manipulation of the rosaramicin gene cluster or use of various genes from the rosaramicin gene cluster in accordance with the methods of the invention. The invention can be used to generate a focused library of analogs around a macrolide lead candidate to fine-tune the compound for optimal properties. Genetic engineering methods of the invention can be directed to modify positions of the molecule previously inert to chemical modifications. Known techniques allow one to manipulate a known macrolide gene cluster either to produce the macrolide compound synthesized by that gene cluster at
30 higher levels than occur in nature or in hosts that otherwise do not produce the macrolide. Known techniques allow one to produce molecules that are structurally related to, but distinct from, the macrolide compounds produced from known macrolide gene clusters. Cloning, analysis, and manipulation by recombinant DNA technology of

genes that encode rosaramicin gene products can be performed according to known techniques.

Thus, in a first aspect the invention provides an isolated, purified or enriched nucleic acid comprising a sequence selected from the group consisting of SEQ ID NO: 1; the sequences complementary to SEQ ID NO: 1; fragments comprising at least 100, 200, 300, 500, 1000, 2000 or more consecutive nucleotides of SEQ ID NO: 1; and fragments comprising at least 100, 200, 300, 500, 1000, 2000 or more consecutive nucleotides of the sequences complementary to SEQ ID NO: 1. Preferred
10 embodiments of this aspect include isolated, purified or enriched nucleic acids capable of hybridizing to the above sequences under conditions of moderate or high stringency; isolated, purified or enriched nucleic acid comprising at least 100, 200, 300, 500, 1000, 2000 or more consecutive bases of the above sequences; and isolated, purified or enriched nucleic acid having at least 70%, 75%, 80%, 85%, 90%, 95%, 97% or 99% homology to the above sequences as determined by analysis with BLASTN version 2.0 with the default parameters.

Further embodiments of this aspect of the invention include an isolated, purified or enriched nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and the sequences complementary thereto; an isolated, purified or enriched nucleic acid
20 comprising at least 50, 75, 100, 200, 500, 800 or more consecutive bases of a sequence selected from the group consisting of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and the sequences complementary thereto; and an isolated, purified or enriched nucleic acid capable of hybridizing to the above listed nucleic acids under conditions of moderate or high stringency, and isolated, purified or enriched nucleic acid having at least 70%, 75%, 80%, 85%, 90%, 95%, 97% or 99% homology to the nucleic acid of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 as determined by analysis with BLASTN version 2.0 with the default parameters.

In a second embodiment, the invention provides an isolated or purified
30 polypeptide comprising a sequence selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38; an isolated or purified polypeptide comprising at least 50, 75, 100, 200, 300 or more consecutive amino acids of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22,

24, 26, 28, 30, 32, 34, 36, 38; and an isolated or purified polypeptide having at least 70%, 75%, 80%, 85%, 90%, 95%, 97%, or 99% homology to the polypeptide of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 as determined by analysis with BLASTP version 2.2.2 with the default parameters. In a further aspect, the invention provides a polypeptide comprising one or two or three or five or more or the above polypeptide sequences.

The invention also provides recombinant DNA expression vectors containing the above nucleic acids. The polynucleotides and the methods of the invention enable one skilled in the art to create recombinant host cells with the ability to produce macrolides.

10 Thus, the invention provides a method of preparing a macrolide compound, said method comprising transforming a heterologous host cell with a recombinant DNA vector that encodes at least one of the above nucleic acids, and culturing said host cell under conditions such that a macrolide is produced. In one aspect, the method is practiced with a *Streptomyces* host cell. In another aspect, the macrolide produced is rosaramicin. In another aspect, the macrolide produced is a compound related in structure to rosaramicin. The invention also provides a method for producing a rosaramicin compound by culturing *Micromonospora carbonacea* under conditions allowing for expression of its endogenous rosaramicin biosynthetic locus.

20 The invention also encompasses a method of invention for detecting by, *in silico* hybridization or traditional hybridization, putative macrolide gene clusters or macrolide-producing microorganisms using compositions of the invention. In one embodiment, a polypeptide encoding one or more of the polyketide synthase proteins (SEQ ID NOS: 10, 12, 14, 16 and 18) or fragments thereof are used as probes to detect putative macrolide gene clusters by *in silico* hybridization.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be further understood from the following description with reference to the following figures:

30 Figure 1 is a block diagram of a computer system which implements and executes software tools for the purpose of comparing a query to a subject, wherein the subject is selected from the reference sequences of the invention.

Figures 2A, 2B, 2C and 2D are flow diagrams of a sequence comparison software that can be employed for the purpose of comparing a query to a subject,

wherein the subject is selected from the reference sequences of the invention, wherein Figure 2A is the query initialization subprocess of the sequence comparison software, Figure 2B is the subject datasource initialization subprocess of the sequence comparison software, Figure 2C illustrates the comparison subprocess and the analysis subprocess of the sequence comparison software, and Figure 2D is the Display/Report subprocess of the sequence comparison software.

Figure 3 is a flow diagram of the comparator algorithm (238) of Figure 2C which is one embodiment of a comparator algorithm that can be used for pairwise determination of similarity between a query/subject pair.

10 Figure 4 is a flow diagram of the analyzer algorithm (244) of Figure 2C which is one embodiment of an analyzer algorithm that can be used to assign identity to a query sequence, based on similarity to a subject sequence, where the subject sequence is a reference sequence of the invention.

Figure 5 is a graphical depiction of the rosaramicin biosynthetic locus showing, at the top of the figure, the regions covered by the three deposited cosmid clones 010CK, 010CF and 010CJ; a scale in kilobase pairs; the positioning of the open reading frames on a continuous black line representing the continuous DNA sequence (SEQ ID NO: 1); and the relative position and orientation of 19 ORFs referred to by number at the bottom of figure.

20 Figure 6 illustrates the construction of the rosaramicin backbone by the Type 1 polyketide synthase enzymes (PKS) in the rosaramicin biosynthetic locus.

Figure 7 illustrates a mechanism for the biosynthesis of rosaramicin.

Figures 8A and 8B represent a Clustal amino acid alignment of the eight ketosynthase (KS) domains found in the rosaramicin PKS enzyme complex. Key residues are highlighted.

Figures 9A and 9B represent a Clustal amino acid alignment of the eight acyl transferase (AT) domains in the rosaramicin PKS enzyme complex. Key residues are highlighted. Regions important in substrate recognition are indicated by "s" above the alignment.

30 Figure 10 represents a Clustal amino acid alignment of the 3 DH domains in the rosaramicin PKS enzyme complex. Key residues are highlighted.

Figure 11 represents a Clustal amino acid alignment comparing the single enoyl reductase (ER) domain in the rosaramicin PKS enzyme complex to a prototypical ER

domain of the erythromycin PKS, *i.e.* 6-deoxyerythronolide B synthase (DEBS), key residues are highlighted.

Figure 12 represents a Clustal amino acid alignment of the 7 KR domains in the rosaramicin PKS enzyme complex. Key residues are highlighted.

Figure 13 represents a Clustal amino acid alignment of the 8 ACP domains in the rosaramicin PKS enzyme complex. The key active site serine residue is highlighted.

Figure 14 represents a Clustal amino acid alignment comparing the single thioesterase (Te) domain in the rosaramicin PKS enzyme complex to a prototypical Te domain of the erythromycin PKS, DEBS.

Figure 15 represents a Clustal amino acid alignment that demonstrates the overall high degree of homology between the second AT domain of ORF7 with two other ethylmalonyl-CoA-specific AT domains from the tylosin and niddamycin PKS complexes.

Figure 16 is a LCMS graph showing the production of a compound of the molecular weight of rosaramicin.

DETAILED DESCRIPTION OF THE INVENTION:

Throughout the description and the figures, the biosynthetic locus for rosaramicin from *Micromonospora carbonacea* is sometimes referred to as ROSA. The ORFs in ROSA are assigned a putative function sometimes referred to throughout the description and figures by reference to a four-letter designation, as indicated in Table I.

Table 1

Families	ORF#	Function
ABCC	1	ABC transporter; contains repeated domain
DATF	17	dehydratase/aminotransferase; SMAT family (secondary metabolism aminotransferase); transaminase
GTFA	11	glycosyl transferase
MTFA	12	methyltransferase, SAM-dependent; N,N-dimethyltransferases
MTRA	19	resistance methyltransferase; 23S ribosomal
NBPA	16	unknown, nucleotide (ATP/GTP) binding protein; may be involved in regulated proteolysis
OXRB	10	oxidoreductase; similar to NDP-hexose-3,4-isomerases (tautomerase)

OXRC	3, 4	oxidoreductase; cytP450 monooxygenase, hydroxylase; oxygen-binding site motif: LLxAGx(D,E); heme-binding pocket motif: GxGxHxCxGxxLxR, the cysteine is invariable and coordinates the heme
OXRH	13	oxidoreductase, NAD(P)-dependent; similar to crotonyl CoA reductases (CCR); similarity to some quinone oxidoreductases, zinc-containing alcohol dehydrogenases
PKSH	5-9	polyketide synthase, type I
REGM	15	regulator; similar to TylR global activator of the tylosin locus and the carbomycin AcyB2 positive regulator
REGS	14	regulator, may be positive regulator; similar to spiramycin SrmR, which specifically activates the production of spiramycin
SURA	18	sugar reductase; iron-sulfur (4Fe-4S) protein; may be involved in 1,2-migration of the amino group from C4 to C3 via the Schiff's base intermediate
TESA	2	thioesterase

The terms "macrolide producer" and "macrolide-producing organism" refer to a microorganism that carries the genetic information necessary to produce a macrolide compound, whether or not the organism is known to produce a macrolide compound. The terms "rosaramicin producer" and "rosaramicin-producing organism" refer to a microorganism that carries the genetic information necessary to produce a rosaramicin compound, whether or not the organism is known to produce a rosaramicin product. The terms apply equally to organisms in which the genetic information to produce the macrolide or rosaramicin compound is found in the organism as it exists in its natural environment, and to organisms in which the genetic information is introduced by recombinant techniques. For the sake of particularity, specific organisms contemplated herein include organisms of the family *Micromonosporaceae*, of which preferred genera include *Micromonospora*, *Actinoplanes* and *Dactylosporangium*; the family *Streptomycetaceae*, of which preferred genera include *Streptomyces* and *Kitasatospora*; the family *Pseudonocardiaceae*, of which preferred genera are *Amycolatopsis* and *Saccharopolyspora*; and the family *Actinosynnemataceae*, of which preferred genera include *Saccharothrix* and *Actinosynnema*; however the terms are intended to encompass all organisms containing genetic information necessary to produce a macrolide compound.

The term rosaramicin biosynthetic gene product refers to any enzyme or polypeptide involved in the biosynthesis of rosaramicin. The term "rosaramicin" is intended to encompass the compounds sometimes referred to as 4'-deoxycirramycin A1, rosamicin, izenamicin A1, juvenimicin A3, 6108A3, M 4365A2, Sch 14947, antibiotic 6108A3, antibiotic M 4365A2 and antibiotic Sch 14947. For the sake of

particularity, the rosaramicin biosynthetic pathway is associated with *Micromonospora carbonacea*. However, it should be understood that this term encompasses rosaramicin biosynthetic enzymes (and genes encoding such enzymes) isolated from any microorganism of the genus *Micromonospora* or *Streptomyces*, and furthermore that these genes may have novel homologues in related actinomycete microorganisms or non-actinomycete microorganisms that fall within the scope of the invention. Representative rosaramicin biosynthetic gene products include the polypeptides listed in SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 or homologues thereof.

10 The term "isolated" means that the material is removed from its original environment, e.g. the natural environment if it is naturally-occurring. For example, a naturally-occurring polynucleotide or polypeptide present in a living organism is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

20 The term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual nucleic acids obtained from a library have been conventionally purified to electrophoretic homogeneity. The purified nucleic acids of the present invention have been purified from the remainder of the genomic DNA in the organism by at least 10^4 to 10^6 fold. However, the term "purified" also includes nucleic acids which have been purified from the remainder of the genomic DNA or from other sequences in a library or other environment by at least one order of magnitude, preferably two or three orders of magnitude, and more preferably four or five orders of magnitude.

30 "Recombinant" means that the nucleic acid is adjacent to "backbone" nucleic acid to which it is not adjacent in its natural environment. "Enriched" nucleic acids represent 5% or more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. "Backbone" molecules include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids, and other vectors or nucleic acids used to maintain or manipulate a nucleic acid of interest. Preferably, the enriched nucleic acids represent 15% or more, more

preferably 50% or more, and most preferably 90% or more, of the number of nucleic acid inserts in the population of recombinant backbone molecules.

"Recombinant" polypeptides or proteins refer to polypeptides or proteins produced by recombinant DNA techniques, *i.e.* produced from cells transformed by an exogenous DNA construct encoding the desired polypeptide or protein. "Synthetic" polypeptides or proteins are those prepared by chemical synthesis.

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as, where applicable, intervening regions (introns) between
10 individual coding segments (exons).

A DNA or nucleotide "coding sequence" or "sequence encoding" a particular polypeptide or protein, is a DNA sequence which is transcribed and translated into a polypeptide or protein when placed under the control of appropriate regulatory sequences.

"Oligonucleotide" refers to a nucleic acid, generally of at least 10, preferably 15 and more preferably at least 20 nucleotides, preferably no more than 100 nucleotides, that are hybridizable to a genomic DNA molecule, a cDNA molecule, or an mRNA molecule encoding a gene, mRNA, cDNA or other nucleic acid of interest.

A promoter sequence is "operably linked to" a coding sequence recognized by
20 RNA polymerase which initiates transcription at the promoter and transcribes the coding sequence into mRNA.

"Plasmids" are designated herein by a lower case p preceded or followed by capital letters and/or numbers. The starting plasmids herein are commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described herein are known in the art and will be apparent to the skilled artisan.

"Digestion" of DNA refers to enzymatic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction
30 enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinary skilled artisan. For analytical purposes, typically 1 µg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 µl of buffer solution. For the purpose of

isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the gel electrophoresis may be performed to isolate the desired fragment.

We have now discovered the genes and proteins involved in the biosynthesis of the 16-member macrolide rosamycin. Nucleic acid sequences encoding proteins involved in the biosynthesis of rosamycin are provided in the accompanying
10 sequence listing as SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39. Polypeptides involved in the biosynthesis of rosamycin are provided in the accompanying sequence listing as SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38.

One aspect of the present invention is an isolated, purified, or enriched nucleic acid comprising one of the sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, the sequences complementary thereto, or a fragment comprising at least 100, 200, 300, 400, 500, 600, 700, 800 or more consecutive bases of one of the sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 or the sequences complementary thereto.
20 The isolated, purified or enriched nucleic acids may comprise DNA, including cDNA, genomic DNA, and synthetic DNA. The DNA may be double stranded or single stranded, and if single stranded may be the coding (sense) or non-coding (anti-sense) strand. Alternatively, the isolated, purified or enriched nucleic acids may comprise RNA.

As discussed in more detail below, the isolated, purified or enriched nucleic acids of one of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 may be used to prepare one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 respectively or fragments comprising at least 50, 75, 100, 200, 300, 500 or more consecutive amino acids of one
30 of the polypeptides of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38.

Accordingly, another aspect of the present invention is an isolated, purified or enriched nucleic acid which encodes one of the polypeptides of SEQ ID NOS: 2, 4, 6,

8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 or fragments comprising at least 50, 75, 100, 150, 200, 300 or more consecutive amino acids of one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38. The coding sequences of these nucleic acids may be identical to one of the coding sequences of one of the nucleic acids of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 or a fragment thereof or may be different coding sequences which encode one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 or fragments comprising at least 50, 75, 100, 150, 200, 300 consecutive amino acids of one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 as a result of the redundancy or degeneracy of the genetic code. The genetic code is well known to those of skill in the art and can be obtained, for example, from Stryer, *Biochemistry*, 3rd edition, W. H. Freeman & Co., New York.

The isolated, purified or enriched nucleic acid which encodes one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, may include, but is not limited to: (1) only the coding sequences of one of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39; (2) the coding sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and additional coding sequences, such as leader sequences or proprotein; and (3) the coding sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and non-coding sequences, such as introns or non-coding sequences 5' and/or 3' of the coding sequence. Thus, as used herein, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide that includes only coding sequence for the polypeptide as well as a polynucleotide that includes additional coding and/or non-coding sequence.

The invention relates to polynucleotides based on SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 but having polynucleotide changes that are "silent", for example changes which do not alter the amino acid sequence encoded by the polynucleotides of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39. The invention also relates to polynucleotides which have nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38. Such nucleotide changes may be introduced

using techniques such as site directed mutagenesis, random chemical mutagenesis, exonuclease III deletion, and other recombinant DNA techniques.

The isolated, purified or enriched nucleic acids of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, the sequences complementary thereto, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 consecutive bases of one of the sequence of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, or the sequences complementary thereto may be used as probes to identify and isolate DNAs encoding the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 respectively. In such procedures, a genomic DNA library is constructed from a sample microorganism or a sample containing a microorganism capable of producing a macrolide. The genomic DNA library is then contacted with a probe comprising a coding sequence or a fragment of the coding sequence, encoding one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or a fragment thereof under conditions which permit the probe to specifically hybridize to sequences complementary thereto. In a preferred embodiment, the probe is an oligonucleotide of about 10 to about 30 nucleotides in length designed based on a nucleic acid of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39. Genomic DNA clones which hybridize to the probe are then detected and isolated. Procedures for preparing and identifying DNA clones of interest are disclosed in Ausubel *et al.*, Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. 1997; and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual 2d Ed., Cold Spring Harbor Laboratory Press, 1989. In another embodiment, the probe is a restriction fragment or a PCR amplified nucleic acid derived from SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39.

The isolated, purified or enriched nucleic acids of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, the sequences complementary thereto, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 consecutive bases of one of the sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, or the sequences complementary thereto may be used as probes to identify and isolate related nucleic acids. In some embodiments, the related nucleic acids may be genomic DNAs (or

cDNAs) from potential macrolide producers. In such procedures, a nucleic acid sample containing nucleic acids from a potential macrolide-producer or rosaramycin-producer is contacted with the probe under conditions that permit the probe to specifically hybridize to related sequences. The nucleic acid sample may be a genomic DNA (or cDNA) library from the potential macrolide-producer. Hybridization of the probe to nucleic acids is then detected using any of the methods described above.

Hybridization may be carried out under conditions of low stringency, moderate stringency or high stringency. As an example of nucleic acid hybridization, a polymer membrane containing immobilized denatured nucleic acids is first prehybridized for 30 minutes at 45 °C in a solution consisting of 0.9 M NaCl, 50 mM NaH₂PO₄, pH 7.0, 5.0 mM Na₂EDTA, 0.5% SDS, 10X Denhardt's, and 0.5 mg/ml polyriboadenylic acid. Approximately 2×10^7 cpm (specific activity $4-9 \times 10^8$ cpm/ug) of ³²P end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in 1X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM Na₂EDTA) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at T_m-10°C for the oligonucleotide probe where T_m is the melting temperature. The membrane is then exposed to autoradiographic film for detection of hybridization signals.

By varying the stringency of the hybridization conditions used to identify nucleic acids, such as genomic DNAs or cDNAs, which hybridize to the detectable probe, nucleic acids having different levels of homology to the probe can be identified and isolated. Stringency may be varied by conducting the hybridization at varying temperatures below the melting temperatures of the probes. The melting temperature of the probe may be calculated using the following formulas:

For oligonucleotide probes between 14 and 70 nucleotides in length the melting temperature (T_m) in degrees Celcius may be calculated using the formula:

$T_m = 81.5 + 16.6(\log [Na^+]) + 0.41(\text{fraction G+C}) - (600/N)$ where N is the length of the oligonucleotide.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation $T_m = 81.5 + 16.6(\log [Na^+]) + 0.41(\text{fraction G + C}) - (0.63\% \text{ formamide}) - (600/N)$ where N is the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 0.1 mg/ml denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's

reagent, 0.5% SDS, 0.1 mg/ml denatured fragmented salmon sperm DNA, 50% formamide. The composition of the SSC and Denhardt's solutions are listed in Sambrook et al., *supra*.

Hybridization is conducted by adding the detectable probe to the hybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured by incubating at elevated temperatures and quickly cooling before addition to the hybridization solution. It may also be desirable to similarly denature single stranded probes to eliminate or diminish formation of secondary structures or oligomerization. The filter is contacted with the hybridization solution for a sufficient
10 period of time to allow the probe to hybridize to cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization may be carried out at 15-25 °C below the T_m . For shorter probes, such as oligonucleotide probes, the hybridization may be conducted at 5-10 °C below the T_m . Preferably, the hybridization is conducted in 6X SSC, for shorter probes. Preferably, the hybridization is conducted in 50% formamide containing solutions, for longer probes.

All the foregoing hybridizations would be considered to be examples of hybridization performed under conditions of high stringency.

Following hybridization, the filter is washed for at least 15 minutes in 2X
20 SSC, 0.1% SDS at room temperature or higher, depending on the desired stringency. The filter is then washed with 0.1X SSC, 0.5% SDS at room temperature (again) for 30 minutes to 1 hour.

Nucleic acids which have hybridized to the probe are identified by conventional autoradiography and non-radioactive detection methods.

The above procedure may be modified to identify nucleic acids having decreasing levels of homology to the probe sequence. For example, to obtain nucleic acids of decreasing homology to the detectable probe, less stringent conditions may be used. For example, the hybridization temperature may be decreased in increments of 5 °C from 68 °C to 42 °C in a hybridization buffer having a Na^+ concentration of
30 approximately 1M. Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate stringency" conditions above 50°C and "low stringency" conditions below 50°C. A specific example of "moderate stringency" hybridization conditions is when

the above hybridization is conducted at 55°C. A specific example of "low stringency" hybridization conditions is when the above hybridization is conducted at 45°C.

Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42 °C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50 °C. These conditions are considered to be "moderate stringency" conditions above 25% formamide and "low stringency" conditions below 25% formamide. A specific example of "moderate stringency" hybridization conditions is when the above hybridization is conducted at 30% formamide. A specific example of "low stringency" hybridization conditions is when the above hybridization is conducted at 10% formamide.

Nucleic acids which have hybridized to the probe are identified by conventional autoradiography and non-radioactive detection methods.

For example, the preceding methods may be used to isolate nucleic acids having a sequence with at least 97%, at least 95%, at least 90%, at least 85%, at least 80%, or at least 70% homology to a nucleic acid sequence selected from the group consisting of the sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, fragments comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases thereof, and the sequences complementary thereto. Homology may be measured using BLASTN version 2.0 with the default parameters. For example, the homologous polynucleotides may have a coding sequence that is a naturally occurring allelic variant of one of the coding sequences described herein. Such allelic variant may have a substitution, deletion or addition of one or more nucleotides when compared to the nucleic acids of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, or the sequences complementary thereto.

Additionally, the above procedures may be used to isolate nucleic acids which encode polypeptides having at least 99%, 95%, at least 90%, at least 85%, at least 80%, or at least 70% homology to a polypeptide having the sequence of one of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or fragments comprising at least 50, 75, 100, 150, 200, 300 consecutive amino acids

thereof as determined using the BLASTP version 2.2.2 algorithm with default parameters.

Another aspect of the present invention is an isolated or purified polypeptide comprising the sequence of one of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 or fragments comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof. As discussed herein, such polypeptides may be obtained by inserting a nucleic acid encoding the polypeptide into a vector such that the coding sequence is operably linked to a sequence capable of driving the expression of the encoded polypeptide in a suitable host cell. For example, the expression vector may comprise a promoter, a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for modulating expression levels, an origin of replication and a selectable marker.

Promoters suitable for expressing the polypeptide or fragment thereof in bacteria include the *E.coli lac* or *trp* promoters, the *lacI* promoter, the *lacZ* promoter, the T3 promoter, the T7 promoter, the *gpt* promoter, the λ P_R promoter, the λ P_L promoter, promoters from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), and the acid phosphatase promoter. Fungal promoters include the α factor promoter. Eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, heat shock promoters, the early and late SV40 promoter, LTRs from retroviruses, and the mouse metallothionein-I promoter. Other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses may also be used.

Mammalian expression vectors may also comprise an origin of replication, any necessary ribosome binding sites, a polyadenylation site, splice donors and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. In some embodiments, DNA sequences derived from the SV40 splice and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Vectors for expressing the polypeptide or fragment thereof in eukaryotic cells may also contain enhancers to increase expression levels. Enhancers are cis-acting elements of DNA, usually from about 10 to about 300 bp in length that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the

late side of the replication origin bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and the adenovirus enhancers.

In addition, the expression vectors preferably contain one or more selectable marker genes to permit selection of host cells containing the vector. Examples of selectable markers that may be used include genes encoding dihydrofolate reductase or genes conferring neomycin resistance for eukaryotic cell culture, genes conferring tetracycline or ampicillin resistance in *E. coli*, and the *S. cerevisiae* TRP1 gene.

10 In some embodiments, the nucleic acid encoding one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or fragments comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof is assembled in appropriate phase with a leader sequence capable of directing secretion of the translated polypeptides or fragments thereof. Optionally, the nucleic acid can encode a fusion polypeptide in which one of the polypeptide of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof is fused to heterologous peptides or polypeptides, such as N-terminal identification peptides which impart desired characteristics such as increased stability or simplified purification or detection.

20 The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is ligated to the desired position in the vector following digestion of the insert and the vector with appropriate restriction endonucleases. Alternatively, appropriate restriction enzyme sites can be engineered into a DNA sequence by PCR. A variety of cloning techniques are disclosed in Ausbel et al. Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. 1997 and Sambrook et al., Molecular Cloning: A Laboratory Manual 2d Ed., Cold Spring Harbour Laboratory Press, 1989. Such procedures and others are deemed to be within the scope of those skilled in the art.

30 The vector may be, for example, in the form of a plasmid, a viral particle, or a phage. Other vectors include derivatives of chromosomal, nonchromosomal and synthetic DNA sequences, viruses, bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. A variety of cloning

and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989).

Particular bacterial vectors which may be used include the commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017), pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden), pGEM1 (Promega Biotec, Madison, WI, USA) pQE70, pQE60, pQE-9 (Qiagen), pD10, phiX174, pBluescript II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene), ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia), pKK232-8 and pCM7. Particular
10 eukaryotic vectors include pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia). However, any other vector may be used as long as it is replicable and stable in the host cell.

The host cell may be any of the host cells familiar to those skilled in the art, including prokaryotic cells or eukaryotic cells. As representative examples of appropriate hosts, there may be mentioned: bacteria cells, such as *E. coli*, *Streptomyces*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, fungal cells, such as yeast, insect cells such as *Drosophila S2* and *Spodoptera Sf9*, animal cells such as CHO, COS or Bowes melanoma, and adenoviruses. The selection of an appropriate host is
20 within the abilities of those skilled in the art.

The vector may be introduced into the host cells using any of a variety of techniques, including electroporation transformation, transfection, transduction, viral infection, gene guns, or Ti-mediated gene transfer. Where appropriate, the engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter may be induced by appropriate means (e.g., temperature shift or chemical induction) and the cells may be cultured for an additional period to allow them to produce the desired polypeptide or
30 fragment thereof.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract is retained for further purification. Microbial cells employed for expression of proteins can be disrupted by any convenient

method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known to those skilled in the art. The expressed polypeptide or fragment thereof can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the polypeptide. If desired, high performance liquid chromatography (HPLC) can be employed for final purification steps.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts (described by Gluzman, Cell, 23:175(1981)), and other cell lines capable of expressing proteins from a compatible vector, such as the C127, 3T3, CHO, HeLa and BHK cell lines.

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Depending upon the host employed in a recombinant production procedure, the polypeptide produced by host cells containing the vector may be glycosylated or may be non-glycosylated.

Polypeptides of the invention may or may not also include an initial methionine amino acid residue.

Alternatively, the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or fragments comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof can be synthetically produced by conventional peptide synthesizers. In other embodiments, fragments or portions of the polynucleotides may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides.

Cell-free translation systems can also be employed to produce one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or fragments comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof using mRNAs transcribed from a DNA construct comprising a promoter operably linked to a nucleic acid encoding the polypeptide or fragment

thereof. In some embodiments, the DNA construct may be linearized prior to conducting an *in vitro* transcription reaction. The transcribed mRNA is then incubated with an appropriate cell-free translation extract, such as a rabbit reticulocyte extract, to produce the desired polypeptide or fragment thereof.

The present invention also relates to variants of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or fragments comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof. The term "variant" includes derivatives or analogs of these polypeptides. In particular, the variants may differ in amino acid sequence from the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

The variants may be naturally occurring or created *in vitro*. In particular, such variants may be created using genetic engineering techniques such as site directed mutagenesis, random chemical mutagenesis, Exonuclease III deletion procedures, and standard cloning techniques. Alternatively, such variants, fragments, analogs, or derivatives may be created using chemical synthesis or modification procedures.

Other methods of making variants are also familiar to those skilled in the art. These include procedures in which nucleic acid sequences obtained from natural isolates are modified to generate nucleic acids that encode polypeptides having characteristics which enhance their value in industrial or laboratory applications. In such procedures, a large number of variant sequences having one or more nucleotide differences with respect to the sequence obtained from the natural isolate are generated and characterized. Preferably, these nucleotide differences result in amino acid changes with respect to the polypeptides encoded by the nucleic acids from the natural isolates.

For example, variants may be created using error prone PCR. In error prone PCR, DNA amplification is performed under conditions where the fidelity of the DNA polymerase is low, such that a high rate of point mutation is obtained along the entire length of the PCR product. Error prone PCR is described in Leung, D.W., *et al.*, Technique, 1:11-15 (1989) and Caldwell, R. C. & Joyce G.F., PCR Methods Applic., 2:28-33 (1992). Variants may also be created using site directed mutagenesis to

generate site-specific mutations in any cloned DNA segment of interest.

Oligonucleotide mutagenesis is described in Reidhaar-Olson, J.F. & Sauer, R.T., *et al.*, Science, 241:53-57 (1988). Variants may also be created using directed evolution strategies such as those described in US patent nos. 6,361,974 and 6,372,497. The variants of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, may be (i) variants in which one or more of the amino acid residues of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code.

Conservative substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the following replacements: replacements of an aliphatic amino acid such as Ala, Val, Leu and Ile with another aliphatic amino acid; replacement of a Ser with a Thr or vice versa; replacement of an acidic residue such as Asp or Glu with another acidic residue; replacement of a residue bearing an amide group, such as Asn or Gln, with another residue bearing an amide group; exchange of a basic residue such as Lys or Arg with another basic residue; and replacement of an aromatic residue such as Phe or Tyr with another aromatic residue.

Other variants are those in which one or more of the amino acid residues of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 includes a substituent group.

Still other variants are those in which the polypeptide is associated with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).

Additional variants are those in which additional amino acids are fused to the polypeptide, such as leader sequence, a secretory sequence, a proprotein sequence or a sequence which facilitates purification, enrichment, or stabilization of the polypeptide.

In some embodiments, the fragments, derivatives and analogs retain the same biological function or activity as the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38. In other embodiments, the fragment, derivative or analogue includes a fused heterologous sequence which

facilitates purification, enrichment, detection, stabilization or secretion of the polypeptide that can be enzymatically cleaved, in whole or in part, away from the fragment, derivative or analogue.

Another aspect of the present invention are polypeptides or fragments thereof which have at least 70%, at least 80%, at least 85%, at least 90%, or more than 95% homology to one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or a fragment comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof. Homology may be determined using a program, such as BLASTP version 2.2.2 with the default parameters, which
10 aligns the polypeptides or fragments being compared and determines the extent of amino acid identity or similarity between them. It will be appreciated that amino acid "homology" includes conservative substitutions such as those described above.

The polypeptides or fragments having homology to one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or a fragment comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof may be obtained by isolating the nucleic acids encoding them using the techniques described above.

Alternatively, the homologous polypeptides or fragments may be obtained through biochemical enrichment or purification procedures. The sequence of
20 potentially homologous polypeptides or fragments may be determined by proteolytic digestion, gel electrophoresis and/or microsequencing. The sequence of the prospective homologous polypeptide or fragment can be compared to one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or a fragment comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof using a program such as BLASTP version 2.2.2 with the default parameters.

The polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or fragments, derivatives or analogs thereof comprising at least 40, 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof invention may
30 be used in a variety of applications. For example, the polypeptides or fragments, derivatives or analogs thereof may be used to catalyze certain biochemical reactions. In particular, the polypeptides of the TESA family, namely SEQ ID NO: 4 or fragments, derivatives or analogs thereof; the PKSH family, namely SEQ ID NOS: 10, 12, 14, 16,

18 or fragments, derivatives or analogs thereof; the OXRH family, namely SEQ ID NO: 26 or fragments, derivatives or analogs thereof may be used in any combination, *in vitro* or *in vivo*, to direct or enhance the synthesis or modification of a polyketide, polyketide substructure, or precursor thereof. Polypeptides of the MTFA family, namely SEQ ID NO: 24 or fragments, derivatives or analogs thereof may be used, *in vitro* or *in vivo*, to catalyze methylation reactions that modify compounds that are either endogenously produced by the host, supplemented to the growth medium, or are added to a cell-free, purified or enriched preparation of MTFA polypeptide.

10 Polypeptides of the OXRC family, namely SEQ ID NOS: 6, 8 or fragments, derivatives or analogs thereof; the OXRB family, namely SEQ ID NO: 20 or fragments, derivatives or analogs thereof; the OXRH family, namely SEQ ID NO: 26 or fragments, derivatives or analogs thereof may be used, *in vitro* or *in vivo*, to catalyze oxidation reactions that modify compounds that are either endogenously produced by the host, supplemented to the growth medium, or are added to a cell-free, purified or enriched preparation of said polypeptide. Polypeptides of the NBPA family, namely SEQ ID NO: 32 or fragments, derivatives or analogs thereof; the OXRB family, namely SEQ ID NO: 20 or fragments, derivatives or analogs thereof; the DATF family, namely SEQ ID NO: 34 or fragments, derivatives or analogs thereof; the SURA family, namely SEQ ID NO: 36 or fragments, derivatives or analogs thereof; the MTFA family, namely SEQ ID NO: 24 or

20 fragments, derivatives or analogs thereof; the GTFA family, namely SEQ ID NO: 22 or fragments, derivatives or analogs thereof may be used, *in vitro* or *in vivo*, to catalyze biochemical reactions involved in activating, modifying, or transferring sugar moieties. Polypeptides of the ABCC family, namely SEQ ID NO: 2 or fragments, derivatives or analogs thereof; the MTRA family, namely SEQ ID NO: 38 or fragments, derivatives or analogs thereof may be used to confer to microorganisms or eukaryotic cells resistance to polyketides, macrolides, rosaramicin, or compounds related to rosaramicin. Polypeptides of the REGS family, namely SEQ ID NO: 28 or fragments, derivatives or analogs thereof; the REGM family, namely SEQ ID NO: 30 or fragments, derivatives or analogs thereof may be used to increase the yield of polyketides,

30 macrolides, rosaramicin, or compounds related to rosaramicin in either naturally producing organisms or heterologously producing recombinant organisms.

The polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or fragments, derivatives or analogues thereof comprising at

least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof, may also be used to generate antibodies which bind specifically to the polypeptides or fragments, derivatives or analogues. The antibodies generated from SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 may be used to determine whether a biological sample contains *Micromonospora carbonacea* or a related microorganism.

In such procedures, a biological sample is contacted with an antibody capable of specifically binding to one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof. The ability of the biological sample to bind to the antibody is then determined. For example, binding may be determined by labeling the antibody with a detectable label such as a fluorescent agent, an enzymatic label, or a radioisotope. Alternatively, binding of the antibody to the sample may be detected using a secondary antibody having such a detectable label thereon. A variety of assay protocols which may be used to detect the presence of an rosamycin-producer or of *Micromonospora carbonacea* or of polypeptides related to SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, in a sample are familiar to those skilled in the art. Particular assays include ELISA assays, sandwich assays, radioimmunoassays, and Western Blots. Alternatively, antibodies generated from SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, may be used to determine whether a biological sample contains related polypeptides that may be involved in the biosynthesis of natural products of the rosamycin class or other macrolides.

Polyclonal antibodies generated against the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptide itself. In this manner, even a sequence encoding only a fragment of the polypeptide can be used to generate antibodies that may bind to the whole native polypeptide. Such antibodies can then be used to isolate the polypeptide from cells expressing that polypeptide.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kholer and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof. Alternatively, transgenic mice may be used to express humanized antibodies to these polypeptides or fragments thereof.

Antibodies generated against the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof may be used in screening for similar polypeptides from a sample containing organisms or cell-free extracts thereof. In such techniques, polypeptides from the sample is contacted with the antibodies and those polypeptides which specifically bind the antibody are detected. Any of the procedures described above may be used to detect antibody binding. One such screening assay is described in "Methods for measuring Cellulase Activities", Methods in Enzymology, Vol 160, pp. 87-116.

As used herein, the term "nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39" encompass the nucleotide sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, fragments of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, nucleotide sequences homologous to SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, or homologous to fragments of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, and sequences complementary to all of the preceding sequences. The fragments include portions of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 consecutive nucleotides of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39. Preferably, the fragments are novel fragments.

Homologous sequences and fragments of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 80%, 75% or 70% identity to these sequences. Homology may be determined using any of the computer programs and parameters described herein, including BLASTN and TBLASTX with the default parameters. Homologous sequences also include RNA sequences in which uridines replace the thymines in the nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39.

10 The homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error. It will be appreciated that the nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 can be represented in the traditional single character format in which G, A, T and C denote the guanine, adenine, thymine and cytosine bases of the deoxyribonucleic acid (DNA) sequence respectively, or in which G, A, U and C denote the guanine, adenine, uracil and cytosine bases of the ribonucleic acid (RNA) sequence (see the inside back cover of Stryer, *Biochemistry*, 3rd edition, W. H. Freeman & Co., New York) or in any other format which records the identity of the nucleotides in a sequence.

20 "Polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38" encompass the polypeptide sequences of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 which are encoded by the nucleic acid sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, polypeptide sequences homologous to the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or fragments of any of the preceding sequences. Homologous polypeptide sequences refer to a polypeptide sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75% or 70% identity to one of the polypeptide sequences of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38. Polypeptide sequence homology may be determined using any of the computer programs and
30 parameters described herein, including BLASTP version 2.2.1 with the default parameters or with any user-specified parameters. The homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error. The polypeptide fragments comprise at least 5, 10,

15, 20, 25, 30, 35, 40, 50, 75, 100 or 150 consecutive amino acids of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38. Preferably the fragments are novel fragments. It will be appreciated that the polypeptide codes of the SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 can be represented in the traditional single character format or three letter format (see the inside back cover of Stryer, *Biochemistry*, 3rd edition, W.H. Freeman & Co., New York) or in any other format which relates the identity of the polypeptides in a sequence.

10 It will be readily appreciated by those skilled in the art that the nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, and the polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 can be stored, recorded and manipulated on any medium which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, and the polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38.

20 Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random Access Memory (RAM), or Read Only Memory (ROM) as well as other types of media known to those skilled in the art.

The nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, a subset thereof, the polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, and a subset thereof may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, one or more of the nucleic acid codes of SEQ ID
30 NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, and one or more of the polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 may be stored as ASCII or text in a word processing file, such as MicrosoftWORD or WORDPERFECT in a variety of database programs

familiar to those of skill in the art, such as DB2 or ORACLE. In addition, many computer programs and databases may be used as sequence comparers, identifiers or sources of query nucleotide sequences or query polypeptide sequences to be compared to one or more of the nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, and one or more of the polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38.

The following list is intended not to limit the invention but to provide guidance to programs and databases useful with one or more of the nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, and the polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38. The program and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group) Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul *et al.*, *J. Mol. Biol.* 215:403 (1990)), FASTA (Person and Lipman, *Proc. Natl. Acad. Sci. USA*, 85:2444 (1988)), FASTDB (Brutlag *et al. Comp. App. Biosci.* 6:237-245, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius².DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMM (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi (Molecular Simulations Inc.), QuanteMM (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WetLab (Molecular Simulations Inc.), WetLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwents' World Drug Index database, the BioByteMasterFile database, the Genbank database, and the Gensyqn database. Many other programs and databases would be apparent to one of skill in the art given the present disclosure.

Embodiments of the present invention include systems, particularly computer systems that store and manipulate the sequence information described herein. As

used herein, "a computer system", refers to the hardware components, software components, and data storage components used to analyze one or more of the nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, and the polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38.

Preferably, the computer system is a general purpose system that comprises a processor and one or more internal data storage components for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently
10 available computer systems are suitable.

The computer system of Figure 1 illustrates components that may be present in a conventional computer system. One skilled in the art will readily appreciate that not all components illustrated in Figure 1 are required to practice the invention and, likewise, additional components not illustrated in Figure 1 may be present in a computer system contemplated for use with the invention. Referring to the computer system of Figure 1, the components are connected to a central system bus 116. The components include a central processing unit 118 with internal 118 and/or external cache memory 120, system memory 122, display adapter 102 connected to a monitor 100, network adapter 126 which may also be referred to as a network interface,
20 internal modem 124, sound adapter 128, IO controller 132 to which may be connected a keyboard 140 and mouse 138, or other suitable input device such as a trackball or tablet, as well as external printer 134, and/or any number of external devices such as external modems, tape storage drives, or disk drives 136. One or more host bus adapters 114 may be connected to the system bus 116. To host bus adapter 114 may optionally be connected one or more storage devices such as disk drives 112 (removable or fixed), floppy drives 110, tape drives 108, digital versatile disk DVD drives 106, and compact disk CD ROM drives 104. The storage devices may operate in read-only mode and / or in read-write mode. The computer system may optionally include multiple central processing units 118, or multiple banks of memory 122.
30 Arrows 142 in Figure 1 indicate the interconnection of internal components of the computer system. The arrows are illustrative only and do not specify exact connection architecture.

Software for accessing and processing the one or more of the nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, and the polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 (such as sequence comparison software, analysis software as well as search tools, annotation tools, and modeling tools etc.) may reside in main memory 122 during execution.

In one embodiment, the computer system further comprises a sequence comparison software for comparing the nucleic acid codes of a query sequence stored on a computer readable medium to a subject sequence which is also stored on a computer readable medium; or for comparing the polypeptide code of a query sequence stored on a computer readable medium to a subject sequence which is also stored on computer readable medium. A "sequence comparison software" refers to one or more programs that are implemented on the computer system to compare nucleotide and/or protein sequences with other nucleotide and/or sequences stored within the data storage means. The design of one example of a sequence comparison software is provided in Figures 2A, 2B, 2C and 2D.

The sequence comparison software will typically employ one or more specialized comparator algorithms. Protein and/or nucleic acid sequence similarities may be evaluated using any of the variety of sequence comparator algorithms and programs known in the art. Such algorithms and programs include, but are no way limited to, TBLASTN, BLASTN, BLASTP, FASTA, TFASTA, CLUSTAL, HMMER, MAST, or other suitable algorithm known to those skilled in the art. (Pearson and Lipman, 1988, *Proc. Natl. Acad. Sci USA* 85(8): 2444-2448; Altschul *et al*, 1990, *J. Mol. Biol.* 215(3):403-410; Thompson *et al.*, 1994, *Nucleic Acids Res.* 22(2):4673-4680; Higgins *et al.*, 1996, *Methods Enzymol.* 266:383-402; Altschul *et al.*, 1990, *J. Mol. Biol.* 215(3):403-410; Altschul *et al.*, 1993, *Nature Genetics* 3:266-272; Eddy S.R., *Bioinformatics* 14:755-763, 1998; Bailey TL *et al*, *J Steroid Biochem Mol Biol* 1997 May;62(1):29-44). One example of a comparator algorithm is illustrated in Figure 3. Sequence comparator algorithms identified in this specification are particularly contemplated for use in this aspect of the invention.

The sequence comparison software will typically employ one or more specialized analyzer algorithms. One example of an analyzer algorithm is illustrated in Figure 4. Any appropriate analyzer algorithm can be used to evaluate similarities,

determined by the comparator algorithm, between a query sequence and a subject sequence (referred to herein as a query/subject pair). Based on context specific rules, the annotation of a subject sequence may be assigned to the query sequence. A skilled artisan can readily determine the selection of an appropriate analyzer algorithm and appropriate context specific rules. Analyzer algorithms identified elsewhere in this specification are particularly contemplated for use in this aspect of the invention.

Figures 2A, 2B, 2C and 2D together provide a flowchart of one example of a sequence comparison software for comparing query sequences to a subject sequence. The software determines if a gene or set of genes represented by their nucleotide
10 sequence, polypeptide sequence or other representation (the query sequence) is significantly similar to the one or more of the nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and the corresponding polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 of the invention (the subject sequence). The software may be implemented in the C or C++ programming language, Java, Perl or other suitable programming language known to a person skilled in the art.

One or more query sequence(s) are accessed by the program by means of input from the user 210, accessing a database 208 or opening a text file 206 as illustrated in the query initialization subprocess (Figure 2A). The query initialization
20 subprocess allows one or more query sequence(s) to be loaded into computer memory 122, or under control of the program stored on a disk drive 112 or other storage device in the form of a query sequence array 216. The query array 216 is one or more query nucleotide or polypeptide sequences accompanied by some appropriate identifiers.

A dataset is accessed by the program by means of input from the user 228, accessing a database 226, or opening a text file 224 as illustrated in the subject datasource initialization subprocess (Figure 2B). The subject data source initialization process refers to the method by which a reference dataset containing one or more sequence selected from the nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and the corresponding polypeptide codes
30 of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 is loaded into computer memory 122, or under control of the program stored on a disk drive 112 or other storage device in the form of a subject array 234. The subject array

234 comprises one or more subject nucleotide or polypeptide sequences accompanied by some appropriate identifiers.

The comparison subprocess of Figure 2C illustrates a process by which the comparator algorithm 238 is invoked by the software for pairwise comparisons between query elements in the query sequence array 216, and subject elements in the subject array 234. The "comparator algorithm" of Figure 2C refers to the pair-wise comparisons between a query sequence and subject sequence, *i.e.* a query/subject pair from their respective arrays 216, 234. Comparator algorithm 238 may be any algorithm that acts on a query/subject pair, including but not limited to homology
10 algorithms such as BLAST, Smith Waterman, Fasta, or statistical representation/probabilistic algorithms such as Markov models exemplified by HMMER, or other suitable algorithm known to one skilled in the art. Suitable algorithms would generally require a query/subject pair as input and return a score (an indication of likeness between the query and subject), usually through the use of appropriate statistical methods such as Karlin Altschul statistics used in BLAST, Forward or Viterbi algorithms used in Markov models, or other suitable statistics known to those skilled in the art.

The sequence comparison software of Figure 2C also comprises a means of analysis of the results of the pair-wise comparisons performed by the comparator
20 algorithm 238. The "analysis subprocess" of Figure 2C is a process by which the analyzer algorithm 244 is invoked by the software. The "analyzer algorithm" refers to a process by which annotation of a subject is assigned to the query based on query/subject similarity as determined by the comparator algorithm 238 according to context-specific rules coded into the program or dynamically loaded at runtime. Context-specific rules are what the program uses to determine if the annotation of the subject can be assigned to the query given the context of the comparison. These rules allow the software to qualify the overall meaning of the results of the comparator algorithm 238.

In one embodiment, context-specific rules may state that for a set of query
30 sequences to be considered representative of a rosaramicin biosynthetic locus, the comparator algorithm 238 must determine that the set of query sequences contains at least five query sequences that show a statistical similarity to a subject sequence corresponding to the polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18,

20, 22, 24, 26, 28, 30, 32, 34, 36, 38. Of course preferred context specific rules may specify a wide variety of thresholds for identifying rosaramicin biosynthetic genes or rosaramicin-producing organisms without departing from the scope of the invention. Some thresholds contemplate that at least one query sequence in the set of query sequences show a statistical similarity to the nucleic acid code corresponding to 5, 6, 7, 8 or more of the polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38. Other context specific rules set the level of homology required in each of the group and may be set at 70%, 80%, 85%, 90%, 95% or 98% in regards to any one or more of the subject sequences.

10 In another embodiment context-specific rules may state that for a query sequence to be considered indicative of a macrolide, the comparator algorithm 238 must determine that the query sequence shows a statistical similarity to subject sequences corresponding to a nucleic acid sequence code for a polypeptide of SEQ ID NO: 10, 12, 14, 16 and 18, polypeptides having at least 75% homology to a polypeptide of SEQ ID NOS: 10, 12, 14, 16 and 18 and fragment comprising at least 400 consecutive amino acids of the polypeptides of SEQ ID NOS: 10, 12, 14, 16 and 18. Of course preferred context specific rules may specify a wide variety of thresholds for identifying a macrolide protein without departing from the scope of the invention. Some context specific rules set level of homology required of the query sequence at
20 70%, 80%, 85%, 90%, 95% or 98%.

Thus, the analysis subprocess may be employed in conjunction with any other context specific rules and may be adapted to suit different embodiments. The principal function of the analyzer algorithm 244 is to assign meaning or a diagnosis to a query or set of queries based on context specific rules that are application specific and may be changed without altering the overall role of the analyzer algorithm 244.

Finally the sequence comparison software of Figure 2 comprises a means of returning of the results of the comparisons by the comparator algorithm 238 and analyzed by the analyzer algorithm 244 to the user or process that requested the comparison or comparisons. The "display / report subprocess" of Figure 2D is the
30 process by which the results of the comparisons by the comparator algorithm 238 and analyses by the analyzer algorithm 244 are returned to the user or process that requested the comparison or comparisons. The results 240, 246 may be written to a file 252, displayed in some user interface such as a console, custom graphical

interface, web interface, or other suitable implementation specific interface, or uploaded to some database such as a relational database, or other suitable implementation specific database. Once the results have been returned to the user or process that requested the comparison or comparisons the program exits.

The principle of the sequence comparison software of Figure 2 is to receive or load a query or queries, receive or load a reference dataset, then run a pair-wise comparison by means of the comparator algorithm 238, then evaluate the results using an analyzer algorithm 244 to arrive at a determination if the query or queries bear significant similarity to the reference sequences, and finally return the results to the user or calling program or process.

Figure 3 is a flow diagram illustrating one embodiment of comparator algorithm 238 process in a computer for determining whether two sequences are homologous. The comparator algorithm receives a query/subject pair for comparison, performs an appropriate comparison, and returns the pair along with a calculated degree of similarity.

Referring to Figure 3, the comparison is initiated at the beginning of sequences 304. A match of (x) characters is attempted 306 where (x) is a user specified number. If a match is not found the query sequence is advanced 316 by one character with respect to the subject, and if the end of the query has not been reached 318 another match of (x) characters is attempted 306. Thus if no match has been found the query is incrementally advanced in entirety past the initial position of the subject. Once the end of the query is reached 318, the subject pointer is advanced by 1 character and the query pointer is set to the beginning of the query 320. If the end of the subject has been reached and still no matches have been found a null homology result score is assigned 324 and the algorithm returns the pair of sequences along with a null score to the calling process or program. The algorithm then exits 326. If instead a match is found 308, an extension of the matched region is attempted 310 and the match is analyzed statistically 312. The extension may be unidirectional or bidirectional. The algorithm continues in a loop extending the matched region and computing the homology score, giving penalties for mismatches taking into consideration that given the chemical properties of the amino acid side chains (in the case of comparisons) not all mismatches are equal. For example a mismatch of a lysine with an arginine both of which have basic side chains receive a lesser penalty

than a mismatch between lysine and glutamate which has an acidic side chain. The extension loop stops once the accumulated penalty exceeds some user specified value, or of the end of either sequence is reached 312. The maximal score is stored 314, and the query sequence is advanced 316 by one character with respect to the subject, and if the end of the query has not been reached 318 another match of (x) characters is attempted 306. The process continues until the entire length of the subject has been evaluated for matches to the entire length of the query. All individual scores and alignments are stored 314 by the algorithm and an overall score is computed 324 and stored. The algorithm returns the pair of sequences along with

10 local and global scores to the calling process or program. The algorithm then exits 326.

One example of comparator algorithm 238 algorithm may be represented in pseudocode as follows:

```

INPUT:      Q[m]: query, m is the length
           S[n]: subject, n is the length
           x:  x is the size of a segment
START:
for each i in [1,n] do
  for each j in [1,m] do
20    if ( j + x - 1 ) <= m and ( i + x - 1 ) <= n then
      if Q(j, j+x-1) = S(i, i+x-1) then
        k=1;
        while Q(j, j+x-1+k) = S(i, i+x-1+k) do
          k++;
        Store highest local homology
      Compute overall homology score
      Return local and overall homology scores
END.
```

30

The comparator algorithm 238 may be written for use on nucleotide sequences, in which case the scoring scheme would be implemented so as to calculate scores and apply penalties based on the chemical nature of nucleotides. The comparator algorithm 238 may also provide for the presence of gaps in the scoring method for nucleotide or polypeptide sequences.

BLAST is one implementation of the comparator algorithm 238. HMMER is another implementation of the comparator algorithm 238 based on Markov model analysis. In a HMMER implementation a query sequence would be compared to a mathematical model representative of a subject sequence or sequences rather than

40 using sequence homology.

Figure 4 is a flow diagram illustrating an analyzer algorithm 244 process for detecting the presence of a rosaramicin biosynthetic locus. The analyzer algorithm of Figure 4 may be used in the process by which the annotation of a subject is assigned to the query based on their similarity as determined by the comparator algorithm 238 and according to context-specific rules coded into the program or dynamically loaded at runtime. Context sensitive rules are what determines if the annotation of the subject can be assigned to the query given the context of the comparison. Context specific rules set the thresholds for determining the level and quality of similarity that would be accepted in the process of evaluating matched pairs.

10 The analyzer algorithm 244 receives as its input an array of pairs that had been matched by the comparator algorithm 238. The array consists of at least a query identifier, a subject identifier and the associated value of the measure of their similarity. To determine if a group of query sequences includes sequences diagnostic of a rosaramicin biosynthetic gene cluster, a reference or diagnostic array 406 is generated by accessing a data source and retrieving rosaramicin specific information 404 relating to nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and the corresponding polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38. Diagnostic array 406 consists at least of subject identifiers and their associated annotation. Annotation may include
20 reference to the protein families ABCC, DATF, GTFA, MTFA, MTRA, NBPA, OXRB, OXRC, OXRH, PKSH, REGM, REGS, SURA and TESA. Annotation may also include information regarding presence in loci of a specific structural class or may include previously computed matches to other databases, for example databases of motifs.

 Once the algorithm has successfully generated or received the two necessary arrays 402, 406, and holds in memory any context specific rules, each matched pair as determined by the comparator algorithm 238 can be evaluated. The algorithm will perform an evaluation 408 of each matched pair and based on the context specific rules confirm or fail to confirm the match as valid 410. In cases of successful confirmation of the match 410 the annotation of the subject is assigned to
30 the query. Results of each comparison are stored 412. The loop ends when the end of the query / subject array is reached. Once all query / subject pairs have been evaluated against one or more of the nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and the polypeptide codes of SEQ

ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 in the subject array, a final determination can be made if the query set of ORFs represents a rosaramicin locus 416. The algorithm then returns the overall diagnosis and an array of characterized query / subject pairs along with supporting evidence to the calling program or process and then terminates 418.

The analyzer algorithm 244 may be configured to dynamically load different diagnostic arrays and context specific rules. It may be used for example in the comparison of query/subject pairs with diagnostic subjects for other biosynthetic pathways, such as macrolide biosynthetic pathways.

10 Thus one embodiment of the present invention is a computer readable medium having stored thereon a sequence selected from the group consisting of a nucleic acid code of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and a polypeptide code of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38. Another aspect of the present invention is a computer readable medium having recorded thereon one or more nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, preferably at least 2, 5, 10, 15, or 20 nucleic acid codes of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39. Another aspect of the invention is a computer readable medium having recorded thereon one or more of the polypeptide
20 codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, preferably at least 2, 5, 10, 15 or 20 polypeptide codes of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38.

Another embodiment of the present invention is a computer system comprising a processor and a data storage device wherein said data storage device has stored thereon a reference sequence selected from the group consisting of a nucleic acid code of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and a polypeptide code of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38.

30 Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random Access Memory (RAM), or Read Only Memory (ROM) as well as other types of media known to those skilled in the art.

The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples.

EXAMPLE 1: Identification and sequencing of a rosaramicin biosynthetic locus in *Micromonospora carbonacea* var. *aurantiaca* NRRL 2997

Micromonospora carbonacea var. *aurantiaca* NRRL 2997 was obtained from the Agricultural Research Service collection (National Center for Agricultural Utilization Research, 1815 N. University Street, Peoria, Illinois 61604) and cultured using
10 standard microbiological techniques (Kieser *et al.*, *supra*). This organism was propagated on oatmeal agar medium at 28 degrees Celsius for several days. For isolation of high molecular weight genomic DNA, cell mass from three freshly grown, near confluent 100 mm petri dishes was used. The cell mass was collected by gentle scraping with a plastic spatula. Residual agar medium was removed by repeated washes with STE buffer (75 mM NaCl; 20 mM Tris-HCl, pH 8.0; 25 mM EDTA). High molecular weight DNA was isolated by established protocols (Kieser *et al.* *supra*) and its integrity was verified by field inversion gel electrophoresis (FIGE) using the preset program number 6 of the FIGE MAPPER™ power supply (BIORAD). This high
20 molecular weight genomic DNA served for the preparation of a small size fragment genomic sampling library (GSL), as well as a large size fragment cluster identification library (CIL). Both libraries contained randomly generated *M. carbonacea* genomic DNA fragments and, therefore, are representative of the entire genome of this organism.

For the generation of the GSL library, genomic DNA was randomly sheared by sonication. DNA fragments having a size range between 1.5 and 3 kb were fractionated on a agarose gel and isolated using standard molecular biology techniques (Sambrook *et al.*, *supra*). The ends of the obtained DNA fragments were repaired using T4 DNA polymerase (Roche) as described by the supplier. This
30 enzyme creates DNA fragments with blunt ends that can be subsequently cloned into an appropriate vector. The repaired DNA fragments were subcloned into a derivative of pBluescript SK+ vector (Stratagene) which does not allow transcription of cloned DNA fragments. This vector was selected as it contains a convenient polylinker region surrounded by sequences corresponding to universal sequencing primers such as T3,

T7, SK, and KS (Stratagene). The unique *EcoRV* restriction site found in the polylinker region was used as it allows insertion of blunt-end DNA fragments. Ligation of the inserts, use of the ligation products to transform *E. coli* DH10B (Invitrogen) host and selection for recombinant clones were performed as previously described (Sambrook et al., *supra*). Plasmid DNA carrying the *M. carbonacea* genomic DNA fragments was extracted by the alkaline lysis method (Sambrook et al., *supra*) and the insert size of 1.5 to 3 kb was confirmed by electrophoresis on agarose gels. Using this procedure, a library of small size random genomic DNA fragments is generated that covers the entire genome of the studied microorganism. The number of individual clones that can be generated is infinite but only a small number is further analyzed to sample the microorganism's genome.

A CIL library was constructed from the *M. carbonacea* high molecular weight genomic DNA using the SuperCos-1 cosmid vector (Stratagene™). The cosmid arms were prepared as specified by the manufacturer. The high molecular weight DNA was subjected to partial digestion at 37 degrees Celsius with approximately one unit of *Sau3AI* restriction enzyme (New England Biolabs) per 100 micrograms of DNA in the buffer supplied by the manufacturer. This procedure generates random fragments of DNA ranging from the initial undigested size of the DNA to short fragments of which the length is dependent upon the frequency of the enzyme DNA recognition site in the genome and the extent of the DNA digestion by the enzyme. At various timepoints, aliquots of the digestion were transferred to new microfuge tubes and the enzyme was inactivated by adding a final concentration of 10 mM EDTA and 0.1% SDS. Aliquots judged by FIGE analysis to contain a significant fraction of DNA in the desired size range (30-50kb) were pooled, extracted with phenol/chloroform (1:1 vol:vol), and pelleted by ethanol precipitation. The 5' ends of *Sau3AI* DNA fragments were dephosphorylated using alkaline phosphatase (Roche) according to the manufacturer's specifications at 37 degrees Celcius for 30 min. The phosphatase was heat inactivated at 70 degrees Celcius for 10 min and the DNA was extracted with phenol/chloroform (1:1 vol:vol), pelleted by ethanol precipitation, and resuspended in sterile water. The dephosphorylated *Sau3AI* DNA fragments were then ligated overnight at room temperature to the SuperCos-1 cosmid arms in a reaction containing approximately four-fold molar excess SuperCos-1 cosmid arms. The ligation products were packaged using Gigapack® III XL packaging extracts (Stratagene™) according to

the manufacturer's specifications. The CIL library consisted of 864 isolated cosmid clones in *E. coli* DH10B (Invitrogen). These clones were picked and inoculated into nine 96-well microtiter plates containing LB broth (per liter of water: 10.0 g NaCl; 10.0 g tryptone; 5.0 g yeast extract) which were grown overnight and then adjusted to contain a final concentration of 25% glycerol. These microtiter plates were stored at – 80 degrees Celcius and served as glycerol stocks of the CIL library. Duplicate microtiter plates were arrayed onto nylon membranes as follows. Cultures grown on microtiter plates were concentrated by pelleting and resuspending in a small volume of LB broth. A 3 X 3 grid (96-pin) was arrayed onto nylon membranes. These
10 membranes representing the complete CIL library were then layered onto LB agar and incubated overnight at 37 degrees Celcius to allow the colonies to grow. The membranes were layered onto filter paper pre-soaked with 0.5 N NaOH/1.5 M NaCl for 10 min to denature the DNA and then neutralized by transferring onto filter paper pre-soaked with 0.5 M Tris (pH 8)/1.5 M NaCl for 10 min. Cell debris was gently scraped off with a plastic spatula and the DNA was crosslinked onto the membranes by UV irradiation using a GS GENE LINKER™ UV Chamber (BIORAD). Considering an average size of 8 Mb for an actinomycete genome and an average size of 35 kb of genomic insert in the CIL library, this library represents roughly a 4-fold coverage of the microorganism's entire genome.

20 The GSL library was analyzed by sequence determination of the cloned genomic DNA inserts. The universal primers KS or T7, referred to as forward (F) primers, were used to initiate polymerization of labeled DNA. Extension of at least 700 bp from the priming site can be routinely achieved using the TF, BDT v2.0 sequencing kit as specified by the supplier (Applied Biosystems). Sequence analysis of the small genomic DNA fragments (Genomic Sequence Tags, GSTs) was performed using a 3700 ABI capillary electrophoresis DNA sequencer (Applied Biosystems). The average length of the DNA sequence reads was ~700 bp. Further analysis of the obtained GSTs was performed by sequence homology comparison to various protein sequence databases. The DNA sequences of the obtained GSTs were translated into
30 amino acid sequences and compared to the National Center for Biotechnology Information (NCBI) nonredundant protein database and the proprietary Ecopia natural product biosynthetic gene Decipher™ database using previously described algorithms (Altschul et al., *supra*). Sequence similarity with known proteins of defined function in

the database enables one to make predictions on the function of the partial protein that is encoded by the translated GST.

A total of 437 *M. carbonacea* GSTs were generated using the forward sequencing primer and analyzed by sequence comparison using the Blast algorithm (Altschul et al., *supra*). Sequence alignments displaying an E value of at least e^{-5} were considered as significantly homologous and retained for further evaluation. GSTs showing similarity to a gene of interest can be at this point selected and used to identify larger segments of genomic DNA from the CIL library that include the gene(s) of interest. Polyketide natural products are often synthesized by type I polyketide
10 synthases (PKSs). Several forward GST reads were identified as portions of PKS genes. For example, one such GST encoded an internal portion of a PKS acyl transferase (AT) domain in the antisense orientation relative to the sequencing primer. The GSL clone from which this GST was obtained was also sequenced using the reverse sequencing primer and was found to encode the N-terminal portion of a PKS ketosynthase (KS) domain in the sense orientation relative to the sequencing primer. Based on the sequence of the forward read of this GSL clone, a 20mer oligonucleotide was designed for use as a probe to identify and isolate CIL clones which harbored the sequences of interest.

Hybridization oligonucleotide probes were radiolabeled with P^{32} using T4
20 polynucleotide kinase (New England Biolabs) in 15 microliter reactions containing 5 picomoles of oligonucleotide and 6.6 picomoles of $[\gamma\text{-}P^{32}]\text{ATP}$ in the kinase reaction buffer supplied by the manufacturer. After 1 hour at 37 degrees Celcius, the kinase reaction was terminated by the addition of EDTA to a final concentration of 5 mM. The specific activity of the radiolabeled oligonucleotide probes was estimated using a Model 3 Geiger counter (Ludlum Measurements Inc., Sweetwater, Texas) with a built-in integrator feature. The radiolabeled oligonucleotide probes were heat-denatured by incubation at 85 degrees Celcius for 10 minutes and quick-cooled in an ice bath immediately prior to use.

The CIL library membranes were pretreated by incubation for at least 2
30 hours at 42 degrees Celcius in Prehyb Solution (6X SSC; 20mM NaH_2PO_4 ; 5X Denhardt's; 0.4% SDS; 0.1 mg/ml sonicated, denatured salmon sperm DNA) using a hybridization oven with gentle rotation. The membranes were then placed in Hyb Solution (6X SSC; 20mM NaH_2PO_4 ; 0.4% SDS; 0.1 mg/ml sonicated, denatured

salmon sperm DNA) containing 1×10^6 cpm/ml of radiolabeled oligonucleotide probe and incubated overnight at 42 degrees Celcius using a hybridization oven with gentle rotation. The next day, the membranes were washed with Wash Buffer (6X SSC, 0.1% SDS) for 45 minutes each at 46, 48, and 50 degrees Celcius using a hybridization oven with gentle rotation. The membranes were then exposed to X-ray film to visualize and identify the positive cosmid clones. Positive clones were identified, cosmid DNA was extracted from 30 ml cultures using the alkaline lysis method (Sambrook et al., *supra*) and the inserts were entirely sequenced using a shotgun sequencing approach (Fleischmann et al., *Science*, 269:496-512).

10 Sequencing reads were assembled using the Phred-Phrap™ algorithm (University of Washington, Seattle, USA) recreating the entire DNA sequence of the cosmid insert. Reiterations of hybridizations of the CIL library with probes derived from the ends of the original cosmid allow indefinite extension of sequence information on both sides of the original cosmid sequence until the complete sought-after gene cluster is obtained. Three overlapping cosmid clones that were either directly identified by the original oligonucleotide probe (derived from the GSL clone) or by probes derived from the ends of the original cosmids have been completely sequenced to provide over 60 Kb of genetic information. Subsequently, the forward and reverse reads of the GSL clone from which the original oligonucleotide probe was derived were mapped to a
20 region of the rosaramicin biosynthetic locus that encodes a portion of the PKS gene identified herein as ORF 7, more specifically nucleotides encoding amino acids 1531 to 2416 approximately. This corresponds to a GSL clone with an insert size of approximately 2.6kb, in good agreement with the selected size range of 1.5- 3kb described above. The sequence of these cosmids and analysis of the proteins encoded by them undoubtedly demonstrated that the gene cluster obtained was indeed responsible for the production of a glycosylated macrolide consistent with the known structure of rosaramicin, which was not previously reported to be produced by *M. carbonacea* var *aurantiaca* NRRL 2997.

30 Example 2: Genes and proteins involved in biosynthesis of rosaramicin

The rosaramicin locus includes the 60196 base pairs provided in SEQ ID NO: 1 and contains the 19 ORFs provided SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39. More than 19 kilobases of DNA sequence were

analyzed on each side of the rosaramicin locus and these regions contain primary metabolic genes. The accompanying sequence listing provides the nucleotide sequence of the 19 ORFs regulating the biosynthesis of rosaramicin and the corresponding deduced polypeptides, wherein ORF 1 (SEQ ID NO: 3) represents the polynucleotide drawn from residues 1 to 1683 (sense strand) of SEQ ID NO: 1; ORF 2 (SEQ ID NO: 5) represents the polynucleotide drawn from residues 2522 to 1728 (antisense strand) of SEQ ID NO: 1; ORF 3 (SEQ ID NO: 7) represents the polynucleotide drawn from residues 3861 to 2629 (antisense strand) of SEQ ID NO: 1; ORF 4 (SEQ ID NO: 9) represents the polynucleotide drawn from residues 4365 to 5573 (sense strand) of SEQ ID NO: 1; ORF 5 (SEQ ID NO: 11) represents the polynucleotide drawn from residues 5702 to 19117 (sense strand) of SEQ ID NO: 1; ORF 6 (SEQ ID NO: 13) represents the polynucleotide drawn from residues 19144 to 24921 (sense strand) of SEQ ID NO: 1; ORF 7 (SEQ ID NO: 15) represents the polynucleotide drawn from residues 24993 to 36230 (sense strand) of SEQ ID NO: 1; ORF 8 (SEQ ID NO: 17) represents the polynucleotide drawn from residues 36292 to 41016 (sense strand) of SEQ ID NO: 1; ORF 9 (SEQ ID NO: 19) represents the polynucleotide drawn from residues 41049 to 46403 (sense strand) of SEQ ID NO: 1; ORF 10 (SEQ ID NO: 21) represents the polynucleotide drawn from residues 46400 to 47794 (sense strand) of SEQ ID NO: 1; ORF 11 (SEQ ID NO: 23) represents the polynucleotide drawn from residues 47794 to 49083 (sense strand) of SEQ ID NO: 1; ORF 12 (SEQ ID NO: 25) represents the polynucleotide drawn from residues 49092 to 49814 (sense strand) of SEQ ID NO: 1; ORF 13 (SEQ ID NO: 27) represents the polynucleotide drawn from residues 49868 to 51226 (sense strand) of SEQ ID NO: 1; ORF 14 (SEQ ID NO: 29) represents the polynucleotide drawn from residues 51506 to 53416 (sense strand) of SEQ ID NO: 1; ORF 15 (SEQ ID NO: 31) represents the polynucleotide drawn from residues 54569 to 53358 (antisense strand) of SEQ ID NO: 1; ORF 16 (SEQ ID NO: 33) represents the polynucleotide drawn from residues 54897 to 56342 (sense strand) of SEQ ID NO: 33; ORF 17 (SEQ ID NO: 35) represents the polynucleotide drawn from residues 56408 to 57634 (sense strand) of SEQ ID NO: 1; ORF 18 (SEQ ID NO: 37) represents the polynucleotide drawn from residues 57657 to 59123 (sense strand) of SEQ ID NO: 1; ORF 19 (SEQ ID NO: 39) represents the polynucleotide drawn from residues 59363 to 60196 (sense strand) of SEQ ID NO: 1.

Some open reading frames listed herein initiate with non-standard initiation codons (e.g. GTG – Valine or CTG - Leucine) rather than the standard initiation codon ATG, namely ORFs 1, 6, 7, 10, 14 and 18. All ORFs are listed with the appropriate M, V or L amino acids at the amino-terminal position to indicate the specificity of the first codon of the ORF. It is expected, however, that in all cases the biosynthesized protein will contain a methionine residue, and more specifically a formylmethionine residue, at the amino terminal position, in keeping with the widely accepted principle that protein synthesis in bacteria initiates with methionine (formylmethionine) even when the encoding gene specifies a non-standard initiation codon (e.g. Stryer, Biochemistry 3rd edition, 1998, W.H. Freeman and Co., New York, pp. 752-754).

Three deposits, namely *E. coli* DH10B (O10CK) strain, *E. coli* DH10B (O10CF) strain and *E. coli* DH10B (O10CJ) strain each harbouring a cosmid clone of a partial biosynthetic locus for rosaramicin from *Micromonospora carbonacea* subsp. *aurantiaca* have been deposited with the International Depositary Authority of Canada, Bureau of Microbiology, Health Canada, 1015 Arlington Street, Winnipeg, Manitoba, Canada R3E 3R2 on July 10, 2002 and were assigned deposit accession number IDAC 100702-1, 100702-2 and 100702-3 respectively. The *E. coli* strain deposits are referred to herein as “the deposited strains”.

The cosmids harbored in the deposited strains comprise a complete biosynthetic locus for rosaramicin. The sequence of the polynucleotides comprised in the deposited strains, as well as the amino acid sequence of any polypeptide encoded thereby are controlling in the event of any conflict with any description of sequences herein.

The deposit of the deposited strains has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The deposited strains will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposited strains are provided merely as convenience to those skilled in the art and are not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112. A license may be required to make, use or sell the deposited strains, and compounds derived therefrom, and no such license is hereby granted.

The order and relative position of the 19 open reading frames and the corresponding polypeptides of the biosynthetic locus for rosaramicin are provided in

Figure 5. The arrows represent the orientation of the ORFs of the rosaramicin biosynthetic locus. The top line in Figure 5 provides a scale in kilobase pairs. The black bars depict the part of the locus covered by each of the deposited cosmids O10CK, O10CF and O10CJ.

In order to identify the function of the genes in the rosaramicin locus, SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 were compared, using the BLASTP version 2.2.1 algorithm with the default parameters, to sequences in the National Center for Biotechnology Information (NCBI) nonredundant protein database and the DECIPHER™ database of microbial genes, pathways and natural products (Ecopia BioSciences Inc. St.-Laurent, QC, Canada).

The accession numbers of the top GenBank hits of this BLAST analysis are presented in Table 2 along with the corresponding E value. The E value relates the expected number of chance alignments with an alignment score at least equal to the observed alignment score. An E value of 0.00 indicates a perfect homolog or nearly perfect homolog. The E values are calculated as described in Altschul *et al.* J. Mol. Biol., October 5; 215(3) 403-10. The E value assists in the determination of whether two sequences display sufficient similarity to justify an inference of homology.

Table 2

ORF no.	Family	#aa	GenBank homology	probability	% identity	% similarity	proposed function of GenBank match
1	ABCC	560	S25202,550aa	1e-144	284/542 (52.4%)	337/542 (62.18%)	spiramycin-resistance protein, Streptomyces ambofaciens
			AAC32027.1,551aa	1e-142	275/543 (50.64%)	341/543 (62.8%)	carbomycin resistance protein, Streptomyces thermotolerans
			S67863,569aa	1e-126	259/553 (46.84%)	323/553 (58.41%)	oleandomycin resistance protein, Streptomyces antibioticus
2	TESA	264	BAB69315.1,255aa	3e-52	109/232 (46.98%)	131/232 (56.47%)	thioesterase, Streptomyces avermitilis
			S49055,253aa	9e-48	110/238 (46.22%)	129/238 (54.2%)	thioesterase, Streptomyces fradiae
			AAC01736.1,254aa	8e-47	101/229 (44.1%)	125/229 (54.59%)	thioesterase, Amycolatopsis mediterranei
3	OXRC	410	S49051,417aa	1e-168	282/394 (71.57%)	335/394 (85.03%)	cytochrome P450, Streptomyces fradiae
			BAB83674.1,402aa	2e-73	161/384 (41.93%)	222/384 (57.81%)	cytochrome P450 monooxygenase, Streptomyces virginiae
			B40634,412aa	3e-73	161/393 (40.97%)	218/393 (55.47%)	erythromycin monooxygenase, Saccharopolyspora erythraea
4	OXRC	402	AAA92553.1,407aa	4e-95	187/400 (46.75%)	242/400 (60.5%)	cytochrome P450, Streptomyces antibioticus
			AAD28449.1,410aa	8e-85	166/399 (41.6%)	229/399 (57.39%)	cytochrome P450, Streptomyces lavendulae
			S51594,397aa	9e-81	161/398 (40.45%)	223/398 (56.03%)	cytochrome P450, Micromonospora griseorubida
5	PKSH	4471	BAA76543.1,4290aa	0.0	1488/2961 (50.25%)	1726/2961 (58.29%)	polyketide synthase, Micromonospora griseorubida
			T17409,4613aa	0.0	1405/3027 (46.42%)	1696/3027 (56.03%)	polyketide synthase, Streptomyces venezuelae
			AAB66504.1,4472aa	0.0	1275/2459 (51.85%)	1483/2459 (60.31%)	tylactone polyketide synthase, Streptomyces fradiae
6	PKSH	1925	AAB66505.1,1864aa	0.0	923/1883 (49.02%)	1087/1883 (57.73%)	tylactone polyketide synthase, Streptomyces fradiae
			AAC46025.1,1839aa	0.0	847/1897 (44.65%)	1027/1897 (54.14%)	polyketide synthase, Streptomyces caelestis
			AAK73514.1,10917aa	0.0	834/1810 (46.08%)	1015/1810 (56.08%)	AmphC polyketide synthase, Streptomyces nodosus

7	PKSH	3745	AAB66506.1,3729aa	0.0	2004/3782 (52.99%)	2283/3782 (60.36%)	tylactone polyketide synthase, Streptomyces fradiae
			T17410,3739aa	0.0	1417/3144 (45.07%)	1730/3144 (55.03%)	polyketide synthase, Streptomyces venezuelae
			AAG13918.1,3562aa	0.0	1269/3054 (41.55%)	1588/3054 (52%)	6-deoxyerythronolide B synthase, Micromonospora megalomicea
8	PKSH	1574	AAB66507.1,1611aa	0.0	756/1521 (49.7%)	869/1521 (57.13%)	tylactone polyketide synthase, Streptomyces fradiae
			BAB69192.1,6146aa	0.0	723/1527 (47.35%)	839/1527 (54.94%)	polyketide synthase, Streptomyces avermitilis
			AAG23266.1,3170aa	0.0	721/1556 (46.34%)	847/1556 (54.43%)	polyketide synthase, Saccharopolyspora spinosa
9	PKSH	1784	AAB66508.1,1841aa	0.0	956/1832 (52.18%)	1079/1832 (58.9%)	tylactone polyketide synthase, Streptomyces fradiae
			BAB69307.1,3352aa	0.0	726/1546 (46.96%)	855/1546 (55.3%)	polyketide synthase, Streptomyces avermitilis
			AAF71766.1,9477aa	0.0	725/1614 (44.92%)	869/1614 (53.84%)	Nysl polyketide synthase, Streptomyces noursei
10	OXRB	464	CAA57471.1,423aa	6e-55	161/422 (38.15%)	182/422 (43.13%)	NDP hexose 3,4 isomerase, Streptomyces fradiae
			AAF73456.1,443aa	3e-36	138/430 (32.09%)	166/430 (38.6%)	AknT, Streptomyces galliaeus
			AAD15266.1,438aa	1e-27	127/421 (30.17%)	150/421 (35.63%)	dnrQ, Streptomyces peucetius
11	GTFA	429	CAA57472.2,452aa	1e-148	258/421 (61.28%)	315/421 (74.82%)	glycosyltransferase, Streptomyces fradiae
			AAC68677.1,426aa	1e-126	237/426 (55.63%)	289/426 (67.84%)	glycosyl transferase, Streptomyces venezuelae
			CAA05642.1,426aa	1e-125	230/424 (54.25%)	293/424 (69.1%)	glycosyltransferase, Streptomyces antibioticus
12	MTFA	240	CAA57473.2,255aa	8e-74	132/234 (56.41%)	155/234 (66.24%)	N-methyltransferase, Streptomyces fradiae
			CAA05643.1,246aa	2e-72	130/233 (55.79%)	150/233 (64.38%)	methyltransferase, Streptomyces antibioticus
			AAC68678.1,237aa	1e-67	125/235 (53.19%)	147/235 (62.55%)	N,N-dimethyltransferase, Streptomyces venezuelae
13	OXRH	452	NP_630556.1,447aa	0.0	344/440 (78.18%)	386/440 (87.73%)	crotonyl CoA reductase, Streptomyces coelicolor
			S72400,447aa	0.0	342/440 (77.73%)	381/440 (86.59%)	trans-2-enoyl-CoA reductase, Streptomyces collinus
			CAA57474.2,449aa	0.0	344/427 (80.56%)	375/427 (87.82%)	crotonyl CoA reductase, Streptomyces fradiae
14	REGS	636	S25203,604aa	1e-92	219/583 (37.56%)	290/583 (49.74%)	smrR regulator, Streptomyces ambifaciens
			NP_625307.1,634aa	1e-63	189/612 (30.88%)	264/612 (43.14%)	hypothetical protein, Streptomyces coelicolor
			NP_630273.1,569aa	4e-12	69/211 (32.7%)	93/211 (44.08%)	putative regulatory protein, Streptomyces coelicolor

15	REGM	403	AAF29380.1,430aa JC2032,387aa	1e-99 1e-78	192/386 (49.74%) 170/387 (43.93%)	236/386 (61.14%) 219/387 (56.59%)	TyIR regulator, Streptomyces fradiae regulatory protein, Streptomyces sp.
16	NBPA	481	NP_629920.1,497aa NP_217241.1,495aa NP_337300.1,556aa	1e-160 1e-135 1e-135	293/431 (67.98%) 261/438 (59.59%) 261/438 (59.59%)	331/431 (76.8%) 303/438 (69.18%) 303/438 (69.18%)	hypothetical protein, Streptomyces coelicolor putative HflX GTP-binding protein, Mycobacterium tuberculosis GTP-binding protein, Mycobacterium tuberculosis
17	DATF	408	T51108,393aa AAC68684.1,415aa AAB84075.1,401aa	1e-149 1e-144 1e-140	259/392 (66.07%) 259/392 (66.07%) 245/396 (61.87%)	304/392 (77.55%) 303/392 (77.3%) 298/396 (75.25%)	dehydratase, Streptomyces antibioticus 4-dehydrase, Streptomyces venezuelae EryCIV, Saccharopolyspora erythraea
18	SURA	488	AAC68683.1,485aa T51109,485aa CAA72085.1,489aa	0.0 1e-179 1e-168	313/480 (65.21%) 310/479 (64.72%) 294/474 (62.03%)	367/480 (76.46%) 369/479 (77.04%) 352/474 (74.26%)	putative reductase, Streptomyces venezuelae probable reductase, Streptomyces antibioticus EryCV, Saccharopolyspora erythraea
19	MTRA	277	T17407,322aa S28985,278aa P43433,311aa	5e-95 2e-73 4e-66	167/251 (66.53%) 143/249 (57.43%) 130/247 (52.63%)	198/251 (78.88%) 173/249 (69.48%) 164/247 (66.4%)	rRNA methyltransferase, Streptomyces venezuelae lincomycin resistance protein, Streptomyces lincolnensis mycinamycin resistance protein, Micromonospora griseorubida

Example 3: Formation of rosaramicin

The chemical structure of rosaramicin is a 16-membered macrolide having an epoxide, an aldehyde and a deoxyamino sugar. The rosaramicin locus includes five polyketide synthase (PKS) Type I genes. ORF 5 represents a PKS Type I gene having a domain arrangement of KS-AT-ACP-KS-AT-KR-ACP-KS-AT-DH-KR-ACP. ORF 6 represents a PKS Type 1 gene having a domain arrangement of KS-AT-DH-KR-ACP. ORF 7 represents a PKS Type I gene having a domain arrangement of KS-AT-KR-ACP-KS-AT-DH-ER-KR-ACP. ORF 8 represents a PKS Type I gene having a domain arrangement of KS-AT-KR-ACP. ORF 9 represents a PKS Type I gene having a domain arrangement of KS-AT-KR-ACP-Te.

While not intending to be limited to any particular mode of action or biosynthetic scheme, the gene products of the invention can explain the synthesis of rosaramicin. ORFs 5, 6, 7, 8, and 9 constitute a polyketide synthase system that assembles the core polyketide precursor of rosaramicin. Figure 6 highlights schematically the series of reactions catalyzed by this polyketide synthase system based on the correlation between the deduced domain architecture and the polyketide core of rosaramicin. Type I PKS domains and the reactions they carry out are well known to those skilled in the art and well documented in the literature, see for example, Hopwood (1997) Chem. Rev. Vol 97 pp. 2465-2497.

Figure 7 depicts a proposed biochemical pathway involving the OXRB, DATF, SURA, MTFA gene products for the formation of the deoxyamino sugar. This sugar is transferred to the core polyketide precursor of rosaramicin by the GTFA gene product. Also depicted in Figure 7 are the oxidation reactions carried out by two cytochrome P450 monooxygenases OXRC1 and OXRC2, referring to ORFs 3 and 4, respectively. OXRC1 is expected to catalyze the formation of an aldehyde while OXRC2 is expected to catalyze the formation of an epoxide. While Figure 7 proposes one scheme in regard to timing of the glycosylation and oxidation reactions catalyzed by the GTFA, OXRC1 and OXRC2, the invention does not reside in the actual timing and order of the reactions, which may be different than that depicted in Figure 7.

Figures 8 to 10 are amino acid alignments comparing the rosaramicin PKS domains. The domains which occur only once in the rosaramicin PKS, namely the enoylreductase (ER) and thioesterase (Te) domains, are compared to prototypical domains from the erythromycin PKS system (DEBS). Where applicable, key active site

residues and motifs for the various polyketide synthase domains as described in Kakavas et al. (1997) J. Bacteriol. Vol 179 pp. 7515-7522 are indicated in Figures 8 to 14. In each of the clustal alignments a line above the alignment is used to mark strongly conserved positions. In addition, three characters, namely * (asterisk), : (colon) and . (period) are used, wherein "*" indicates positions which have a single, fully conserved residue; ":" indicates that one of the following strong groups is fully conserved: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, and FYW; and "." Indicates that one of the following weaker groups is fully conserved: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, FVLIM, and HFY.

10 Of particular relevance with respect to PKS domain function, the KS domain in the loading module (ORF5|KS1) contains a Gln (Q) in place of the active site Cys (C) residue (Figure 8) and that the KR domain of the first module of ORF7 (ORF7|KR1) contains several amino acid substitutions in the key cofactor-binding motif (Figure 12). Figure 15 shows the high degree of overall homology between ethylmalonyl-CoA-specific AT domains from the tylosin PKS (TYLO) and the niddamycin PKS (NIDD) and the second AT domain of rosaramicin ORF 7. This high degree of homology is indicative of their shared substrate specificity.

 REGS and REGM are involved in regulation of gene expression. ABCC, a membrane transport protein and MTRA, a rRNA methyltransferase, are involved in
20 resistance to and/or export of rosaramicin. The TESA gene product represents a free-standing thioesterase enzyme that is expected to play a "proofreading" role in the assembly of the rosaramicin core polyketide precursor. The OXRH gene product represents a crotonyl CoA reductase that is involved in the formation of the acyl-CoA precursor used by the loading module of ORF 5 and/or the second module of ORF 7. The step involving crotonyl CoA reductase, *ie.* the OXRH gene product, is expected to be a rate-limiting step in the biosynthesis of rosaramicin (Stassi D.L. *et al.*, *Proc Natl Acad Sci* 95(13), 7305-9, June 23, 1998) and it is expected that increasing the levels of the OXRH enzyme will have a beneficial effect on the yield of rosaramicin. The NBPA gene product is a nucleotide binding protein (*i.e.*, contains a GTP/ATP binding motif)
30 and is expected to activate a sugar by tethering it to a nucleotide, usually TTP. Therefore, the NBPA gene product is expected to be involved in the first step in the pathway leading to the formation of the deoxyamino sugar of rosaramicin.

Example 4: Fermentation of *Micromonospora carbonacea*, *aurantiaca* and detection of rosaramicin:

Micromonospora carbonacea aurantiaca NRRL 2997 was cultured on a 30ml media A plate (glucose 1.0%, dextrin 4.0%, sucrose 1.5%, casein enzymatic hydrolysate 1.0%, MgSO_4 0.1%, CaCO_3 0.2%, and agar 2.2g/100ml) at 30°C for 14 days. The cells and agar were added to 25 ml of 95 % ethanol and incubated at room temperature for 2h under agitation. The ethanol phase was collected and the extraction step was repeated under the same conditions. The ethanol was evaporated from the pooled extracts and the residue was freeze-dried. The residue was then resuspended in 1.0 ml of water.

SPE of extracts: The C-18 solid phase column (Burdick & Jackson) was conditioned before use by sequential washing with 3ml of distilled water, 3 ml of methanol, and finally 3 ml of distilled water. The residue previously resuspended in 1.0 ml of water was loaded on the conditioned solid phase extraction system (SPE). Following passage of the sample through the SPE column washes were performed first, with 5 ml of water to remove polar materials, and then with 70% acetone and 30% methanol to elute a secondary metabolite-containing fraction which was then freeze-dried. This organic fraction was dissolved in 300ul of 50% acetonitrile—distilled water.

Chemical analysis: Chemical analysis of the organic fraction from the SPE column was performed by HPLC-ES-MS (Waters, ZQ systems). The extracts (50.0 ul) were separated on a C18 symmetry analytical column (2.1X150mm) with HPLC 2690 system (Waters) using a 60-min linear gradient from 30% acetonitrile-5mM ammonium acetate to 95% acetonitrile-5mM ammonium acetate at a flow rate of 150ul min⁻¹. UV and visible light absorption spectra (220 to 500nm) were acquired with a PDA (Waters) by using the column effluents prior to their analysis by ES-MS. The electrospray source was switched between positive ion mode and negative ion mode at 0.3 s intervals to acquire both positive and negative ion spectra. The cone voltage was 25.0 V. The capillary was maintained at 3.0 V. The source temperature was kept at 100°C. The desolvation temperature was kept at 400°C and the desolvation gas flow was 479 litre.h⁻¹. The data collection and analysis were performed with MassLynx V3.5 program (Waters).

Figure 8 is a HPLC-ES-MS analysis of rosaramicin showing a UV spectra at a retention time of 24.4 minutes and a MS spectra showing a molecular ion consistent with rosaramicin at retention time 24.4 minutes (mass of 582.57 $[M+H]^+$).

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and the accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

10 It is further to be understood that all sizes and all molecular weight or mass values are approximate, and are provided for description.

Patents, patent publications, procedures and publications cited throughout this application are incorporated herein in their entirety for all purposes.

CLAIMS:

1. An isolated, purified or enriched nucleic acid comprising a nucleic acid sequence selected from the group consisting of:
 - (a) a nucleic acid of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39;
 - (b) a nucleic acid encoding a polypeptide of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38;
 - (c) a nucleic acid having at least 75% homology to a nucleic acid of (a) or (b) as determined by analysis with BLASTN version 2.0 with the default parameters;
 - (d) a nucleic acid complementary to a nucleic acid of (a), (b) or (c).
2. An isolated, purified or enriched nucleic acid capable of hybridizing to a nucleic acid of claim 1 under conditions of high stringency.
3. An isolated, purified or enriched nucleic acid capable of hybridizing to the nucleic acid of claim 1 under conditions of moderate stringency.
- 20 4. An isolated, purified or enriched nucleic acid comprising the sequence of at least two nucleic acids of claim 1.
5. An isolated, purified or enriched nucleic acid comprising the sequence of at least three nucleic acids of claim 1.
6. An isolated, purified or enriched nucleic acid comprising a nucleic acid that hybridizes under stringent conditions to any one of rosaramicin open reading frames (ORFs) 1 to 19 (SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39) and can substitute for the ORF to which it specifically hybridizes to direct the
30 synthesis of a rosaramicin compound or analogue.
7. An isolated, purified or enriched nucleic acid that hybridizes under stringent conditions to any one of rosaramicin ORFs 1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 14 or 15 (SEQ

ID NOS: 3, 5, 9, 11, 13, 15, 17, 19, 21, 25, 29, and 31) and can substitute for the ORF to which it specifically hybridizes to direct the synthesis of a rosaramicin compound or analogue.

8. An isolated nucleic acid of claim 1 that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide selected from the group comprising SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20.

9. The isolated nucleic acid of claim 1 that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide selected from the group consisting of SEQ ID NOS: 22, 24, 26, 28, 30, 32, 34, 36, 38.

10. An isolated gene cluster comprising ORFs encoding polypeptides sufficient to direct the synthesis of a rosaramicin compound or analogue.

11. The isolated gene cluster of claim 10 wherein the gene cluster is present in a bacterium.

12. The isolated gene cluster of claim 10 wherein the gene cluster contains a nucleic acid of any one of rosaramicin ORFs 1 to 19 (SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39) present in the E. coli strains DH10B having accession nos. IDAC 100702-1, 100702-2 and 100702-3.

13. An isolated polypeptide comprising a polypeptide sequence selected from any one of:

- (a) a polypeptide of any one of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38; and
- (b) a polypeptide which is at least 75% identical in amino acid sequence to a polypeptide of any one of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 as determined by analysis with BLASTP with the default parameters.

14. The isolated polypeptide of claim 13 wherein the polypeptide sequence selected from any one of:
- a) a polypeptide of any one of rosaramicin ORFs 1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 14 or 15 (SEQ ID NOS: 2, 4, 8, 10, 12, 14, 16, 18, 20, 24, 28 and 30); and
 - b) a polypeptide which is at least 75% identical in amino acid sequence to a polypeptide of any one of rosaramicin ORFs 1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 14 or 15 (SEQ ID NOS: 2, 4, 8, 10, 12, 14, 16, 18, 20, 24, 28 and 30) as determined by analysis with BLASTP with the default parameters.
- 10 15. A polypeptide comprising at least two polypeptides of claim 14.
16. A polypeptide comprising at least three polypeptides of claim 14.
17. A polypeptide comprising at least five or more polypeptides of claim 14.
18. An expression vector comprising a nucleic acid of claim 1.
19. A host cell transformed with an expression vector of claim 18.
- 20 20. The host cell of claim 19, wherein the cell is transformed with an exogenous nucleic acid comprising a gene cluster encoding polypeptides sufficient to direct the assembly of a rosaramicin compound or analogue.
21. A method of chemically modifying a biological molecule that is a substrate for a polypeptide encoded by a rosaramicin biosynthesis gene cluster, said method comprising contacting the biological molecule with a polypeptide of claim 13, wherein said polypeptide chemically modifies said biological molecule.
- 30 22. The method of chemically modifying a biological molecule that is a substrate for a polypeptide encoded by a rosaramicin biosynthesis gene cluster, said method comprising contacting the biological molecule with at least two different polypeptides of claim 13.

23. An isolated or purified antibody capable of specifically binding to a polypeptide having a sequence selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38.
24. A method of making a polypeptide having a sequence selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 comprising introducing a nucleic acid encoding said polypeptide, said nucleic acid being operably linked to a promoter, into a host cell.
- 10 25. A method of making a rosaramicin compound or analog comprising the step of providing a bacterium containing a gene cluster with sufficient genes to produce a rosaramicin compound of analogue and culturing the bacterium under conditions allowing for expression of the sufficient genes to produce a rosaramicin compound, wherein the gene cluster contains at least one nucleic acid of claim 1.
26. A method of making a rosaramicin compound or analog comprising culturing a *Micromonospora carbonacea* bacterium under conditions allowing for expression of rosaramicin ORFs 1 to 19 (SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39) present in the E. coli strains DH10B having accession nos. IDAC
20 100702-1, 100702-2 and 100702-3.
27. A computer readable medium having stored thereon a sequence selected from the group consisting of a nucleic acid code of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51 and a polypeptide code of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50.
28. A computer system comprising a processor and a data storage device wherein said data storage device has stored thereon a sequence selected from the group
30 consisting of a nucleic acid code of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51 and a polypeptide code of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50.

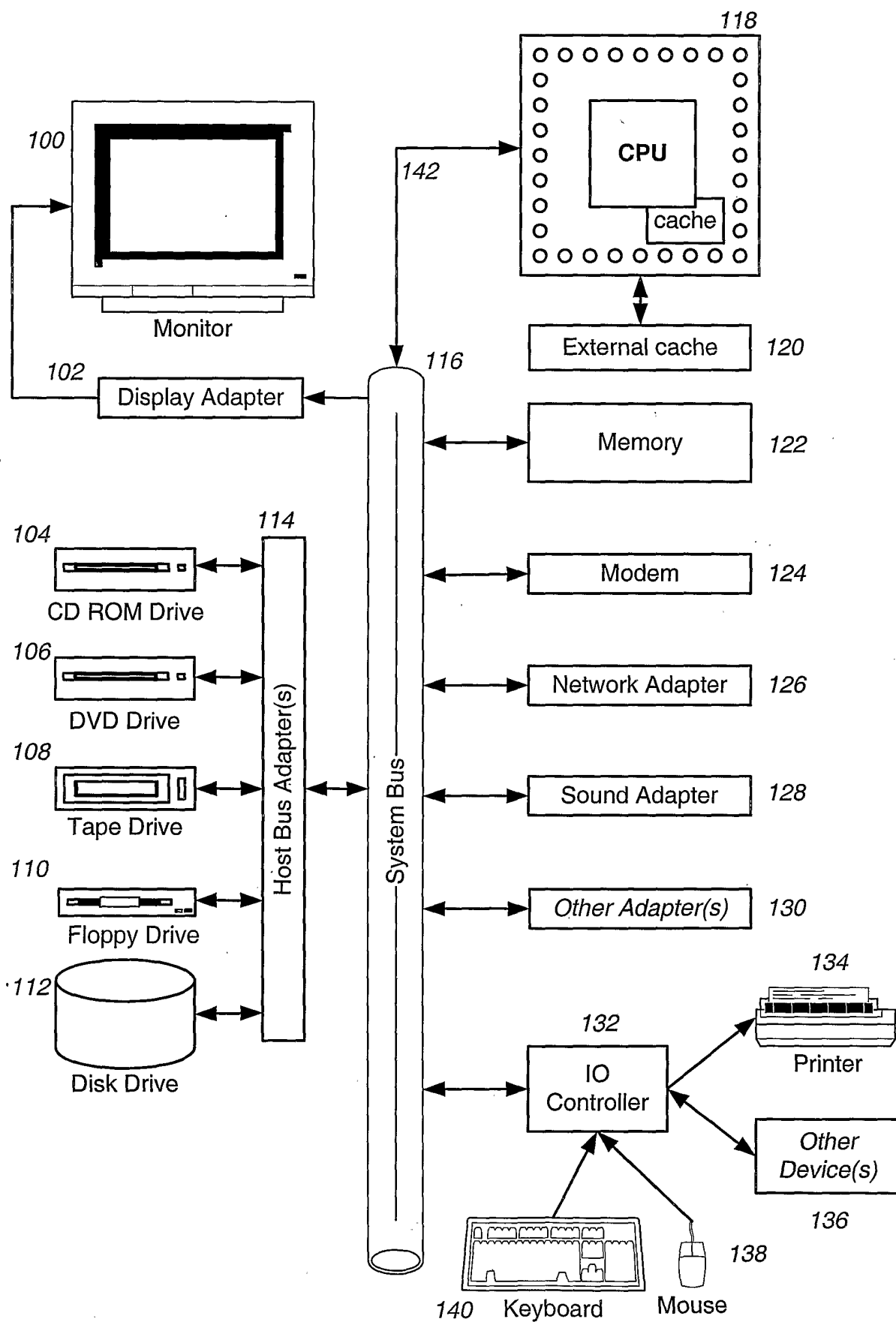


Figure 1

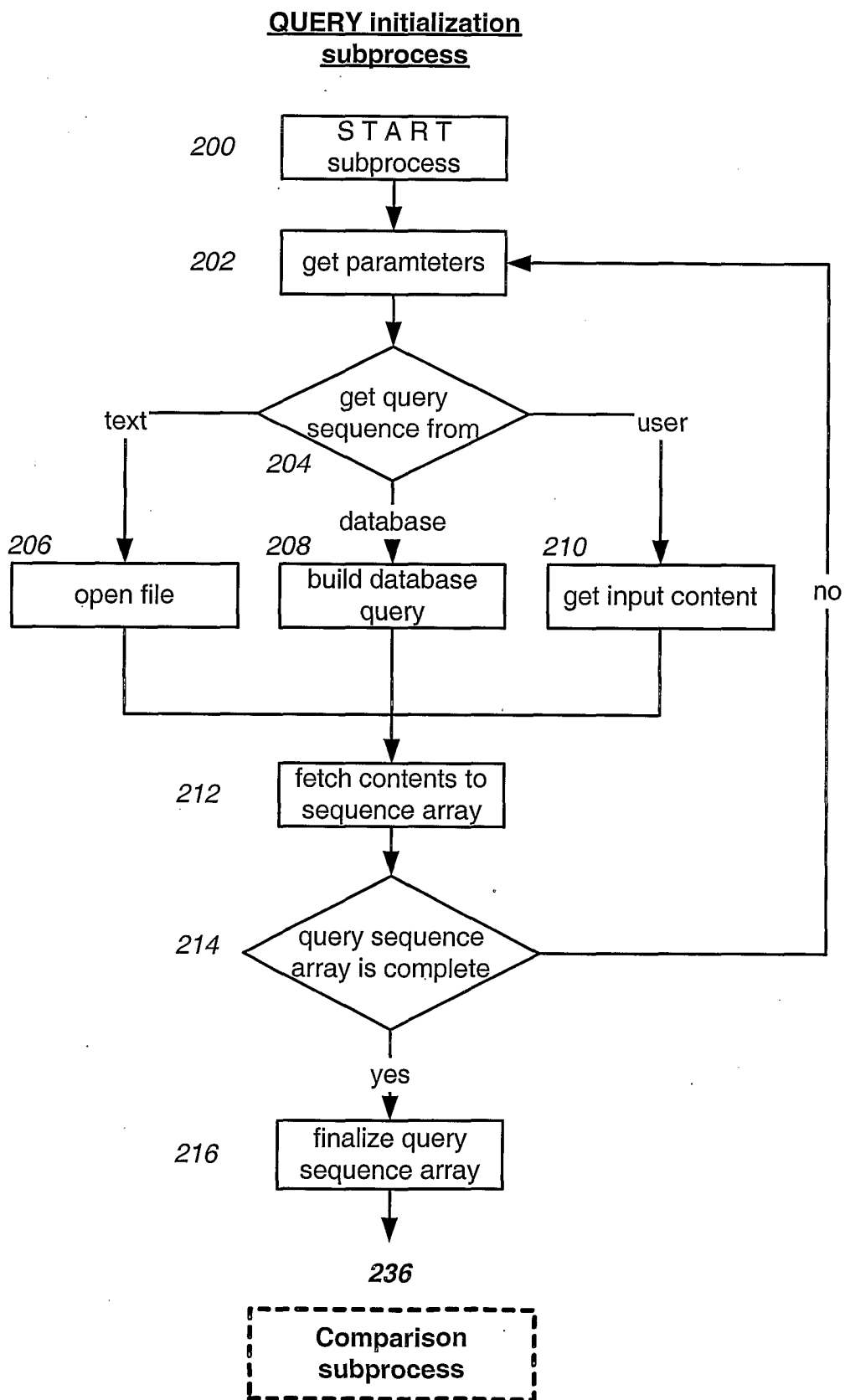


Figure 2-A

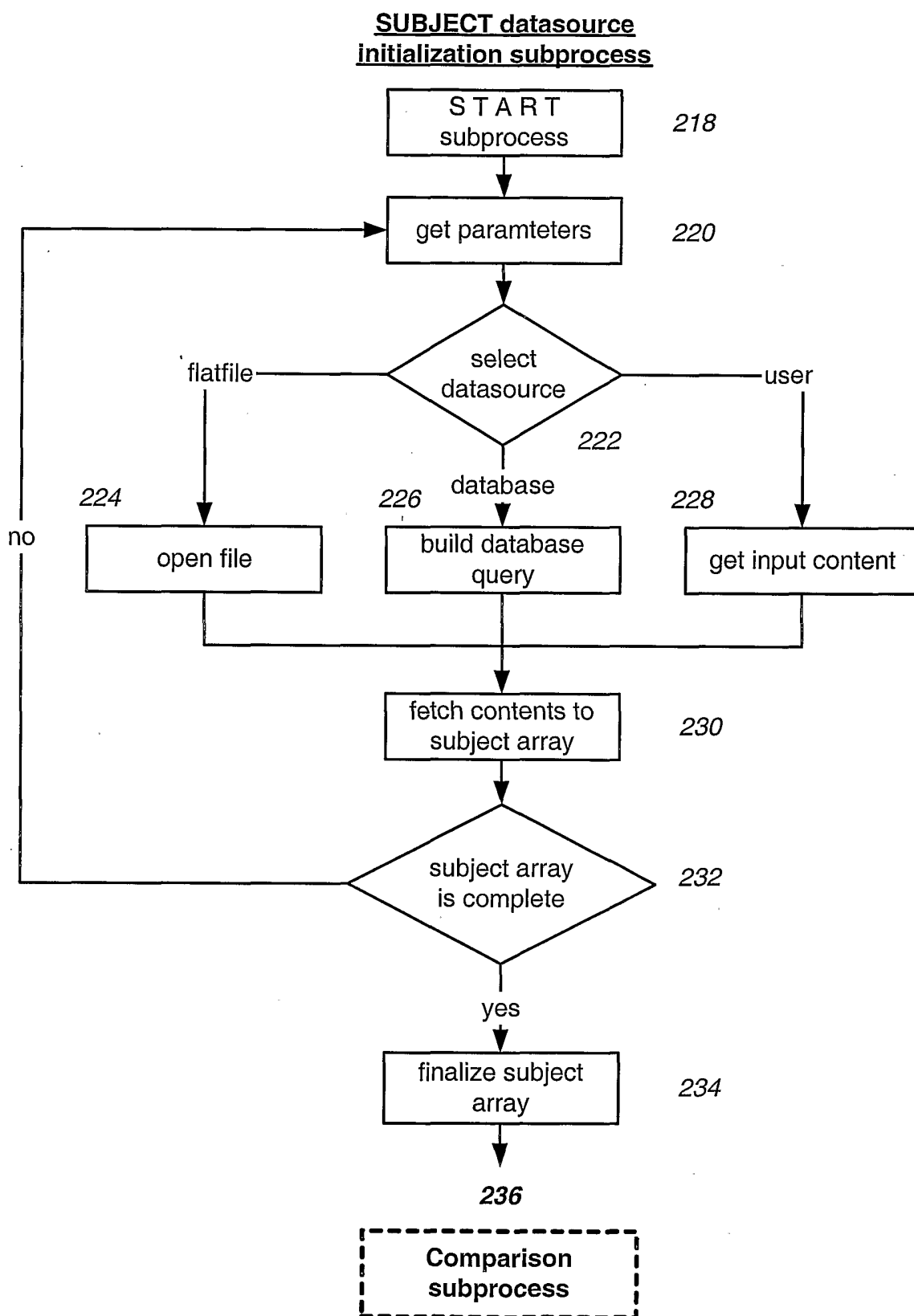


Figure 2-B

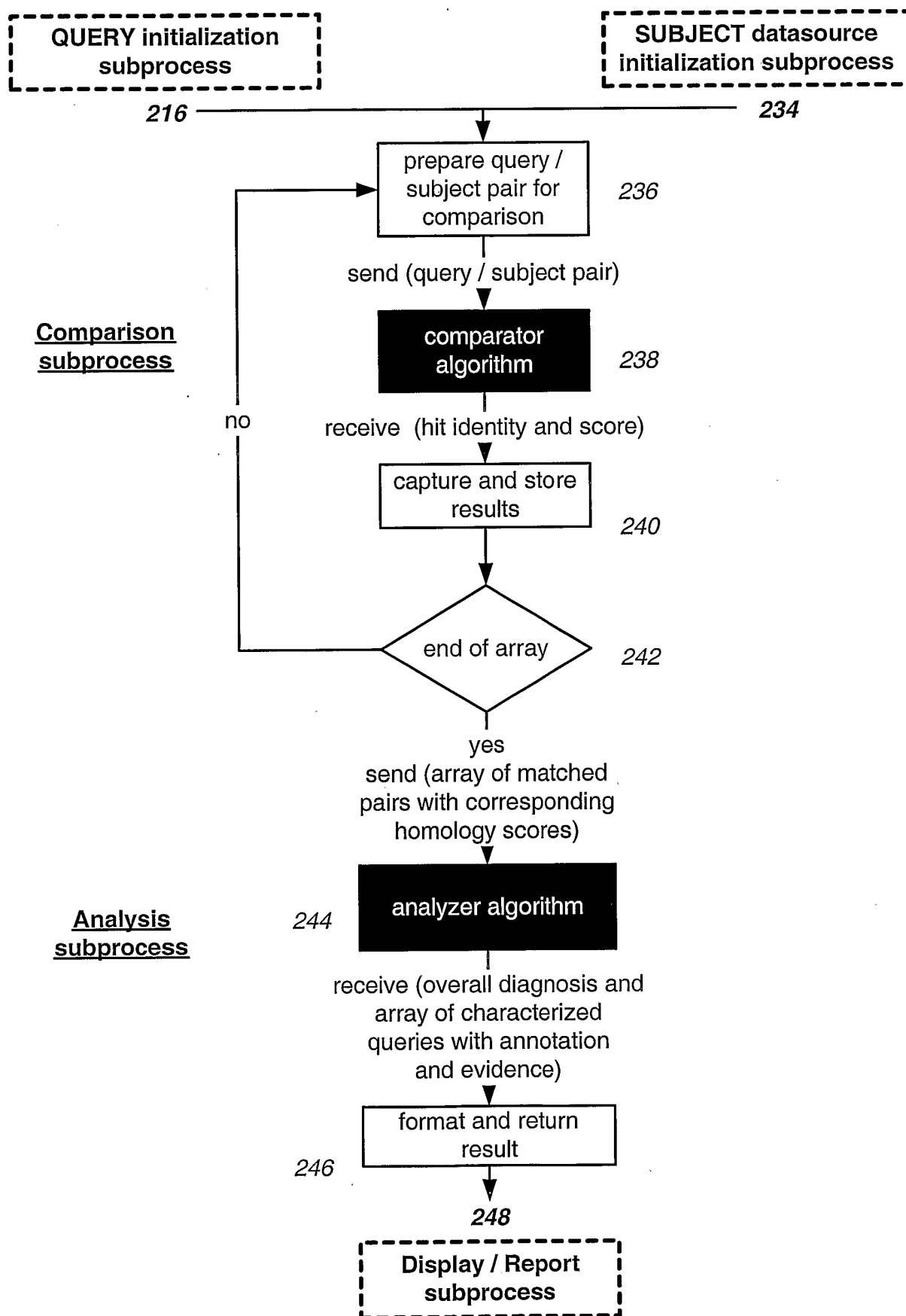


Figure 2-C

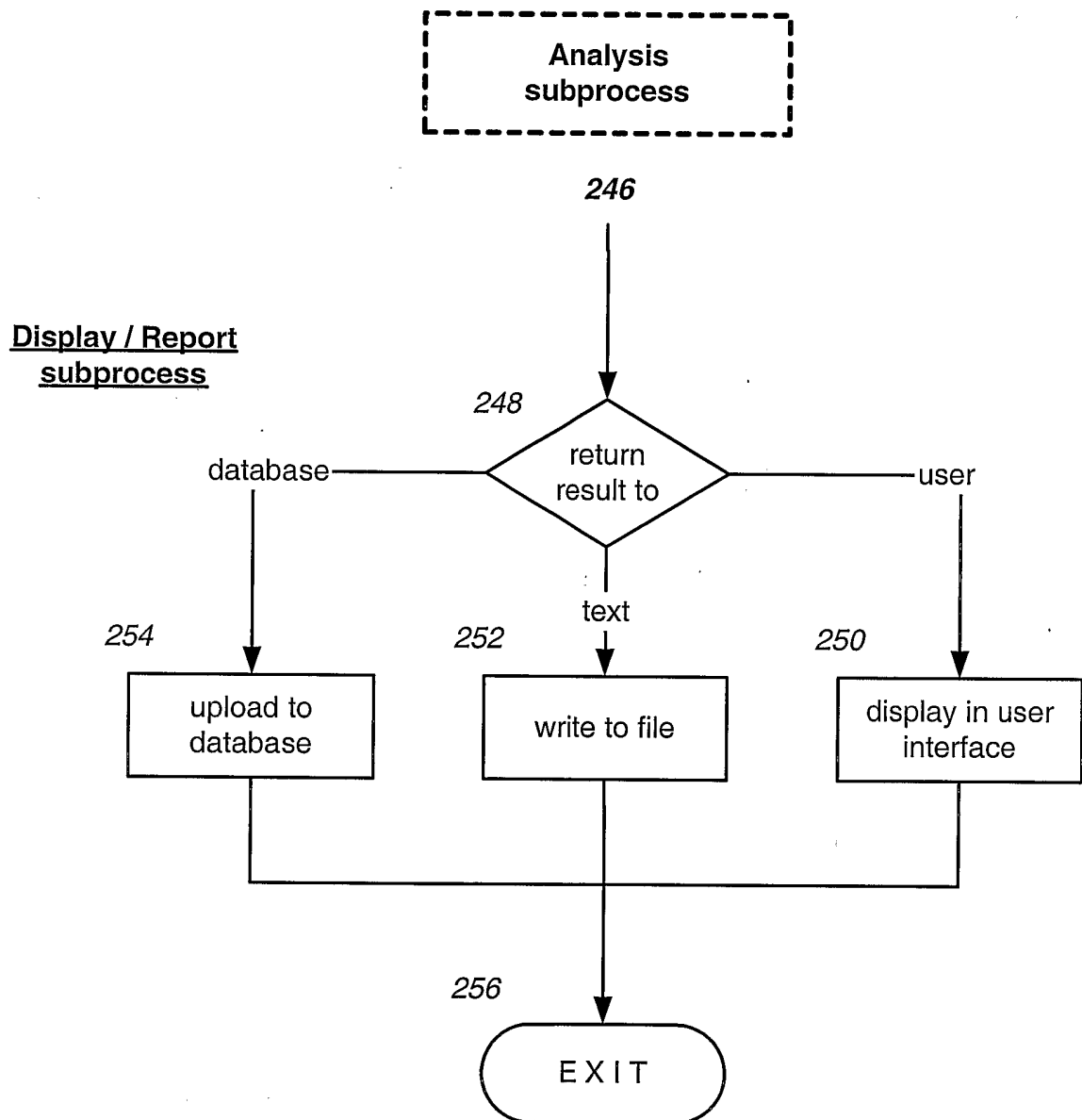


Figure 2-D

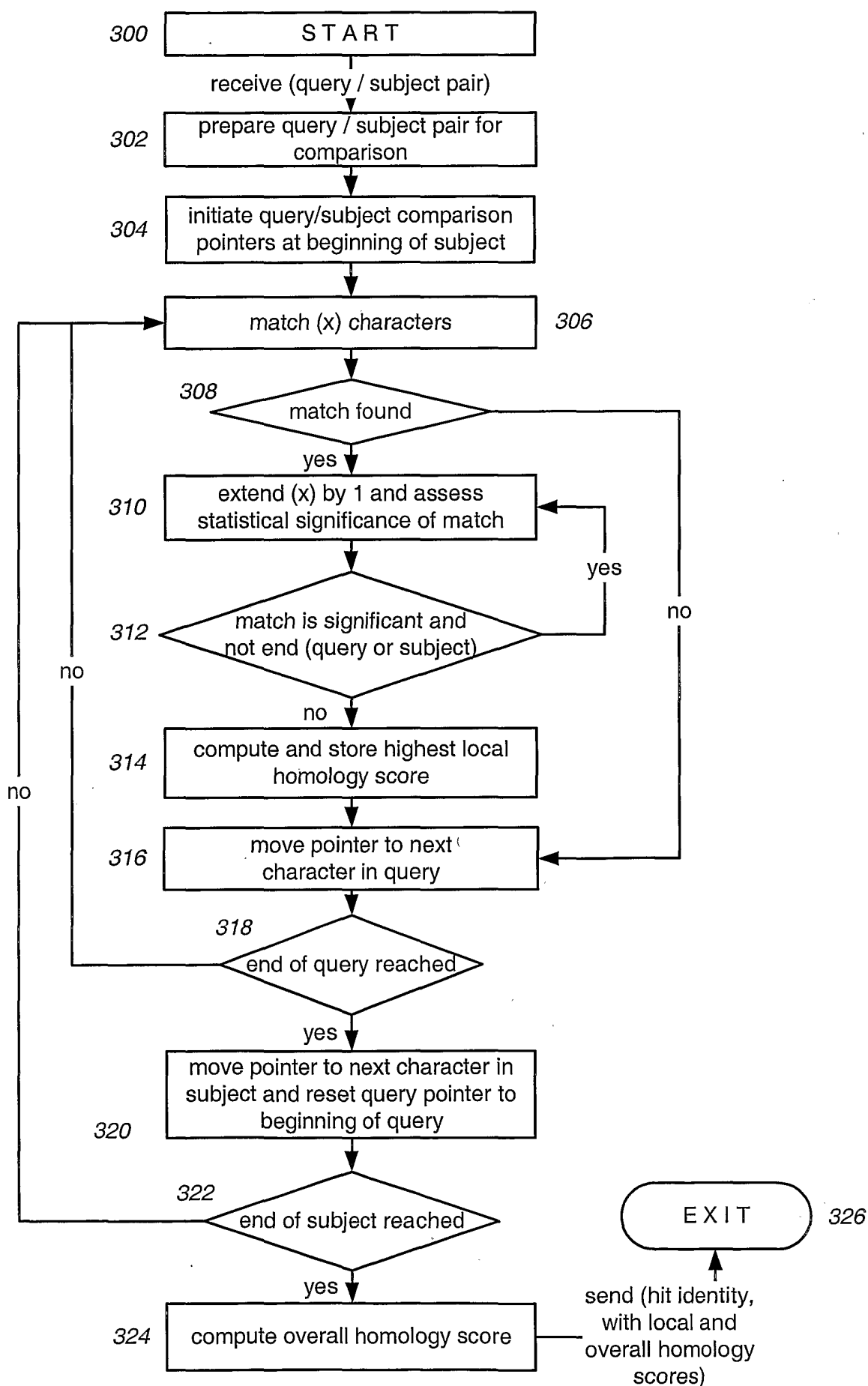


Figure 3
6/21

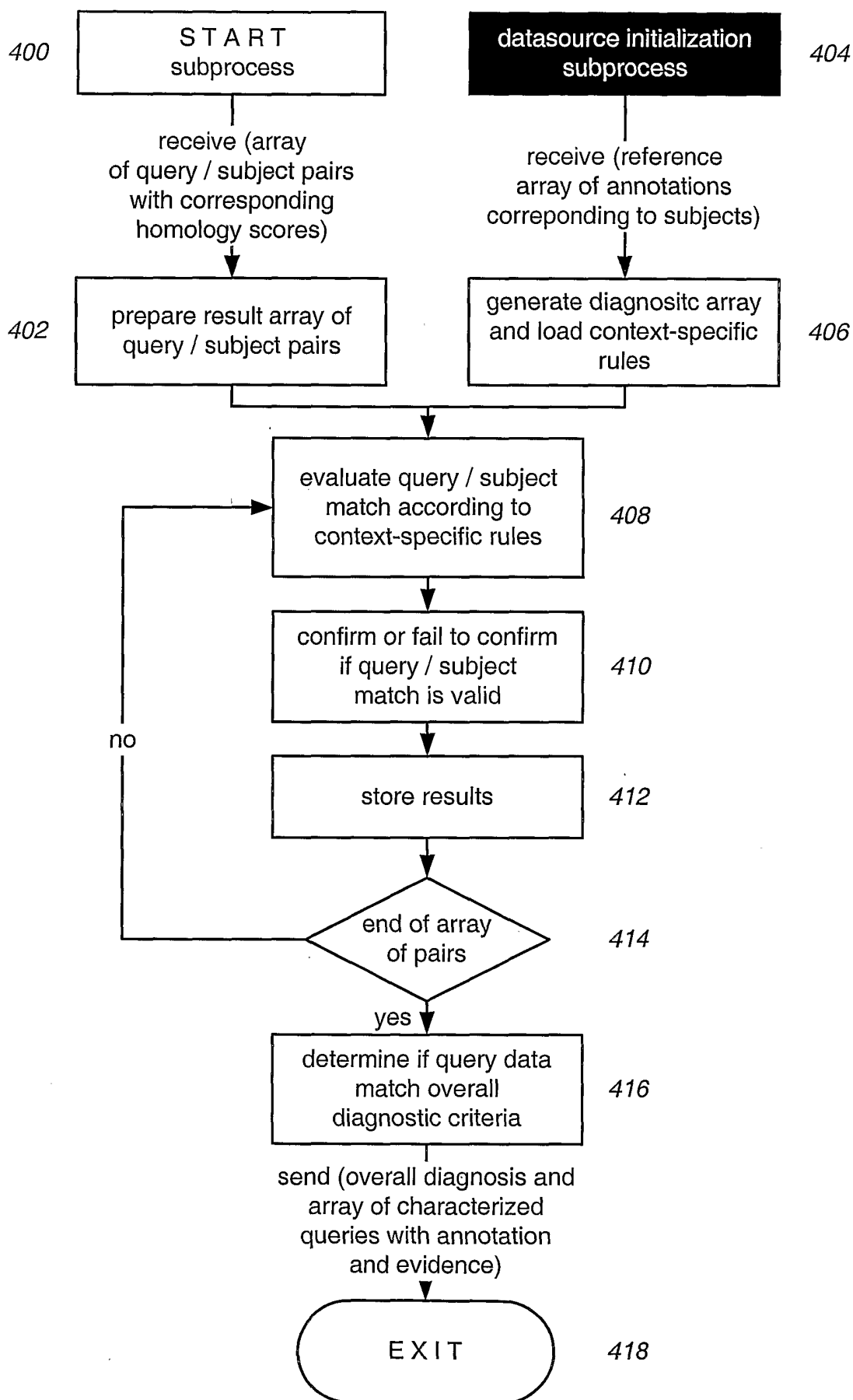


Figure 4

Figure 5

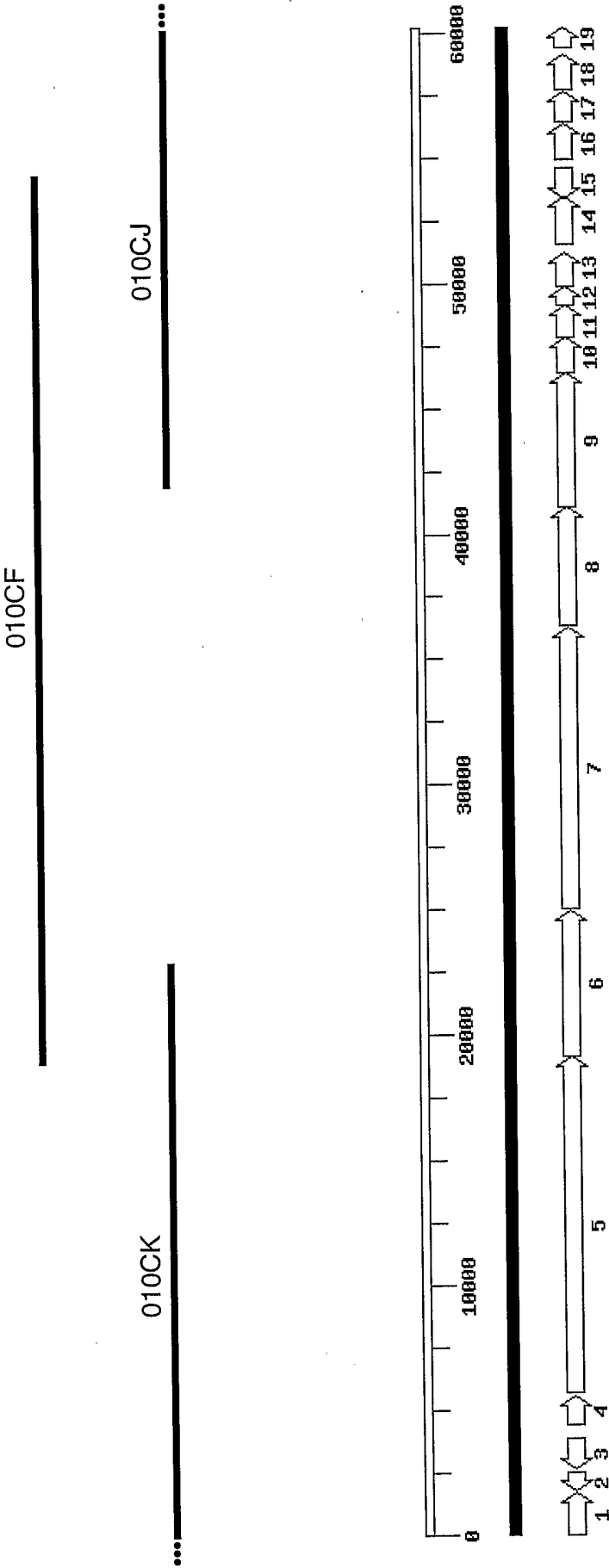


Figure 6

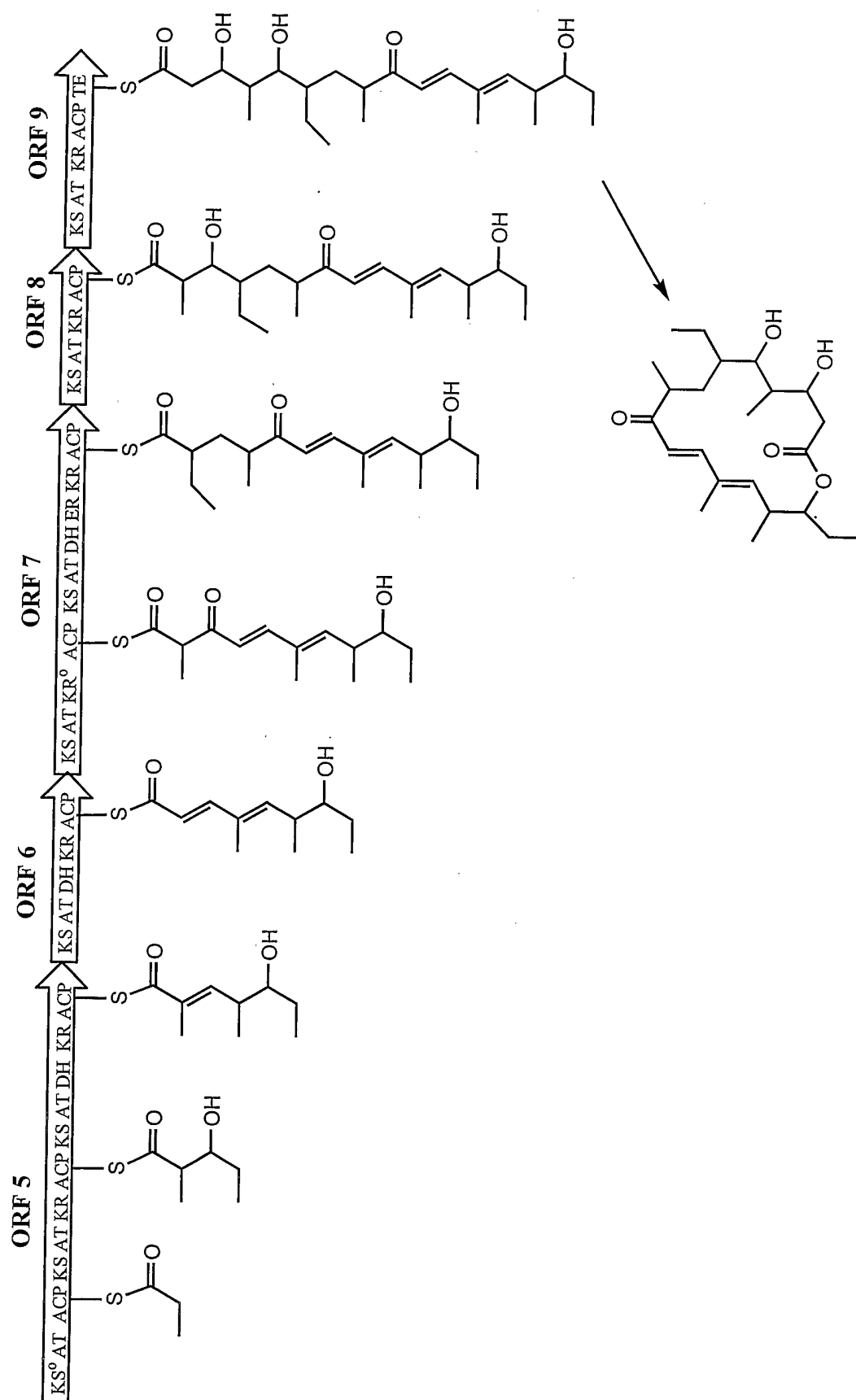


Figure 7

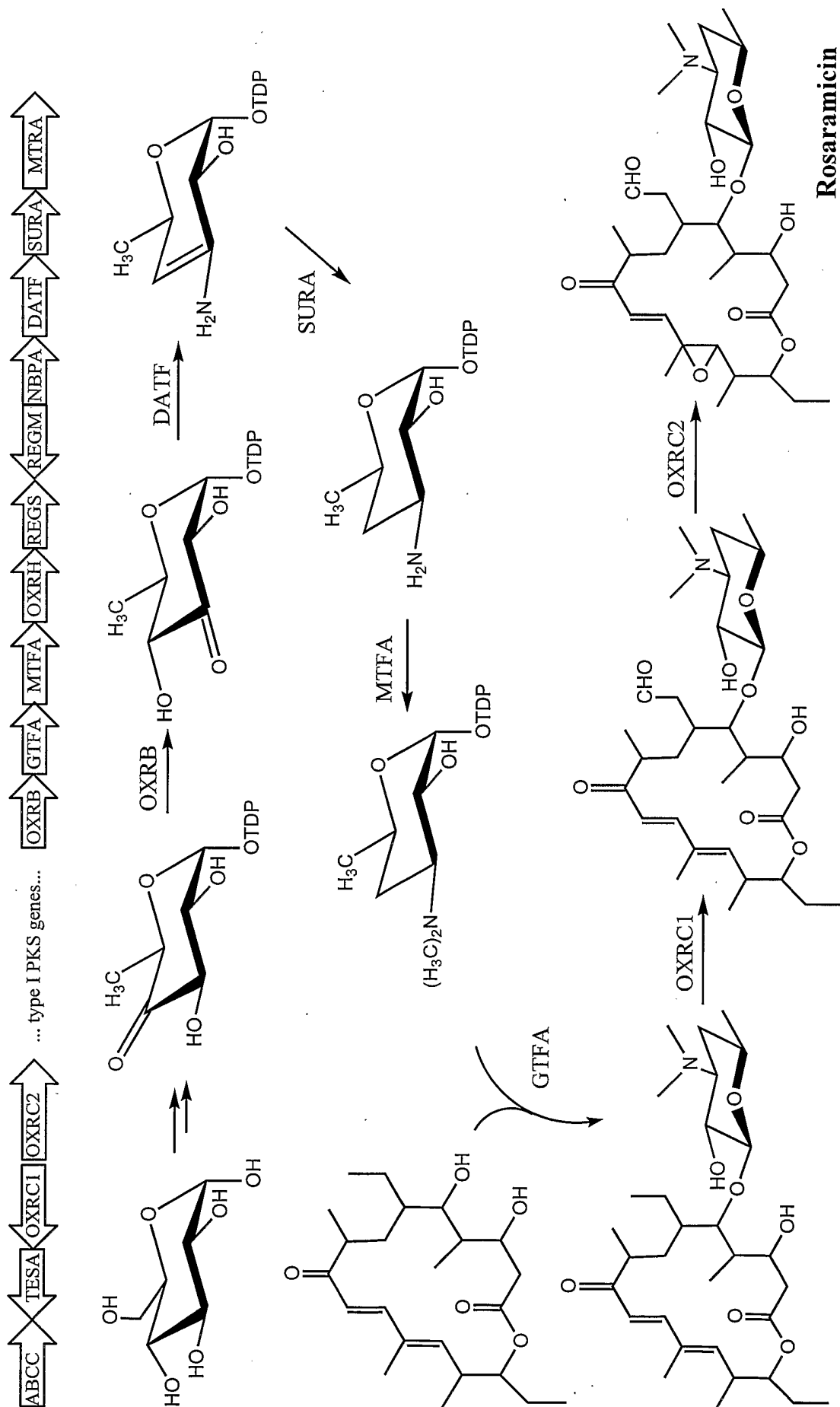


Figure 8A

ORF5	KS1	DIIVIGMSCRLPG-APSIIEFWDLCLCSGRSAVDRQP-DGGWR-----
ORF5	KS2	PVVVVGMCGRFPGGVVCAEGLWDLVLGGGDAVSGFPVDRGWDVEGLFDPVRGVVGKSYVR
ORF5	KS3	PIAIVGMACRYPGGADTPEKLWDLVLAGADVIGPAPDDRWDVDVDFVPGAAGKSYAR
ORF6	KS1	PIAIVGMSCRLPGGVSTPEDLWRLVEAGTDAISGFPPDRGWDVGRLYDPPDSTGTSYVR
ORF7	KS1	PVVVVGMCGRFPGGVVCAEGLWDLVLGGGDAVSGFPVDRGWDVEGLFDPVRGVVGKSYVR
ORF7	KS2	PIAIVGMACRFPGGVADPDDLWDLAKGRDAISRFPNTRGWDVDGLYDPPDPEAPGRTYVR
ORF8	KS1	PIAIVGMACRYPGGVGSPEELWELVASGTDAISFPDDHGWGDALYDPPDPEAAGRITYCR
ORF9	KS1	PIAVVGMACRYPGGVSSPEDLWRLVETGTDAIGGFPTDRGWDVDAVYDPPDPSRNTTYCR
		: : : * * * * * . : : * * : * . . . * : * *
ORF5	KS1	--AVIDGKGESDAAFFGMSPRQAAAVDPQQRLMLELGWEALENARIRPADLKGSSTGVFV
ORF5	KS2	EGGFVYDAGMFDAEFFGVSPREAVAMPQQRLFLEVSWEALERAGIDPLGLRGSRTGVYV
ORF5	KS3	EGGFVYDAGMFDAEFFGVSPREAVAMPQQRLLETSWEALERAGIDPAGLRGSRTGVYS
ORF6	KS1	EGGFLYDCAEFDPEFFTVSPREALAMPQQRLLEAAWETFERAGIAPDSARGTRTGVYV
ORF7	KS1	EGGFVYDAGMFDAEFFGVSPREAVAMPQQRLFLEVSWEALERAGIDPLGLRGSRTGVYV
ORF7	KS2	EGGFLHDAPDFDAFFGISPREALAMPQQRLLETTWESLERAGIDPTALRGTRTGVFV
ORF8	KS1	EGGFLAGVGDFDAFFGISPREALAMPQQRLLETSWEALERAGIDPPDLSRGSRTGVCV
ORF9	KS1	EGGFLAGAGDFDAFFGVSPHEAVVMDPQQRLLEVSWEALERSGTDPHSLRGSRTGVYV
		. . : . * . * * : * : * : * : * : * : * : * : * : * : * : *
ORF5	KS1	GLTADDYATLLRRSGTPISGHTATGLNRSLTANRLSYLLGLRGPSTVDSAQSSSLVAVH
ORF5	KS2	GVMGQEYGPRLVESGGGFEGLLTGTSPSVVSGRVSIVLGLGEPISVDTACSSSLVALH
ORF5	KS3	GLTHQEYAAARLHEAPQELEGYLLTGKSVSVASGRVSIVLGLGEPISVDTACSSSLVALH
ORF6	KS1	GVMYDDYGSRLSEVPKDLLEGYLNGSAGSVASGRIAYTLGLQGPVTVDTACSSSLVALH
ORF7	KS1	GVMGQEYGPRLVESGGGFEGLLTGTSPSVVSGRVSIVLGLGEPISVDTACSSSLVALH
ORF7	KS2	GTNGQHYMPLLRDGDADDGDLGTGNSASVMSGRLSYVFGLEGPAVTVDTACSSSLVALH
ORF8	KS1	GAWHGGYTDVVGQPPAELEGHLLTGGVVSFTSGRISYALGLEGPALTVDACSSSLVALH
ORF9	KS1	GAAHQGYAVDAGQVPEGAEGFRLTGSADAVLSGRISYLLGLEGPALTVDACSSSLVAVH
		* * * . * . . * : . : * : * : * : * : * : * : * : * : *
ORF5	KS1	LACESLLRGESAVAVVGGVSLILAEESTAAMARMGALSPDGRCTTFDARANGYVRGEGGV
ORF5	KS2	LACQGLRLGECDAVALAGGVTVIAAPGLFVEFSRQGLSGDGRCAFAGGADGTGWGEGAG
ORF5	KS3	LACQGLRLGECDAVALAGGVTVIAAPGLFVEFSRQGLSGDGRCAFAGGADGTGWGEGAG
ORF6	KS1	LAVQALRSGECELALAGGATVLTPTMFVDFARQRLAEDGRCKAFADAADGTGFGEVVG
ORF7	KS1	LACQGLRLGECDAVALAGGVTVIAAPGLFVEFSRQGLSGDGRCAFAGGADGTGWGEGAG
ORF7	KS2	LAVQALRRGECTLALVGGATVMSTPDMLVEFSRQRAMS PDGRSKAFAAAADGVALSEGAA
ORF8	KS1	LAVRALRQGECDLALAGGATVLASPAVFVQFSRQRGLAPDGRCKAFADSADGFGPAEGVG
ORF9	KS1	LAVQALRRGECGLALAGGVAVMPDPAFVFEFSRQRGLAADGRCAFAGGADGTGWAEVVG
		** . . * * * . : * : * * : : . : * . : : * * . : * * * * . *
ORF5	KS1	AMVLKPLIRAIEDGDQVHCVIRGCAVNNDDGGGPSLTHPDREAQEALLRRAYERAGVAPEH
ORF5	KS2	VVVLERLSVARERGRHRLAVVRGSAVNQDGGSNGLTAPSGVAQRRVIGAAALVAAGLGVSD
ORF5	KS3	VVVLERLSVARERGRHRLAVVRGSAVNQDGGSNGLTAPSGVAQRRVIGAAALVAAGLGVSD
ORF6	KS1	MLLVERLSDAVRNRRQVLAVVRGSAVNQDGASNGLTAPNGTAQQLVIRQALTNAGLADE
ORF7	KS1	VVVLERLSVARERGRHRLAVVRGSAVNQDGGSNGLTAPSGVAQRRVIGAAALVAAGLGVSD
ORF7	KS2	MMVVQRLADAEAAAGHEILAVVKGSAVNQDGASNGLTAPNGPSQERVIRQALADAGLRPDQ
ORF8	KS1	MLVVERLSDAVRHGRRLALVTGTAVNQDGASNGLTAPSGPAQEKVLRQALVDARVTAAD
ORF9	KS1	VLVLQRLSDAVRDGRWVLGVIRGSAVNQDGASNGLTAPSGPAQQRVIRQALTDARLGADQ
		: : : * * : : * * * : * * . * * . : * . : : * * * : .
ORF5	KS1	VDYVELHGTGTGKAGDPVEAAALGAVLGVARCDNPLAVGSVKTNVGHLEGAAGITGLLKA
ORF5	KS2	VDVVEAHGTGTRLGDPIEAELLSYGRGRVGG-ALLLSVKSNIHGTQAAAGVAGVIKM
ORF5	KS3	VDVVEAHGTGTRLGDPIEAELLSYGRGRVGG-ALLLSVKSNIHGTQAAAGVAGVIKM
ORF6	KS1	VDAVEAHGTGTRLGDPIEAQALLATYQQRPADRPPLLGLSKSNIGHTQAAAGVAGVIKT
ORF7	KS1	VDVVEAHGTGTRLGDPIEAELLSYGRGRVGG-ALLLSVKSNIHGTQAAAGVAGVIKM
ORF7	KS2	VDAVEAHGTGTALGDPIEAQALLATYGRDRPAGRPLWLGLSKSNIGHTQAAAGIAGVMKV
ORF8	KS1	VDAVEAHGTGTRLGDPIEVRALMNVYAGRPADRPWLGLSKSNIGHTQAAAGVGGVIKT
ORF9	KS1	IDAVEAHGTGTRLGDPIEAQALIAAYGADRTPDRPLWLGLSKSNIGHTQAAAGVGGVIKM
		: * * * * * * * : * * * * * * * : * * * * * * * : * * * * * * *

Figure 8B

ORF5 KS1	VLCVREGVLPPSLNFRTPNPDIRLDELNLRVQTELQWPWPGDGTGRPRVAGVSSFGMGGTN
ORF5 KS2	VMALRAGVVPATLHVDVPSPPLVDWSSGGVELVTEARDWP--VVGRVRRAGVSAFGVSGTN
ORF5 KS3	VMALRAGVVPATLHVDVPSPPLVDWSSGGVELVTEARDWP--VVGRVRRAGVSAFGVSGTN
ORF6 KS1	VLALRHARLPRTLHVDRPSTRVDWSSGAVRLLTEGRPWPDP-HGDRPRRAGVSSFGASGTN
ORF7 KS1	VMALRAGVVPATLHVDVPSPPLVDWSSGGVELVTEARDWP--VVGRVRRAGVSAFGVSGTN
ORF7 KS2	ILALRHDTLPRTLHVDRPTPRVDWASGAVSLLTEPVPWPDP--QGDEPRRAAVSSFGISGTN
ORF8 KS1	VLAMRHGVLPPPTLHVDAPTTEVDWSAGQVALLRAETPWPDP-T-GRPRRAGVSSFGVSGTN
ORF9 KS1	LLAMRAGTLPPTLHADVPTPLVDWSAGVVRLSTGVVPWPA-LPGAPRRAGISAFGVSGTN
	:::.* :* :*: *.. : : : ** . * *.:*:** .***
ORF5 KS1	AHLILEQA
ORF5 KS2	AHLILEQA
ORF5 KS3	AHLILEQA
ORF6 KS1	AHVILESA
ORF7 KS1	AHLILEQA
ORF7 KS2	AHVIVEQA
ORF8 KS1	AHVVLEQA
ORF9 KS1	AHVIVEQP
	**:::*. .

Figure 9A

```

ORF5 | AT1      -VVPVVVSGRSVGALRAYAGRLREVCAGLSDGGGSGGGSGGLVDVGWSLVSSRSVFEHRAV
ORF5 | AT2      -VVPVVVSGRSTAALRAYAGRLREVCAGLSDG-----AGLVNVGWSLVSSRSVFEHRAV
ORF5 | AT3      -VVPVVVSGRSVGALRAYAGRLREVCAGLSDGGGSGGGSGGLVDVGWSLVSSRSVFEHRAV
ORF6 | AT1      --PVWVLSGRTEAALREQARRLHAHLTSRARP-----EPADAVARALARSRATAFAYRAA
ORF7 | AT1      -TVPLFLSARGSAALCAQAARLRARLIEEPDL-----DIAEVGYTLAATRARFEHRAV
ORF7 | AT2      -VVPVVVSGRSVGALRAYAGRLREVCAGLSDGGGSGGGSGGLVDVGWSLVSSRSVFEHRAV
ORF8 | AT1      -VVPLVLSARSQPALAGQARRLRDLLAAPET-----DLASAGLALATARSVFDHRAV
ORF9 | AT1      PLVPWVLSGRTEAALRAQADRLRTHLAHPDA-----DPLD-VGFSLATSRAALEHRAV
          . : * . * * * : . . : * . : * : : * * .

ORF5 | AT1      VFGGGVAEVEVAGLDAVASGAVSSGSVVVGSVASGVAGGGGRRVVFVFPQGQGWQWVGMGAAL
ORF5 | AT2      VFGGGVAEVEVAGLDAVVSAGVASGSVVVGSVASGVAGGGGRRVVFVFPQGQGWQWVGMGAAL
ORF5 | AT3      VFGGGVEEVVAGLGAVASGAVASGSVVVGSVASGVAGGGGRRVVFVFPQGQGWQWVGMGAAL
ORF6 | AT1      VLGRDDTARLDGLHALAAGRSAAG-----LVTGRAVPERRVAFLFTQGGSQRPAGAGREL
ORF7 | AT1      VIGESRAEVGDALAALARGEHP-----SLLRGRAGASDRVAFVFPQGGSQWAEMADGL
ORF7 | AT2      VFGGGVEEVVAGLGAVASGAVASGSVVVGSVASGVAGGGGRRVVFVFPQGQGWQWVGMGAAL
ORF8 | AT1      VTAAGRPQALDALDLAGGEPGP-----AVTGTVAAPTGRVTFVFPQGQTHWAGMGADL
ORF9 | AT1      LVAADDRDGLLRDLVDALAAGEPAAG-----LIRGTVRHRRRTGFLFAGQGGQVRVGMAREL
          : . . : : * . : * . * . * : * : . *

                                     SSSSS      SSSSS
ORF5 | AT1      LDESEVFAESMVECGRALSGFVDWDLLEVVRGGGG---DGSFGRVDVVQPVSWAVMVSLA
ORF5 | AT2      LDESEVFAESMVECGRALSGFVDWDLLEVVRGGAG---EGVWGRVDVVQPVSWAVMVSLA
ORF5 | AT3      LDESEVFAESMVECGRALSGFVDWDLLEVVRGGAG---EGVWGRVDVVQPVSWAVMVSLA
ORF6 | AT1      YARHPAFAQALDGLVLAELDRHLDRPLRAVMLAEPGTEAAALLDDTAYTQPALFALEVALF
ORF7 | AT1      LDRSPAFRASASACDEALRAHLDRVLDVLRVPD---APALSRVDVVQPVLFMTMMVSLA
ORF7 | AT2      LDESEVFAESMVECGRALSGFVDWDLLEVVRGGAG---EGVWGRVDVVQPVSWAVMVSLA
ORF8 | AT1      LDQSPVFAESMRRCEQALSAHTDWKLGEVIRGAAG---SPPLDRVDVLPVSWAVMVSLA
ORF9 | AT1      YEAFPAFADALDQLAARLDRLDRPLLRVLFAPGSDDARLLDGTRYAQAALFAVEVALF
          . * : . * . * : * : . . * . : : * : *

SSS      S
ORF5 | AT1      RLWMSVGVPDAVVGHSQGEVAAAVVGGVLSVADGARVVALRSRVIGEVLAGGGAMVSVG
ORF5 | AT2      RLWMSVGVPDAVVGHSQGEVAAAVVGGVLSVADGARVVALRSRVIGEVLAGGGAMVSVG
ORF5 | AT3      RLWMSVGVPDAVVGHSQGEVAAAVVGGVLSVADGARVVALRSRVIGEVLAGGGAMVSVG
ORF6 | AT1      RLVTSWGLRPDALLGHSVGEITAAYVAGVLTLPDAARLVAVRGRLMADLRAG-GAMAALQ
ORF7 | AT1      AAWRALGVHPSAVVGHSSQGEIAAAHVAGGLSLDDAARIVALRSQAWLR-LAGQGGMVAVS
ORF7 | AT2      RLWMSVGVPDAVVGHSQGEVAAAVVGGVLSVADGARVVALRSRVIGEVLAGGGAMVSVG
ORF8 | AT1      QVWRSLGVEPDVAVVGHSQGEIAAAVVCALTLTPDAARVVALRSQVIGRVLSGRGGMASVQ
ORF9 | AT1      RLVHGWGVRPDVLLGHSVGEIAAAHVAGVLDVDDACELVAARGRLMGELPSG-GAMVAVR
          . * : * . : * * * * : * . : * * : * : * : *

LPVAVVLDRLAGWGGRGLGVAAVNGPSLTVVSGDVDAAVGFVGECERDGVWVRRVA-VDYA
ORF5 | AT2      LPIVDAQERLAGWGGRGLGVAAVNGPSLTVVSGDVDAAVGFVGECERDGVWVRRVA-VDYA
ORF5 | AT3      LPIVDVQERLAGWGGRGLGVAAVNGPSLTVVSGDVDAAVGFVGECERDGVWVRRVA-VDYA
ORF6 | AT1      AAESEVDPLLAGREGELSIAAVNGPQATVIAGDEAAVEEQVALWRDRGRARRRLR-VGHA
ORF7 | AT1      LPVDALRARLARFGDRLSVAAVNSPGTAAVSGYPDALAELVDELTAEGVHAKAIPGVDTA
ORF7 | AT2      LPIVDVQERLAGWGGRGLGVAAVNGPSLTVVSGDVDAAVGFVGECERDGVWVRRVA-VDYA
ORF8 | AT1      LPAREVAGRLAAWAGRLDVAAVNGPQSTVVSGAADAVTELVEAFAEDVRVRRIP-VDYA
ORF9 | AT1      ATEEEVGPLLDGQ--RVAAAVNGPRSVVSGDEEAVLAVAARCAALGHRTRRLN-VSHA
          . * . : : * * . . : : * . . : : * . *

```

Figure 9B

ORF5	AT1	SHSAHVEAVEGMLSGLLGGLCPGRGVVPFYSSVVGGVVDGVGL-DGGYWYRNLRRERVLFS
ORF5	AT2	SHSAHVEAVEGMLSGLLGGLCPGRGVVPFYSSVVGGVVDGVGL-DGGYWYRNLRRERVLFS
ORF5	AT3	SHSAHVEAVEGMLSGLLGGLCPGRGVVPFYSSVVGGVVDGVGL-DGGYWYRNLRRERVLFS
ORF6	AT1	FHSVRMDGMLAEFEKAMGDLRAGEPTIPVVANVRGAIASGTDLRTAGYWIRHAREPVRFL
ORF7	AT1	GHSAQVEVLKDHLMAALAPVSPRSSQIPFYSTVTGGLLDTALL-DAAYWYRNM RDPVEFE
ORF7	AT2	SHSAHVEAVEGMLSGLLGGLCPGRGVVPFYSSVVGGVVDGVGL-DGGYWYRNLRRERVLFS
ORF8	AT1	SHSTQVDRRLRAELLTVLGPVDARPAQVPFYSTVQGGRVDTAGL-DAGYWYRNLRGQVRFE
ORF9	AT1	FHSPHVEAMLEPFRRVARGLTYHAPTIPVVS NATGRLATADALRDPGYWVRHVRQPVRFR
		** : : : : : : : * . . . * * . ** * : * * *
ORF5	AT1	DVVGRVLVDGDFSGFVECSGHPVLAGGVLESVA-----VDPDVRPVVVG
ORF5	AT2	DVVGRVLVDGDFSGFVECSGHPVLAGGVLESVA-----VDPDVRPVVVG
ORF5	AT3	DVVGRVLVDGDFSGFVECSGHPVLAGGVLESVA-----VDPDVRPVVVG
ORF6	AT1	DGMRALRAEGVDTFVELGPDGVL TAMARDCLADPADPVDLADAAEPAGAAEPDRSLLFLP
ORF7	AT1	QATRAMLADGHEGFLEPSHPMLSVSLQGTAA-----DAG--VAATVLG
ORF7	AT2	DVVGRVLVDGDFSGFVECSGHPVLAGGVLESVA-----VDPDVRPVVVG
ORF8	AT1	ETVRVLLDDGHRAFVEAAAHA VLVPAIQELGD-----SAG--VRVAVG
ORF9	AT1	DGVRAARDQGATAFVGLGPDGVL CALAE ECLG-----PTGDVLLLP
		: : * * : . . : *
ORF5	AT1	SLRRDDGGWGRFLT SVGEAFVGGMSVDWKGVFAGAGARLVDLPTYPFQRRHYWAPN
ORF5	AT2	SLRRDDGGWGRFLT SVGEAFVGGMSVDWKGVFAGAGARLVDLPTYPFQRRHYWAPT
ORF5	AT3	SLRRDDGGWGRFLT SVGEAFVGGMSVDWKGVFAGAGARLVDLPTYPFQRRHYWAQT
ORF6	AT1	TLRRDRDDAVAVREALASVHVHGLPVDP-VAPLGDGPLATDLPTYPFQRSRYWL--
ORF7	AT1	TLRRGKGGARWFGMALGLAHAGHIEIDAS-VLFGTDSRRVDLPTYPFQRRERFWYHP
ORF7	AT2	SLRRDDGGWGRFLT SVGEAFVGGMSVDWKGVFAGAGARLVDLPTYPFQRRHYWAPT
ORF8	AT1	SLRREAGGLDRLLASAAEAFTQGVAVDWSRALAGAA R VAVDLPTYAFQRQRYWLEP
ORF9	AT1	VL RPGRPEPATLLAALAGAYAGGAEMDWSRVFAGTGARRVELPTYAFQHRRYWL--
		** . : . . . * : * . * . : ***** : : *

Figure 10

ORF5 DH3	FGMEWEDHPLLGGALSVGGSRSLLLAGHLSLASHAWLTDEAVSGTVLLPGTAFVELALHA
ORF6 DH1	VGLDVAGHPLLAVAVDLPDGAGTVWSGQLCVRTHPWLADHSVWGRTVVPGTALLEIMHRV
ORF7 DH2	AGLSGADHPLLGGAVELPDRGGHVYPARLGVRHHPWLGEHALLGAAILPGAAYAEALWA
	*:. . ***** *:.: . . : .:.* : *.** :*:.* :*:.* *: .
ORF5 DH3	AAAAGCPEVEELRLEAPLVVPARGGVRLQVLVDDPDDGSDRRRAVSFVS
ORF6 DH1	RAEVGCTRVAELTFEAPMVLADDGGVRVRVVVDGPD-ADGARQVRIHS
ORF7 DH2	GRRDGAGRIEELTLDAPLVVADESAAQLRLVVGPD-AEGRRQLTVHS
	*. .: ** :*:.*:. . .:.*:. .* . . * : :.*

Figure 11

ORF7 ER2	QLAVRAGAVHVPRLVRHRPRPDGPLTPPAGAAWRLAAGGQGTLEGLALVPAPDALAPLAP
DEBS ER_mod4	QLALRGDDVFVPRLS---PLAPSALTLPAG-TQRLVPG-DGAIDSVAFEPAPDVEQPLRA
	:*. . *.** * . . .** *** : **..* :*:::*:*: *****. ** .
ORF7 ER2	GQVRVAVRAAGVNFRDTLIALGMYPGTPVLGAEGAGVITEVAPDVAGFAPGDRVLGMWTG
DEBS ER_mod4	GEVRVDVRATGVNFRDVLALLGMYPQKADMGTAAAGVVTAVGPDVDAFAPGDRVLGLFQG
	*:*** ***:*****.*:***** .. :*:*.***:* *.*** .*****: : *
ORF7 ER2	GLGPVAVADARMLARVPRGWSYAEAAVPAVFLTAHYALTRLAGIRPGQSLVHAGAGGV
DEBS ER_mod4	AFAPIAVTDHRLRLARVPDGSADAAVPIAYTTAHYALHDLAGLRAGQSVLIHAAAGGV
	..*:***:* *:***** *** *:***:*** .: ***** ***:*.***:***.***
ORF7 ER2	GMATLQLARHLGVEVYATASRGKWDTLRGLGLDDAHIADSRSLDFAGRFLAATGGRGVDV
DEBS ER_mod4	GMAAVALARRAGAEVLATAGPAKHGTLRALGLDDEHIASSRETGFARKFRERTGGRGVDV
	.: ***: *.** ***. .* ..***** ***.***. .** :* *****
ORF7 ER2	VLNSLAGDFVDASLRLLPRGGHFLELGKADVRDPDRIAADHPGVGYRAFDLVEAGPELVG
DEBS ER_mod4	VLNSLTGELLDESADLLAEDGVFVEMGKTDLRD----AGDFRGR-YAPFDLGEAGDDRLG
	*****:*:::* * **..* *:***:***. *. * * .*** *** : :*
ORF7 ER2	QLLGELMELFAAGVLSPLPLTVRDVRRAREAFRLISQAR
DEBS ER_mod4	EILREVVGLLGAGELDRLPVSAWELGSAPAALQHMSRGR
	::* *: :*.** *. ***** : : * *: :*: :*

ORF5	KR2	PRGNILVTGCTGSLGGRVAVRWLARN-GAEHLVLTSSRGADAPGAAELEDLRALGVEVTM
ORF5	KR3	PDGTVLLTGATGTLGGLLARHLVTTGRARRLLVSRSGPDAPDAGRLTEELTGLGAHVTL
ORF6	KR1	AAGTVLVGTGCTGMLGGAVARHLVRRHGVRRLLLVGRRGPDAPGAAALTRELEELGASVRV
ORF7	KR1	ARGTILVVGDTGPVAALLAGRLLDG-GAAHVVLG-----PAAASTVGLTGGADRVAL
ORF7	KR2	AYGTVLVGTGCTGTLGGAVARHLVARHGVRLVLTGRSGPADGASALVDEL TASGASVTV
ORF8	KR1	PRGTVLVGTGCTGALGAHVARRLAAG-GAAHLVLTSSRGADAPGAAGLVGELRALGAEVTV
ORF9	KR1	PRGTVLVGTGCTGALGAHTARWLARN-GATHLVLTSSRGGNAPGVAALRAELVTLGAEVTV
		. * . : * . * * : . . * * * . : : * * . * :
ORF5	KR2	AACDVADRAALSDVLA----AHPPTAVFHTAGVLHDGVIDTLAAGHIDEVFRPKTAAALL
ORF5	KR3	AACDTTDRAALAGVLGGIPAEHPLTAVVHVAGVLDDGAVQALTPERVDVAVLRPKVDAALH
ORF6	KR1	AACDVGDRGAVTRLLAGVPAAHPLTAVVHSAGLPDDGVLT AQTGERVAAVLRAKADA AVN
ORF7	KR1	IDCDPSDRDALAGLLG----AYRPTTIVVAPPAVALTALAETTPEDFVAAVAAKT'TTAVH
ORF7	KR2	VACDAADRVALRRLLDGIPAAHPLTAVVHAAGVLDDATITALTAGQVDVAVLRPKADAVIN
ORF8	KR1	AVCDVADRAAVAALLAGLPADAPLSAVFHTAGVAHSMPIGETGLTDVAEVFAGKVAGARH
ORF9	KR1	VACDVADREAVAGLLAGIPRAAPLTAVFHAAGVPQVTPPLHETTPELFAQVCAGKVAGAVH
		** ** *: : * : : . : : . . *
ORF5	KR2	LDELTDQH--QELDAFVLFSVVTGVWNGGQAAAYAAANASLDALAERRRAAGLPATSIAWG
ORF5	KR3	LHELTDAG--LPLAAFVLFSGAAGILGRPGQANYAAANTFLDALAQHRRARGLPGVSLAWG
ORF6	KR1	LHELTDH--LDLTAFVLFSVSVAGTIGSAGQAGYAAANAFLDASFASWRQGGQLPATALAWG
ORF7	KR1	LDALAAEAELELDASFVVFSSVAGTIGSGAGHGYYAAGTARLDALVEERRARGLPATAIAWT
ORF7	KR2	LHELTRD--RELDAFVLFSAAALFGSPGQGNYSANGFVDFAFQYRRARGLHVAISLAWG
ORF8	KR1	LDELTRG--HDLDAFVLYSSNAGVWSSGQSAYGAANAALDALAERRRAAGLTATSVAWG
ORF9	KR1	LHELTDG--DLDAFVTFASAAGVWSSGQCAYAAANAALDALAERRRAAGLPATSVAWG
		* . * : * *** : . . * * : * . * . : ** . . * . * : **
ORF5	KR2	LWGGGG-MAEGIG-----EQNLNR-----RGITALDPELGIAALQQ
ORF5	KR3	LWGLASDMTGHLG-----EQDLRR-----MRRSGIAPMTGEEGLALFDL
ORF6	KR1	--PLDGGMAAGLG-----TADVAR-----LRRSGLVPLGVDDALVLFDA
ORF7	KR1	PWADATTAAGGQAPDASAGGHEPDTRAGGPDRELLRRGGTLPDLPGAALDVLRG
ORF7	KR2	LWADSSRMAGHLD-----QEGMRRR-----MARGGVLP LTTDQGLALFDA
ORF8	KR1	LWGS GG-MGEGDA-----EEYLSR-----RGLRPMPPPERGVDALLA
ORF9	KR1	VWGGPG-MGAGAG-----EEYLRR-----RGVRAMPAAALALGR
		* : : : : :

Figure 13

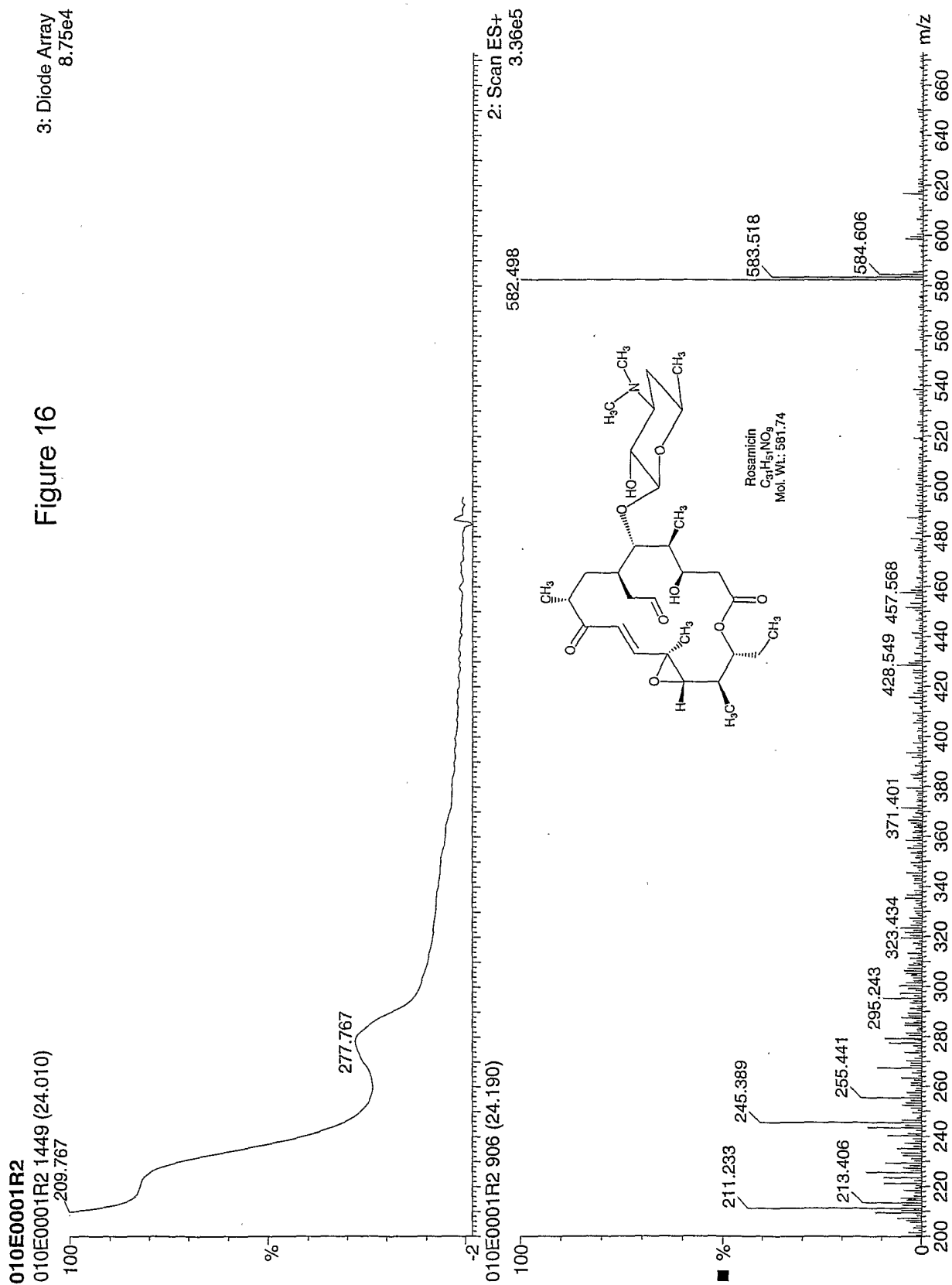
ORF5	ACP1	AVVLGLDSAQAVDPERTFKEHGFESVTAVELCNHLQRTGRLRVPASLVYNHPTPMAAARKLQ
ORF5	ACP2	AAVLRHETVDAVAPTRAFKDAGFDSLTALELRNHLNSTTGLSLPPTVVFDHPTPSTLAKFLE
ORF5	ACP3	AGVLALREAADVDPGRPFREVGFDSLTAVELRNRLGSATGLRLAPSLVFDHPTPSAVAEHLV
ORF6	ACP1	AAVLGHGEAAMLSTQRAFRDAGFDSLTAVDLRNRLGAATGLSLPAAVVFDHPTPAALAAYLR
ORF7	ACP1	AAVLGHDEAEAADPDRAFRRELGFTSVTAVDLRNRLNAATGLNLPASVVFDHPSARVLAAYLR
ORF7	ACP2	AAVLGHADPQAVDADRAFRRELGFDSLTAVELRNRLATASGLRLPATLVFDHPTPEALAEHLL
ORF8	ACP1	AAVLGHGPEHVGPDAAFREIGFDSLTAVDLAKRLRAAVGVPLSATLVFDHPTATAVAEHLA
ORF9	ACP1	AGVLGHDGADDVPADAEFSALGFDSLAAVQLRRRLAEATGLSLSAPVLFDHRTPDALAAHLH
		* * * . * * * *:::* .:* *: :.....* :. . * *

Figure 14

ORF9 Te	TGAGGPALVCCAGTAAASGPREFTAFAAALAGLRDVTVLPQTGFLPGEPLPAGLDVLLDA
DEBS Te	DGPGEVTVICCAGTAAISGPHEFTRLAGALRGIAVRAVPQPGYEEGEPLPSSMAAVAAV
	. :.:***** :.:*** :.*.* *: * .:*.*: *****.: .: .
ORF9 Te	QADAVLAHCAGGPFVLVGH S SAGANMAHALTVRLEARGADPAALVLMIDIYTPAAPGAMGVW
DEBS Te	QADAVIRTQGDKPFVVAG H SAGALMAYALATELLDRGHPPRGVVLIDVYPPGHQDAMNAW
	*****: .. ***:.* **** * **:***:..* ** * .:***:*.*. .***.*
ORF9 Te	REEMPLAWVAERSVVPVDDTRLTAMGAYHRLLLDWAPRPTRAPVLHLYAGEPAGAWPDPRQ
DEBS Te	LEELTATLFDRETVMDDTRLTALGAYDRLTGQWRPRETGLPTLLVSAGEPMGPWPD--D
	:* : :*.*. :***:***.* :* ** * *.* : ******.*** :
ORF9 Te	DWRSRFDGAHTSAEVP G THFSMMTEHAPVTAATVHKWLDEVCP
DEBS Te	SWKPTWPFHDTVAV P GDHFTMVQEHADAIARHIDAWLGGGNS
	.*:.. : * :. ** **:*: *** . * :. **. .

Figure 15

ORF7	AT2	TVPLFLSARGSAALCAQAARLRARLIEEPDLIDIAEVGYTLAATRARFEHRAVVIGESRAE
TYLO	AAB66506	AT2 TVPLLLSGHTEAALREQSTRLLNDLLEHPDEHPADVGYTLITGRAHFGHRAAVIGESREE
NIDD	AAC46026	AT2 TVPLMLSAHSEAALREQARRLCAQLLARPEQRPADVGHALLSTRARFPRRAAVVGESMTE
		****:*.:.*** *: ** *: .*: *:***:* : **:* :*.*:*** *
ORF7	AT2	VGDLAALARGEHHPSLLRG--RAGASDRVAFVFPQGGSQWAEMADGLLDRSPAFRASAS
TYLO	AAB66506	AT2 LLDALKALAEGREHHTVVRGDGTAHPDRRVVVFVFPQGGSQWPSMARDLLDRAPAFRETAK
NIDD	AAC46026	AT2 LAEALDAVAEGGPHPLAATG--TAGTADRVVVFVFPQGGSQWAGMAEGLLERSGAFRSAAD
		: :** *:*. * * * * . **.*****. ** .**:*: *** :*.
ORF7	AT2	ACDEALRAHLDWSVLDVLRVPDAPALSRVDVVQPVLFTMMVSLAAAWRALGVHPSAVVG
TYLO	AAB66506	AT2 ACDAALSVHLDWSVLDVLOEKPDAPPLSRVDVVQPVLFTMMLSLAACWRDLGVHPAAVVG
NIDD	AAC46026	AT2 SCDAALRPYLGWSVLSVLRGEPDAPSLDRVDVVQPVLFTMMVSLAAVWRALGVEPAAVVG
		:** ** :*.***.**: *****.*****.***** ** ***.**:***
ORF7	AT2	HSQGEIAAAHVAGGLSLDDAARIVALRSQAWLRLAGQGGMVAVSLPVDALRARLARFGDR
TYLO	AAB66506	AT2 HSQGEIAAACVAGALSLEDAARIVALRSRAWLTLAGKKGMAAVSLPEARLRERIERFGQR
NIDD	AAC46026	AT2 HSQGEIAAAHVAGALSLLDSARIVALRSRAWLGLAGKGMVAVPMPAEELRPRLVTWGDR
		***** ***.***:*.*****:*** ***:***.***.*** ** *: :*: *
ORF7	AT2	LSVAAVNSPGTAAVSGYPDALAEVLDELTAEGVHAKAIPGVDTAGHSAQVEVLKDHLMAA
TYLO	AAB66506	AT2 LSVAAVNSPGTAAVAGDVALRELLAEVLTAEGIRAKPIPGVDTAGHSAQVDGLKEHLFEV
NIDD	AAC46026	AT2 LAVAAVNSPGSCAVAGDPEALAEVLLTGEGVHARPIPGVDTAGHSPQVDALRAHLLLEV
		*:*****:***: ** **: **.***:*.*****.***: *: **: .
ORF7	AT2	LAPVSPRSSQIPFYSTVTGGLLDLTALLDAAYWYRNMRDPVEFEQATRAMLADGHEGFLEP
TYLO	AAB66506	AT2 LAPVSPRSSDIPFYSTVTGAPLDTERLDAGYWYRNMRPVEFEKAVRALIADGYDLFLEC
NIDD	AAC46026	AT2 LAPVAPRPADIPFYSTVTGGLLDGTELDITYWYRNMRPVEFERATRALIADGHDVFLET
		****:*.:.*****. ** ** *****:*****:*.***:***: **
ORF7	AT2	SPHPMLSVSLQGTAAAGVAATVLGTLRRGKGGARWFGMALGLAHAGHIEIDASVLFGTD
TYLO	AAB66506	AT2 NPHPMLAMSLDETLTDSGGHGTVMHTLRRQKGSAKDFGMALCLAYVNGLEIDGEALFGPD
NIDD	AAC46026	AT2 SPHPMLAVALAQTVTDAGTDAAVLGTLLRRRHGGPRALALAVCRAFAHGVEVDPEAVFGPG
		.*****:*.*: * :*: * .*:*: **** :*..: .*:*: * ..*:*: * ..**..
ORF7	AT2	SRRVDLPTYPFQRERFWYHP
TYLO	AAB66506	AT2 SRRVNPPTYPFQRERYWYHP
NIDD	AAC46026	AT2 ARPVELPTYPFQRERYWCHP
		:* *: *****:* **



SEQUENCE LISTING

<110> ECOPIA BIOSCIENCES INC.

Farnet, Chris

Yang, Xianshu

Staffa, Alfredo

<120> GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN

<130> 3016-3PCT

<160> 39

<170> PatentIn version 3.0

<210> 1

<211> 60196

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

```

<400> 1
gtgccagttc cgacacagga ggcccccttg cggaacagcc cgccgccagc ccattcgcag      60
ctcgtcctga gcgaggtcac gaagcactac gccgagcggg tcgtcctgga ccgcgtttcg      120
ctcaccgtca agccggggga gggggtcggc gtcacgcggc agaacggggtc ggggaagtcg      180
accctgctgc ggctcgtcgc ggggctggag acgccggaca acggcgagtt gaccgtctcg      240
gcgcccgggg gcacgcgcta tctcgcccag cggcttcggc tgccggccgg cggcagcacc      300
gtacgggatg tggtagacca cacgctcgcc gacctgcgag acctggaggc gcggttgccg      360
gccgccgagg cggacctggc caccgccacg cccgagcagt tggacgccta cggcacgctg      420
ctcactgtgt tcgaggcccg cggcggctac caggccgacg cccgggtgga cgccgccctg      480
cacggtctcg gcctggccga gctcgaccgc gatcgcgacg tcgacacgct ctccggcggg      540
gaacggtccc ggctcgcgct cgccgcgacc ctggccgccc cgccggaact gctgctgctc      600
gacgagccca ccaacgacct cgacatcgag gccgtggagt ggctggagga tcacctgcgg      660
tcgcaccggg gcaccgtcgt cgtggtcact cacgaccggg tggtcctgga gtcggtcacg      720
tccaccatcc tcgaggtcga caccgacacc cgggccgtgc accggtacgg cgacggctat      780
gccagctacc tgccggccaa ggccgccctc cgggagagcc gggagcgcgc gtacgcggaa      840
tgggtggccg aggtcgagcg gcagtcccaa ctgcgggagc gggccgggac gatgctccgg      900
tcgatctccc gcaagggacc ggtgcggttc agcggggccc gtgcccaccg ctcccggtcg      960
tcgtcgacgg cgacgtcacg caaggcccgc aacgccaacg agcggtcttc ccggtcgcg      1020
gagaatccgg taccgcgacc cgccgaccgc ttgcgcttca ccgcgtcggg cggcccgat      1080

```

gccacggacg ccgatacccg ccgcgtcgag ttgaccgacg tccgggtggg ccgccgcctg 1140
 cacgtgcccg agctgaccat cggacccgcc gaacgggttg tggtgaccgg acccaacggc 1200
 gcgggtaaga gcaccctgat gcgggtgctc gccggggaac tcgtgcccga cggcggaacg 1260
 gtgcggctgc cggctcgat cggccacctg cgtcaggacg tgacggtcgg gcagcccggg 1320
 cgctctctgc tggagacgta cgcgtcgggt cggccggggc atcccaggga gtacgcggag 1380
 gaggttgctg cccgcgggtct gttccggccc gatgacctgc gcatgccggg cgggacgctc 1440
 tccgtcgggc agcgccggcg gatcgacctg gcccggtgg tcgcccggc ggccgacctg 1500
 ctgctgttgg acgagcccac caaccacttc gcgcccctgc tcgtggagga gctggaacag 1560
 gcgctggacg gctacgccgg agcgtggtc gtggtgacgc acgaccggcg gatgcggagc 1620
 accttcaccg gggctcggct ggaactgcac cagggcggtg ccaccggggc gagccggggc 1680
 tgacgagccg cccgggggtgc cgtggcgcg cgggacggg ggggatctca gccggggctg 1740
 gcacccggca ccgccgtgag ggcggtcgtg gacaccgtg cgagggtggg cgtgacctg 1800
 gcgcacacag cgtcgagctg atcggtgaga tagaagtgc cggccgggaa cgtgcggacc 1860
 atcggtggcg ctgcggtcac ctcgcccccac gcccgggcct cgtcggtggg gacgtgggtg 1920
 tcggcgggcc cggcgagtag ggtgaccggg caacgcagcc tggggcctgg ccggtattcg 1980
 taggcggcgg cggcccggta gtcgttgctg atggcgggga ggagcatgtc cagcagttcc 2040
 ctgtcgtcca ggaggctgga atcggtgccc tggagccggc ggatctctgc gatcagctcg 2100
 tcgtcaaacc ggtagaaccg gtcccgcgc ccgacggacg ggctacggcg gccggaggcg 2160
 aagaggtgca cgagccgacg ggcgtcggcc ggtgggagcc ggcgggcggc ctcaaggcc 2220
 accgtggcgc ccatgctgtg accgaagaag gccaccggtc ggtccggcca ggcgagcagt 2280
 gcgggcagga gcccgctcac cagggcgctg acggactcga tcaagggttc gccgcggcgg 2340
 tcctgccggc ccgggtactg gaccgccagc acgtccacgt cggcgggcag ccggcgggcg 2400
 aacggcaggt acgcgtggc cgcgcccccg gcgtgcggga agcagaacag ccggacggcg 2460
 gggtcggtga cgggcccggta gcggcgtagc cacagctcgg acggatcggc ggacggggac 2520
 atggtgatct gcgtcctcg gtctgctcga cgttcgggtg tcgggtccca ccccgcgcc 2580
 gaagacggcc atgatgtcg gcacggcggc cgtcacggc tcgacgtctt acttcgggtg 2640
 ccgtccgtcg cgtaccacct ggacggggag gcgtcgcgcg gtgagctggg cggcgctcga 2700
 gaactcgacc ccgacgtggg cgatccggaa ctcgggtgaac tggtcgagcg tctggttgag 2760

gaagaccttc gcctccagcc tggccaggaa cgcgccagg cagtgggtgga tgccgtggcc	2820
gaacgccagg tgcttggttcg actcgcgtcg gatgtcgaag gtgtccgggt ccgtgaacac	2880
ctcgggtgtcg cggttcgcgg aggcgatcca ggcgatcacc atctggccct tgcgcatggg	2940
gtggccgagg atgtcgggtgt cctcgttcag gatccggaag atgcagttga acggggaccg	3000
gtagcgcagc gtctcctcga tcacgcccgg cacgaggctg cggtcggcgc ggaccgcggc	3060
ctgtgcctgc ggggtgctcct ccagcaccag gaacaggttg ctgagcagcg tggcgtgga	3120
gatgtgcccc gcggtgagca gcagcgcgac gatgttgacg acttcctcgt cggtcagctt	3180
gcgcccgtcg acctccgcg cacagaggcc gctgatcagg tcgtccttcg gttcggcgcg	3240
cttgtgggcg atctgggcgt acaggaattc ggaccactcc tcgatggcgg gggccaccgt	3300
ctcgggtgaag tcgtccggga ggttgggata ctccagccct tcgttgetga ggatgatgtc	3360
caccactcg cggaacttct cgtgatcctt ggtgggaatg ccgagcagct cggcgatgac	3420
cgtcaccggc agcgggtacg cgaggtcgct ggcgatgtcg atccggctct ggtcgcgtac	3480
ctggtcgagc acgtcggcgg tgatctgcc gatccgcagc tccatctggg cgatccggcg	3540
gggggtgaac gcctggctca ccagcttgcg cagcggcgcg tgccgcggcg ggtcgatgcc	3600
gccgatggtg ccggggccca tcagcagggc cagctccgac ggtacgggaa agaccgaggt	3660
gaagtccgac gagaagatca gcgggttggg ggtcacggtc tggtagtccc ggtaggagaa	3720
cacgtgccag gcctgacggg tctcgtccca ggagacgggc cagttcttcc gcatgtacgc	3780
gaaccagtcc agcagccctt gggcgtcggc gcccttgggc aggtcgatcg gtcccgcgg	3840
ggcgttcggg gtctgcgtca tgggtgtgctc atctcctcgg tggctctcggc cgtcgggcg	3900
aagggaaga gaaccttggg tcgcgagggc gtccggtcgg ggaggggatc ttccgggctg	3960
gcgtgtcac ctgcggcctg ctcggtcgcc tcgccggcat tgacggttgt gctgggcggc	4020
gagtcagcgc tgtggcggcg ggcagggcgg gccctgcact tctccggggc gtcgtaatct	4080
tcggtccgaa tcgtgatggc cgcaaggccg gacctgacat agtgctgtct gcaacgctcg	4140
gagcaccctt tttatcagtt gattgcggtc atttttgtcg acgatcaggg cggttctata	4200
tcgagacttg acatagtctt ctacggattc gtgacaatga tcatcgatcg gtgttggtg	4260
aatcgacgaa aggggcgtgc tgttcgaggg ggcgttgcca agatcaatgc aaaaccgcat	4320
ccttgatcaa tgcggaaccg caccctgcct ggaagagagc tgccatggag catccagtaa	4380
cggccgggtc ctgcaggttc tacccttca gtgaccgtac cgacctgaat atcgatccca	4440
cgtacggcga actgcgctcg aaagagccgg tcgccgcgt ccgcatgccc tacggcgggg	4500

acgcctggct ggtcaccg gacgcccacg ccaagaaggg cctctctgac ccccgactca 4560
 gcattgcagc cggagccggg cgggacgtgc cgcgcgcctc ccccgctctc caggaaccgg 4620
 acggtctgat gggctctccc cccgacgcgc acgcccgact gcgcaggctc gtcgccacgg 4680
 cgttcacgcc gaagcgcgta cgggacatcg ccccgcgctc cgtccagctc gccgacaagc 4740
 ttctcgacga cgtggtcgaa accggggccgc cggccgacct cgtgcagcag ctgcgcgttc 4800
 ccctgccggg gatgatcatc tgcgagatga tgggcatcgg gtacgacgag cagcacctgt 4860
 tccgtgcctt cagcgatgcc ctgatgtcct ccacccgata cacggccgac caggtcgacc 4920
 gcgcggtaga ggacttcgtc gagtacctcg ggggcctcct cgcgcagcgc cgtgcacacc 4980
 gcaccgacga cctcctcggc gccctggctc aggcgcgaga cgacggcgat cggctgaccg 5040
 aggacgaact cgtcatgctc accggcgggc tgctcgtcgg cggccacgag acgaccgcca 5100
 gccagatcgc ctgcgagatc ttctcctcgc tgcgcgaccg gaccaggtag gagcaactcc 5160
 atgcccgtcc ggagttgatc ccacggcgag tcgaggaact gctgcgggtg gcccgcgtct 5220
 gggcctcggg cggccccacc cgcacgcgca ccgaggacct ggaactcaac gggacgacca 5280
 tccgggcccg cgacgccgtc gtcttctcgc tggcgtccgc caatcaggac gacgacgtct 5340
 tcgcgaatgc cgcagacgtc gtgctcgacc gcgaccgaa tccgcacatc gccttcggggc 5400
 acgggccccca ttactgcatc ggggcgtcac tggccagact ggaaatacag gccgccatcg 5460
 gcgccttggc caggcggtt cccggtctcc gcctggcgtc cgaggaaaac gaacttgatt 5520
 ggaacaaggg aatgatggta cgcagcctcg tgtcccttcc ggtgacgtgg tgaccgggcc 5580
 cgggcgcggg atcaggtgac gaacggatca gtacgcatt ggtcggccg gcccgagct 5640
 gacatggcct ggtgaagccg aaaaccatcg gccgcgcgc ctgcaagcgc ccgctgggac 5700
 gatgcgagtt gtgggcgcag acgcgtgcag cgcagccgtc cccgcgggac cgcggatggg 5760
 cttcccagca tcgttcttcg acccaggaga cctcatgacc gtgcagagtg acgtgttgcg 5820
 ccaccgcgat atcgccgtca tcgggatgtc ctgcccgtt cccggcgcg cagcatcga 5880
 ggaattctgg gacctgctgt gcagcgggcg gagcgcggtc gaccgccagc ccgacggcgg 5940
 ttggcggggc gtgatcgatg ggaagggaga atccgacgcc gcgttcttcg gcatgtcccc 6000
 gcgccaggcc gccgcggctc acccgcaaca gcgcctgatg ctggaactcg gctgggaggc 6060
 actggagaac gcccgcatcc ggcccgcga cctgaagggc tccgacactg gcgtcttcgt 6120
 ggggctcacc gccgacgact acgccacctt gctgcgcgc tccggcacgc ccatcagcgg 6180

gcacaccgcg acaggcctga accgtagcct cacggccaac cgtctctcgt acctgctggg 6240
 tctgcgcggc cccagcttca ccgtggactc cgcgcagtcg tcatccctgg tcgccgttca 6300
 cctggcgtgc gaaagcctgc tgcggggcga gagcgcggtc gccgtcgtcg gcggggtgag 6360
 cctcatcctg gcagaggaga gcaccgccgc catggcgcgt atgggggcac tctctcctga 6420
 cgggcgttgc ttcaccttcg acgcccgggc caacggctac gtccgtggcg aggggtggcgt 6480
 ggccatggtc ctcaagccgc tgatccgcgc gatcgaggac ggcgaccagg tgcaactgcgt 6540
 catccggggc tgtgccgtca acaacgacgg cgggtggccc agcctcacc atcccgaccg 6600
 ggaggcccag gaggcattgc tgcgccgggc gtacgagcgg gcgggggtgg cccccaaca 6660
 cgtcgactac gtcgagctgc acggcaccgg gacgaaggcc ggcgacccc tcgaggcggc 6720
 ggccctcggg gcgggtgctgg gtgtcgccc cggctgcgac aaccactcg cggtcggatc 6780
 ggtcaagacc aacgtcggcc acctggaggg gcgggccggc atcacgggcc tgctgaaggc 6840
 ggtgctgtgc gtacgtgagg ggtgctgcc gccgagcctc aacttccgta cgccgaaccc 6900
 ggacatccgc ctgcagagc tgaacctccg ggttcagacg gaactgcagc cgtggccggg 6960
 cgacgggacg ggccgcccgc gtgtcgccgg agtgagttcc ttcggcatgg gcggtacgaa 7020
 tgcgcatctg attctcgagc aggtccgggt ggcggctgag gaaacggctg ttaccgatgc 7080
 cgggtgctgg tccggttcggg tggttccgggt ggtggtgctg ggtcgttcgg tgggggcttt 7140
 gcgggctat gcgggtcgggt tgcgtgaggt gtgcgcgggg ttgtctgacg gtggtggctc 7200
 cgggtggtgg tctggtctgg tggatgtggg ttggtcgttg gtgtcgtcgc ggtcgggtgt 7260
 cgagcatcgg gcggtcgtgt tcgggtgggg tgtcgccgag gtggtggcgg gtttggtatgc 7320
 ggtggcttct ggggcgggtga gttcgggttc ggtggtggtg ggttcgggtg cgtcgggtgt 7380
 tgctggtggt ggtggtcggg tgggtgtttgt gtttcgggt cagggttggc agtggtggg 7440
 tatgggtgcg gctctgttgg acgagtcgga ggtgtttgct gagtcatgg tggagtgtgg 7500
 gcgggcttg tcggggtttg tggattggga tttgttgga gtggtccgc gtggtggggg 7560
 tgacggatcg tttggtcggg ttgatgtggt gcagccggtg tcgtgggcgg tgatggtgtc 7620
 gttggcgcgg ttgtggtgt cgggtgggtgt ggtgccggat gcggtggtgg gtcattcgca 7680
 ggggtgaggt gctgcgccgg tgggtggggg tgtgttgagt gtggtgatg gggcgcggt 7740
 ggtggcgttg cggtcgcggg tgatcggtga ggtgttgagg ggtggtggtg cgatggtgtc 7800
 ggtggggttg ccggtggcgg ttgtgttgga tcggttgagg ggtgggggtg gtcggttggg 7860
 tgtggcggcg gtgaatggtc cgtcgttgac ggtggtgctg ggggatgtgg atgctgctgt 7920

ggggtttgtt ggtgagtgtg agcgggatgg ggtgtgggtg cggcgggtgg cggaggatta	7980
tgcgtcgcat tcggcgcatg tggaggcggg ggaggggatg ctgtcggggg tgttgggtgg	8040
tttgtgtccg gggcgggggtg tggtgccgtt ttattcgtcg gtggtgggtg gtgtgggtga	8100
tgggggtggg ttggatgggt ggtattggta tcggaatctg cgtgagcggg tgttgttttc	8160
ggatgtgggt gggcggcttg ttggggatgg gttttcgggg tttgtggagt gtccggggca	8220
tccgggtgtt gcgggtgggg tgttggagtc ggtggcgggt gtggatccgg atgtgcggcc	8280
ggtgggtggg gggcgcgtgc gccgtgatga tgggtgggtg ggccggtttt tgacgtcggg	8340
gggtgaggcg ttcgtcggcg ggatgagtgt tgactggaag ggtgtgttcg cgggggcggg	8400
cgcgcgggtg gttgacctgc cgacgtatcc gttccaacga cgccactact gggcaccgaa	8460
caccgacggc gcgccagctc cgatcctcga tgatcacgcg gaggcggaga acgaaccagc	8520
cgaatccgag ccagggtatc gggccgagct tctgacgttg gccgagcccg agcaactgaa	8580
ccgactcttg gcgaccgttc gcgccagcac cgccgtcgtt ctgggcctcg actcggcgca	8640
ggcggtcgat ccggagcgca cgttcaagga gcatggattc gaatcgggtc ccgccgtcga	8700
gctctgtaac cacctgcaac gcggcactgg gctgcgggtt ccgcctcgc ttgtatacaa	8760
ccatcccacc ccgatggccg ctgcccggaa gctgcaggaa gaaattcagg gccggcaacc	8820
ggagaacgtc cggcagggtc cctccgctgc tgcgtggat gatccgggtg tgggtgggtgg	8880
gatgggtgt cgttttccgg gtgggggtgt gtgtgcggag ggtttgtggg atttggtgtt	8940
gggggggtgg gatgcgggtg cggggtttcc ggtggatcgg ggttgggatg tggaggggtt	9000
gtttgatccg gtgcgggggt tgggtgggaa gtcgtatgtg cgggaggggg ggtttgtgta	9060
tgacgcgggg atgttcgatg cggagttttt tgggtgtgtc ccgcgtgagg cggtggcgat	9120
ggatccgcag cagcgtttgt ttttggaggt gtcgtgggag gcgttggagc gtgcggggat	9180
tgatccgttg ggtttgcggg gttcgcggac ggggtgtgtat gtgggggtga tgggtcagga	9240
gtatgggccc cggttgggtg agtcgggtgg tgggtttgag ggttatttgt tgacggggac	9300
gtcgcagagt gtggtgtcgg gtcgtgtttc gtatgtgttg gggttggagg gtccgtcgat	9360
ttcggttgat acggcgtgtt cgtcgtcgtt ggtggcgttg catttgccgt gtcaggggtt	9420
gcggttgggt gagtgtgat tggcgttggc ggggtgggtg acggtgattg cggcgccggg	9480
gttgtttgtg gagttttctc ggcagggtgg gttgtcgggt gatgggcggg gtccggcggt	9540
tgcgggtggg gcggatggga cggggtgggg ggagggtgcg ggggtgggtg tgttggagcg	9600

gttgtcggtg gcgcgggagc gtggtcatcg ggtgttggcg gtggtgcggg gttctgcggg 9660
 gaatcaggat ggtgggtcga atggtttgac ggcgcgctcg ggggtggcgc agcgtcgggt 9720
 gattggtgcg gcgttgggtg cggcgggttt ggggtgtgctg gatgtggatg tggaggagggc 9780
 gcatgggacg gggactcggg tgggtgatcc gattgaggct gaggcgttgt tggggtcgta 9840
 tgggcggggg cgtgtgggtg gggcgttgtt gttgggttcg gtgaagtcga atattggtca 9900
 tacgcaggcg gctgcgggtg tggcgggtgt gatcaagatg gtgatggcgt tgcgggcggg 9960
 ggtggtgccg gcgacgttgc atgtggatgt gccgtcccg ttggtggatt ggtcttcggg 10020
 tggggtggag ttggtgacgg aggcgcggga ttggccgggtg gtgggtcgtg tgcgtcgtgc 10080
 ggggtgtgctg gcgtttgggg tgcggggac gaatgcgcac ctgattttgg agcaggcccc 10140
 cgaattcgac gatccggtt ttaccgacac cgacaccgat gctggtgtgg gtaggggtct 10200
 atcgggtggt ccggtggtg ttcgggtcg ttcgacggcg gctttgcgcg cttatgcggg 10260
 ccggttgcgt gaggtgtgcg cgggtctttc cgatggtgcc ggtctggtga atgtgggttg 10320
 gtcgttgggtg tcgtcgcggg cgggtttcga gcatcggcg gtcgtgttg gtgggggtgt 10380
 cgccgagggtg gtggcgggtt tggatgcggg ggtttccggg gcggtggctt cgggttcggg 10440
 ggtggtgggt tcggtggcgt cgggtgttgc tgggtggtgg ggtcgggtgg tgtttgtgtt 10500
 tccgggtcag ggttggcagt ggggtgggtat ggggtgcggc ctgctggacg agtcggagggt 10560
 gtttctgag tcgatggtgg agtgtggtcg ggcgttctcg gggtttgtgg attgggattt 10620
 gttggagggtg gtgcgggggtg gggcgggtga ggggggtgtg ggtcgggttg atgtggtgca 10680
 gccggtgtcg tgggcgggtga tgggtgtcgtt ggcgcgggtt tggatgtcgg tgggtgtggt 10740
 gccggatgcg gtggtgggtc attcgcaggg tgaggttgct gcggcgggtg tgggggtgtg 10800
 gttgagtgtg gctgatgggg cgcgggtggg ggcgttgcgg tcgcgggtaa ttggtgagggt 10860
 gttggccggg ggtggtgcga tgggtgtcgg cggactgccg atcgtggatg cgcaggaaacg 10920
 gttggcgggg tggggtgggtc ggttgggtgt ggcggcgggtg aatggtccgt cgttgacggg 10980
 ggtgtcgggg gatgtggatg ctgctgtggg gtttgttggg gagtgtgagc gggatggggg 11040
 gtgggtgcgg cgggtggcgg tggattatgc gtcgcattcg gcgcatgtgg aggcgggtgga 11100
 ggggatgctg tcggggttgt tgggtggttt gtgtccgggg cggggtgtgg tgccgtttta 11160
 ttctcgggtg gtgggtgggt tgggtgatgg ggtgggtttg gatggtgggt attggtatcg 11220
 gaatctcgt gagcgggtgt tgttttcgga tgtggtgggg cggcttgttg gggatgggtt 11280
 ttccgggttt gtggagtgtt cggggcatcc ggtgttggcg ggtgggtgt tggagtcggg 11340

ggcggtggtg gatccggatg tgcggccggt ggtggtgggg tcgctgcgcc gtgatgatgg 11400
 tgggtggggc cggtttttga cgtcgggtggg tgaggcggtc gtcggcgagg tgagtgttga 11460
 ctggaagggg gtgttcgcgg gggcgggcgc gcggttggtt gacctgccga cgtatccgtt 11520
 ccaacgccgc cactactggg caccgactcc caccaacccc gccaccaacc ccgccacggg 11580
 cgacaccacc accgccgacc cgggtgggtgg cgtgcggtat cggatcacct ggaaaccgtt 11640
 gccgacggac gacccccgac ccctcaccaa ccgctggcta ctcatcgccg acccggggac 11700
 cgccggctcg gagcttgccg cagacatcac agcagcgctc attcgagggg gcgccgaggt 11760
 cgagttgctg gccgtggacc cgctcgcggg ccggggcccgg atcgccgaac tgctcgccac 11820
 cacgacggct gggccgggtgc cgctgtcggg cgccgtgtct cttctcgggc ttgtgcagga 11880
 cgcgcatcct caacacccct ccatcggaat gggcggtgtc tcgtcgctgg cgctggtgca 11940
 ggccatcggg gacgcgggag ccgagactcc tttgtggagc gtcacgcagg gggcggtcgc 12000
 tgtggtgccc caggaggcgc cggatgtgtt cgggtgcgag gtgtgggctg tcgggcgggt 12060
 ggccgccctg gaactgcccg accgctgggg cggcctggtc gaccttcctg ccgtaccgaa 12120
 tgcccggatg ctggaccagc tcgccaacgc cctcgccgga gcggacggcg aggaccagat 12180
 cgcggtacgc ggctcgggga tctacgggag tcgggtgacg cgcgggcgcg gcaactgcgcg 12240
 ccgggaatgg cgccctcgcg ggaacatcct ggtgaccgga ggtacgggaa gtctgggtgg 12300
 ccgggtggcc cgggtggctg ctcgcaacgg tgccgaacac ctcgttctca ccagtcgtcg 12360
 ggggtgccgac gccccggggg cggcagaact ggaagctgat cttcgcgcgc tcggtgtcga 12420
 ggtgaccatg gccgcctgcg atgtagcgga ccgggtgcg ctgtccgacg tcctggcggc 12480
 gcatccgcc actgcggtct tccacacgc cggagtcctg cacgacggtg tgatcgacac 12540
 gctcgccgcc ggacacatcg acgaggtctt ccgtccgaag accgctgccg cgctgctgct 12600
 cgacgaactc acccagcacc aggagctgga cgccttcgtc ctcttctcat cggttaccgg 12660
 agtctggggc aacggcggcc aggcggcgta cgcgggcgcg aacgcacgc tggacgccct 12720
 ggcggagcga cgtcgtgccg caggtcttcc cgccacctcc atagcttggg gactgtgggg 12780
 cggcggtggc atggcgagg ggatcggcga gcagaacctg aaccgccgtg gcatcacggc 12840
 cttggacccg gagctcggca tcgccgtct gcagcaggcc ctcgaccgcg atgacgtgtc 12900
 tgtcacctgc gccgacgtcg actggacggg ttctgctccg cgtcttgccg acctgcgctc 12960
 gggcgggctc ttcgacgggg tgcccaggcc caggagcgcg ctcgatgccc ggaaagtgga 13020

caccgagtcg ccgagcgccg gccttgcgca gcgggtggcg gggatgcccg acgcggaacg 13080
gcagcgggtc ctccctgaaa cgggtgcgggc ggccggccgcg gcggtcctga ggcacgagac 13140
ggatggatgag gtcgcgcccc cccggggcctt caaggacgac ggcttcgact cgctcacggc 13200
gctcgaactg cgcaaccacc tcaacagcac gaccggtctg agtctgcctc cgacgggtgt 13260
cttcgaccac cccaccccggt ccacgttggc gaagtctctg gagggcgctc tcgtcggcgc 13320
ttctgccgag gaagtcccgg tgactgccgc agccgtgccc gtcgacgagc ctattgccat 13380
cgtcggcatg gcctgccgct accccggcgg agccgacact cccgagaagc tctgggacct 13440
cctgctggcc ggtgctgacg tcatcgccc agccccgac gaccggggct gggacgtgga 13500
ctccttcttt gatcccgctg cgggcgcccgc ggggaagtcg tatgcgcggg agggggggtt 13560
tgtgtatgac gcggggatgt tcgatgcgga gttctttggt gtgtcgccgc gtgaggcgg 13620
ggcgatggat ccgcagcagc gcttgttgtt ggagacgtcg tgggaggcgt tggagcgtgc 13680
gggaatcgat ccggcgggtc tgcggggtag ccggaccggc gtgtactccg gcctgaccca 13740
ccaggagtat gccgccgctc tgcacgaggc tccgcaggaa ctcgagggt atctgctcac 13800
cggcaagtcg gtgagcgtcg cgtcgggtcg tgtttcgtat gtgttggggg tggagggtcc 13860
gtcgatttcg gttgatacgg cgtgttcgtc gtcgttggtg gcgttgcat tggcgtgtca 13920
ggggttgcgg ttgggtgagt gtgatgtggc gttggcgggt ggggtgacgg tgattgcggc 13980
gccgggggtt tttgtggagt tttctcgga ggggtgggtt tcgggtgatg ggcggtgtcg 14040
ggcgtttgcg ggtggtgcgg atgggacggg gtggggggag ggtgcggggg tgggtggtgt 14100
ggagcgggtt tcggtggcgc gggagcgtgg tcatcgggtg ttggcgggtg tgcgggggtc 14160
tgcgggtgaat caggatgggt ggtcgaatgg tttgacggcg ccgtcggggg tggcgcagcg 14220
tcgggtgatt ggtgcggcgt tgggtggcgg gggtttgggt gtgtcggatg tggatgtggt 14280
ggaggcgcac gggacgggga ctcggttggg tgatccgatt gaggctgagg cgttggttggg 14340
gtcgtatggg cggggtcgtg tgggtggggc gttgttgttg ggttcggtga agtcgaatat 14400
tggtcatacg caggcggctg cgggtgtggc ggggtgtgat aagatggtga tggcgttgcg 14460
ggcgggggtg gtgccggcga cgttgcatgt ggatgtgccg tcgccgttgg tggattggtc 14520
ttcgggtggg gtggagtgtg tgacggaggc gcgggattgg ccggtggtgg gtcgtgtgcg 14580
tcgtgcgggt gtgtcggcgt ttgggtgtgc ggggacgaat gcgcactga ttttggagca 14640
ggccccgag ttcgacgatc ctgccgattc cgattccgat tccgattccg attccgatgc 14700
cgggtgtcgt gatggcggcg aggggtggtg tggcaggagc ttgtcgggtg ttccgggtgt 14760

ggtgtcgggt cggtcgggtg gggctttgcg ggcgtatgcg ggtcgggtgc gtgagggtgtg 14820
 cgcgggggtt tctgacgggt gtggctccgg tgggtggttct gggttggtgg atgtgggttg 14880
 gtcgttggtg tcgtcgcgggt cgggtgttga gcatcgggcg gtcgtgttcg gtgggggtgt 14940
 ggaggagggt gttgctgggt ttggtgcgggt ggcttctggg gcgggtggctt cgggttcgggt 15000
 ggtggtgggt tcgggtggcg cgggtgttgc tgggtggtgg ggtcgggtgg tgtttgtgtt 15060
 tccgggtcag ggttggcagt ggggtgggtat ggggtgcggcg ctgctggacg agtcggagggt 15120
 gttcgcagag tcgatgggtg agtgtggtcg ggcgttgcg gggtttgtgg attgggattt 15180
 gttggagggt gtgcgcggcg gggcgggtga gggggtgtgg ggtcgggttg atgtggtgca 15240
 gccgggtgtc tgggcgggtga tgggtgtcgtt ggcgcgggtt tggatgtcgg tgggtgtggt 15300
 gccggatgcg gtggtgggtc attcgcagggt tgaggttgct gcggcgggtg tgggggggtgt 15360
 gttgagtgtg gctgatgggg cgcgggtggt ggcgttgcg tcgcgggtga tcggtgagggt 15420
 gttggccgggt ggtggtgcga tgggtgtcgtt cggactgccg atcgtggatg tgcaggaacg 15480
 gttggcgggg tggggtgggt ggttgggtgt ggcggcgggt aatggtccgt cgttgacgggt 15540
 ggtgtcgggg gatgtggatg ctgctgtggg gtttgttggg gagtgtgagc gggatgggggt 15600
 gtgggtgcgg cgggtggcgg tggattatgc gtgcattcgc gcgcattgtg aggcgggtgga 15660
 ggggatgctg tcggggttgt tgggtggttt gtgtccgggg cggggtgtgg tgccgtttta 15720
 ttcgtcgggt gtgggtgggt tggttgatgg ggtgggtttg gatggtgggt attggtatcg 15780
 gaatctgcgt gagcgggtgt tgttttcgga tgtggtgggg cggcttgttg gggatgggtt 15840
 ttccgggttt gtggagtgtt cggggcatcc ggtgttggcg ggtggggtgt tggagtccgt 15900
 ggcgggtggt gatccggatg tcgggcccgt ggtggtgggg tcgctgcgcc gtgatgatgg 15960
 tgggtggggc cggtttctga cgtcgggtgg tgaggcggtc gtcggcggga tgagtgttga 16020
 ctggaagggt gtgttcgcgg gggcgggcgc gcgggttggt gacctgccga cgtatccgtt 16080
 ccaacgacgc cactactggg ccagacctc gcccgctggc gtcgggacgg ccgcggcggc 16140
 ccggttcggc atggagtggg aggaccatcc cctgctcggc ggtgcgctgt cggtcggggg 16200
 ctccaggagc ctgcttctgg ccgggcatct gtcgctcgcc tcgcacgcct ggctgaccga 16260
 ccatgcggtc tccggcaccg tgctgctgcc cggtaaggcc ttctgtggaac tcgccctgca 16320
 cgccgcgct gcggctgggt gtccggaggt cgaggagctg cggctggagg ctcccctggt 16380
 ggtgccggcc aggggcgggg tgcggctcca ggtgctcgtg gacgaccccg acgacggatc 16440

cgaccgccgc gcggttaagcg tgttctcccg ggacgatgcg gcgccggccg agtccgcctg 16500
 gacgcggcac gcggtgggcg tcctggccgc gcggtcgcg cctgcaccgg ctgcgccctg 16560
 gcacaccgac gcctggccac cttcgggcac ggagccggtc gacgtggccg acctgtatga 16620
 gcggttcgcg gcgctgggct acgagtacgg ggaggcggtc gccgggctcc aggggggtctg 16680
 gcggggggac ggcgaggtgt tcgccgaggt gcggctgccc gaccgggtca gcgcggaggc 16740
 cattcgcttc gggctgcatc ccgcgctgct cgacgccgcc ctgcaggggt ggttggcggg 16800
 cgacctcgtc ggcgtccccg agggcagtgt gctgctgccc ttgcctggc agggcgctcg 16860
 gctccacgcc accggcgccg aactctgctg ggttcgcatc ggccgggtcc gtgactcggc 16920
 cgtctgcctg cacgcggtgg acccggccgg tgetccggtc ctctcggttg acgccctggc 16980
 cctgcgtccg ctgctccggg aacgcctcgg gctgcccgcc gatgccggag ccggggcggt 17040
 gtaccgggtc ggctggcggc ggcaggccgc cgttgccggg gcagccgacc ggcggtgggc 17100
 ggtcgtggcc ccgaacggtg ccgaggcgga cggggccgcc gagccgcacc ggtggccggt 17160
 cgccgccgtc gacgtgcaca ccgacgtgga ctgctgctg gcggccctgg acgcgggcgc 17220
 ggaactgccc gccgtcgtec tcgccgaact ccggagggcc gccggctgga gcgtcgacag 17280
 ttgctggtcc gccggcccgt cgcccaacga cggcgcggtg ggcgacggcg cggtgggcga 17340
 cgcccgggcc ggggcccgtc gggcggcgac ccgggcccgg ctggatctgc tgcaacgctg 17400
 gctggccgac gagcggttca tcgcggccag gctcgtggtg gtcaccgaac gggccgtggc 17460
 cgccggggcg gacgaggacg tgccgggcct cgtccacgcg ggactgtggg gcctgctccg 17520
 gtcggcccaa tcggagcacc cggaccgctt cgtgctggtg gacgtcgacg cggacgacag 17580
 ctgctcgcg gcgctgccgt cggccctcgc catggacgcg cccaactgg tggcgcgggc 17640
 cggtcagatc ctgctgcccg agatcgagcc ggtgcggccc gtacccgagc cggagcaggc 17700
 ggaacccgaa ccggggggcg tcctggaccc cgacggcacg gtctgtctca ccggcgcgac 17760
 cggcacgctc ggcgggctgc tcgcccggca cctggtgacc acccgtggtg cgcgccggct 17820
 gctgctggtc agccgcagcg gtccggacgc cccgatgcc ggccggctga ccgaggagct 17880
 gaccgggctc ggcgcccacg tgacgctggc cgctgcgac accacggatc gcgccgcgct 17940
 ggccggcgtc ctgggcgga tccccccga gcatccgctg accgccgtgg tgcacgtggc 18000
 cggcgctactc gacgacgggg cgtgcaggc gtcaccccc gagcgggtcg acgcggtgct 18060
 ccggccgaag gtggacgagg cactgcacct gcacgaactg accgcggggc tgccgctggc 18120
 cgcgctcgtg ctgttctccg gggcggcggg gatcctgggc cggcccggcc aggccaaacta 18180

cgccggcggcg aacaccttcc tggacgccct ggccgcagcac cgacggggccc ggggcctgcc 18240
 cggcgtctcc ctgcctggg gcctgtgggg gctggccagc gacatgacgg gccacctggg 18300
 cgagcaggac ctgcggcgga tgcggcgctc cggcatcgcc ccgatgaccg gcgaggaggg 18360
 cctcgcgctg ttcgacctgg ccctcgacct ggcccgggac gaaccggtgc tcgtaccggc 18420
 ccgactggac ccggcggcgc tgcgccggga gtgggcccgc aacggaccgg gcgccgtccc 18480
 ggtcctgctg cggggtctgg tgccggcggc tccgctccgt cgcgcggccc cgtcggcgcc 18540
 cgccggcggg gcgcccgtgc ccgccgtcgc cgcgccgcag caggcggacg agctgcgcgg 18600
 gcaactggcc gggaaggacg cgcaggccca ggtccggcag ctgctggatc tggtagcgcc 18660
 ccatgtcgcc ggggtgctcg ccctccggga agcggcggac gtggaccggg gcagaccgtt 18720
 ccgcgaggtc ggattcgact cgttgaccgc agtcgaactg cgcaaccggc tgggctcggc 18780
 gaccggcctg cggttggcac cgagcctggg gttcgaccat ccgaccccggt cggccgtggc 18840
 cgagcacctc gtggaccgcc tcgccgccga gggggcggct gacgaggggc cggcggcact 18900
 gaccgggctc gacgcagtgg ccgcggcgct cggcgggatg cggacggacg acgttcgccg 18960
 ggacatcgtc cgcaggcggc tggaggagat gctcgccctg gtcggcgggc cacggtccgg 19020
 gccggcaggt gacgggctgg tggatgccac ggtcgccgag cgaactggact cggcttccga 19080
 cgacgaactc ttccgctga tcgaggagca gctgtgaacc ccgaccgagg agagggccgg 19140
 caggtgaccg cgaacgagga ccggatgcgt gaggacctca agcgggtcac cgcgagctg 19200
 gccgggacgc ggcgacgcct gcgcgagctg gaggacagcg cgcgtgagcc catcgcgac 19260
 gtgggcatga gctgccggtt gccggcgggg gtgagcacgc ccgaggacct gtggcggtg 19320
 gtcgaggccg gtaccgacgc gatctccggc ttccccgacg accggggctg ggatgtcggg 19380
 aggtctacg acccgatcc ggactcgacc ggaacgagct acgtgcgcga gggcggttc 19440
 ctctacgact gcgccgagtt cgaccggag ttcttcaccg tctcgccccg cgaggcgctg 19500
 gccatggacc cgcagcagcg gctgctgctg gaggccgcct gggagacctt cgaacgggcg 19560
 gggatcgccc ccgactcggc ccgcggcacc cgcaccgggg tctacgtcgg ggtgatgtac 19620
 gacgactacg gcagccggct gtcggaggtg ccgaaggacc tggagggcta cctggtcaac 19680
 ggcagcgcg gcaagtgcgc gtcgggccc atcgcgta ca cgtgggggtt gcaggggccc 19740
 gcggtgacgg tcgacacggc ctgctcgctg tcgctggctg cgttgcacct ggccgtgcag 19800
 gcgctgcggg cgggcgagtg tgagctggcc ctggcgggcg gggcgacggt gctcgccacg 19860

ccgacgatgt tcgtcgactt cgcccggcag cgcggtctcg ccgaggacgg ccgttgcaag 19920
 gcgttcgcgg acgccgccga cgggaccggg ttccggcgagg gcgtggggat gctgctggtg 19980
 gaacggctct cggacgcggt ccgcaaccgt cgccagggtgc tggccgtcgt gcggggcagc 20040
 gcggtcaacc aggacggggc gagcaacggc ctgaccgccc cgaacgggtac ggcccagcaa 20100
 ctggtcatcc ggcaggcggt gaccaacgcg gggctggccg cggacgaggt ggacgcggtg 20160
 gaggcacacg gcaccggcac ccggctgggc gatccgatcg aggcgcaggc gctgctggcg 20220
 acgtacggcc agggccggcc ggcggaccgg ccgctcctgc tgggatccct gaagtccaac 20280
 atcggccaca cccaggccgc cgcaggggtc gccgggggtga tcaagaccgt gctggcgctg 20340
 cgtcacgcgc ggtgccccg gacctgcac gtcgatcgcc cctcgaccg ggtggactgg 20400
 tcgtcgggcg cgggtgcggt gctgaccgag ggcggccct ggcccgatca cggcgaccgg 20460
 ccccgccggg ccggggtctc ctggttcggc gcgagcggca ccaacgcgca cgtcatcctg 20520
 gagagcggcc ccggtgcggc ggcggggggc accggggcga cggacctctc ggccccgccg 20580
 gcatccgtcg cccaccatcc ggccacggcc acggccacgg ccccgggcggc gacggtgccc 20640
 actgcccacg aaccggcggg gacggccggc gacgaccccg tctgggtcct gtccggccgg 20700
 accgaggcgg ccctgcgcga gcaggcccg cggtacacg cccacctgac atccggggcg 20760
 cggcccgagc ccgccgacgc cgtggcccg gcgctggcgc gctcccgac cgcgttcgcg 20820
 taccgggccc ccgtgctggg ccgggacgac accgcgcggc tcgacggcct ccacgcgctc 20880
 gcggcgggtc gcagcgccgc ggggctcgtc accgggcggg ccgtgccgga gcggcgcgctg 20940
 gccttcctct tcaccgggca gggcagccag cgaccggggc cgggcccggga actgtacgcc 21000
 cggcatcccg ccttcgcaca ggccctggac ggctcctcg cggaaactga ccggcacctg 21060
 gaccggccgc tgcgcgccgt catgctcgcc gagccgggca ccgaggcggc ggcgctgctg 21120
 gacgacaccg cgtacacca gcccgccctg ttccgctgg aggtggcgct gttccggctg 21180
 gtcacgagct gggggctgcg gcctgacgcc ctgctgggcc actcggtcgg ggagatcacc 21240
 gcggcgtagc tcgcggcggt cctcaccctg ccggacgccg ccgggtggt ggcggtgcgc 21300
 ggtcgactca tggcggaact gcgggcccgc ggtgcgatgg ccgcgctcca ggccgccgag 21360
 agcgaggtcg accccctggt ggcggggcgg gagggcgaac tgtcgatcgc agcggtaaac 21420
 gggccgcagg caaccgtgat cgcgggcgac gaggcggccg tcgaggagca ggtcgcgctg 21480
 tggcgtgacc ggggtgcgcg ggccaggcga ctgcgggtcg gccacgcctt ccaactccgta 21540
 cggatggacg ggatgctcgc cgagttcgag aaggcgatgg gtgatctccg tgccggcgag 21600

ccgacgatcc cctgtggtcgc caacgtcagg ggggagatcg cgtccggcac cgacctccgt 21660
acggccgggt actggatccg gcacgcccgc gagccggtgc gtttcctcga cggcatgcgt 21720
gcgctgcggg ccgagggcgt cgacacgttc gtggaactcg gccccgacgg agtgctcacg 21780
gcgatggcgc gcgactgcct ggcgatccc gccgaccgg tggatctcgc ggacgccgcc 21840
gagcccgccg gggccgcgga gcccgaccgc tccctgctgt tccctgccac cctgcgccgg 21900
gaccgcgacg acgcagtggc cgtgcgggag gccctggcat ccgtccacgt gcacgggctt 21960
cccgtcgacc cggtcgcgcc gctcggcgac gggccgctcg ccaccgacct gccacctac 22020
ccgttccagc ggtcccgccta ctggctcgac ccggtcccg gggcacgcga cctgaccgcc 22080
gtgggcctcg acgtggccgg gcacccgctg ctgcgcgtcg ccgtggacct gcccgacggc 22140
gccggcacgg tctggagcgg tcagctctgc gtgcggacgc atccgtggct cggcgaccac 22200
agcgtgtggg ggcgcacggg ggtgccgggg accgcgctgc tggagatcat gcaccgagt 22260
cgcgccgagg tgggctgcac ccgggtcgcg gaactgacct tcgaggcgcc gatggtgctg 22320
gccgacgacg ggggcgtccg cgtgcgggtc gtcgtcgacg gaccagacgc cgacggggcc 22380
cgccaggctc ggatccactc cgcaccggtg gggcccgagc ctccccactg gaccggcac 22440
gcctcgggcc gcgtcgacag cgccgcgccg gggccggccg ccggcccacc cgcggtggac 22500
gccggccctg gcagcaactg gccgcccag ggggcggagc cgggtggcgt cgagagcgag 22560
tacgagcgct tcgccgacaa cggcatcgga tacggccccg ccttccgagg gctgcgcgcc 22620
gcgtggcgtc gcgggaacga gacgttcgcc gaggtccggc tccccgagg gtacgccgcc 22680
gaggcggggc actacgccgt ccatccggca ctgctggacg cggccctgca cgcatcgtc 22740
ttcgtgacc agtttcccgg tggggcacac gggatgctgc cgttcgcctt caccgacgtg 22800
cgggtgttca gctccggcgc cgaccggctc cgggtgcgca tcgcgcccgc cgatgccgac 22860
tcggtctgcg tgaccgtcgc cgacggcgac gggacgcgg tccctcgccgc agccacctg 22920
gcgttgccgc gggtcgccgc cgaccgatc gcggcgacc tcaccggcca ggcaccgctg 22980
taccggttg agtggtccgc cgtgcggccc gcccgggtgg ccaccggggc gcggttcgcc 23040
gtcgtcgccg cggacgcccc gctgccgtcc ggtgcgctgg gggccggggg gcccgccag 23100
gcgtaccgg acctgggcgc gctggccggc gcgttgccca ccaacggggc accggggccac 23160
gtgctcgctc acttccgccg ccgcgccgac gggccggcag ggcggcagcc cggtgacgtg 23220
ggtgcacgga cccgacgggc gctggccgtc gtccaggagt ggctcgccga cgaccgtttc 23280

accggctcac ggctggtcgt gtcaccagc ggagccgtgg acgccggaac agccgtcacc 23340
 gatccggccg ccgccggggt gtggggcctg ctgcgggtcg cccagaccga gcatccggac 23400
 cggttcgtcc tcgtggacac cgacgaccac ccggattcgc tgcgtgccct ccccggggcg 23460
 atcgttgccg gcgagccgca gctggcactg cgggccggca cggccagcgt tccgggcctg 23520
 gtgcgggtgc cggccggcac cggtgccgcc ccgccgtggg ccgcagccgg caccgtcctc 23580
 gtcaccgggg gcaccggcat gtcggcgccg gcgggtggccc ggcacctggt ccgccggcac 23640
 ggggtccgcc gcctgctgct ggtcggcccg cgcgggcccg acgcacccgg cgcggcggcc 23700
 ctgacccggg aactggagga gctgggagcg tccgtccgcg tcgccgcctg cgacgtcggc 23760
 gatcgtggcg cggtgacgcg cctgttgccc ggggttccc ccgcgcaccc gtcaccgcg 23820
 gtggtgcaact cggccggcct gcccgacgac ggcgtgctga ccgcacagac cggcgagcgg 23880
 gtcgcggcgg tgctccgcgc caaggcggac gcagcggta acctgcacga actcaccgg 23940
 catctcgacc tcaccgcctt cgtgctgttc tcgtcggtag cggggaagat cggcagcgcc 24000
 gggcaggccg ggtacgccgc cgcgaacgcc ttccctcgac cgcttcgcgag ctggcggcag 24060
 ggccaggggc tgcccgccac cgccctggcg tgggggcccgt tggacggcg gatggccgcc 24120
 ggcctcggca ctgcggacgt ggcaaggctg cgccggtccg ggctcgtgcc gtcggcgctg 24180
 gacgacgcgc tcgtttctct cgacgccgcc tgctcccgac cggcgggcgg gtaccacccc 24240
 gtccgcctcg atccggcggt gctgcggctc cacgccccg ccgacagcgc ggtgcccgcc 24300
 gtccctgctg gtccgagccg tgcgcacccg agggacggta cgccggggaa gcctgccgaa 24360
 gccgccctcg ccgcgtgct gaccggcagg tcggcgggcc agcgtacggc gatcctgacc 24420
 gacctggtgc ggacggaggg cgccgcgctt ctcgggcatg gcgaggcggc gatgctgagc 24480
 acgcagcggg ccttccgcga cgccggcttc gactcgtca ccgccgtgga cctccgcaac 24540
 cggctcggcg cggccacggg cctcagcctg ccggccgccg tcgtcttcga ccacccgacc 24600
 ccggcgggccc tggccgccta tctgcggacc gaactggacc gccggtcgcc caccgggcaa 24660
 cagttcccga cggacgccgc cgggtgtctg gccatgctcg accgcctcg ggacggaatc 24720
 gcgacggctg tcagggacga cgccgaccgg acccgcgag ccgacctgt gcgtgtcctg 24780
 ctgcgccagg tcggcgggcc cgggacgggc ccgccccgc acaccgacgg cggctccggc 24840
 ggcgaggtca gcgaccgct ccggaccgcc tccgacgagg aactgttcga cctgctcgac 24900
 agcgatttcc gactggcgta gcgccggccg gagcactgcc cgctcgaatc gaccgacccc 24960
 gggaagacac tcggatcaca gggggaagcg ccgtgtctgt caacaacgaa gacaagcttc 25020

gcgagtatct gcgtcgtgcc atggcggatc tccatgagtc ccgcgagcgg ttgcggcagt 25080
 acgagtcgcc tgctgctgtg gatgatccgg tgggtggtggg ggggatgggt tgcgttttc 25140
 cgggtggggg ggtgctgtcg gagggtttgt gggatttggg gttggggggg ggggatgcgg 25200
 tgcgggggtt tccgggtgat cggggttggg atgtggaggg gttgtttgat ccggtgcggg 25260
 gtgtgggtggg gaagtcgtat gtgcgggagg gggggtttgt gtatgacgcg gggatgttcg 25320
 atgcggagtt ttttggtgtg tcgccgcgtg aggcgggtggc gatggatccg cagcagcgtt 25380
 tgtttttgga ggtgctgtgg gaggcgttgg agcgtgcggg gattgatccg ttgggtttgc 25440
 ggggttcgcg gacgggtgtg tatgtggggg tgatgggtca ggagtatggg ccgcggttgg 25500
 tggagtcggg tgggtgggtt gagggttatt tgttgacggg gacgtcgccg agtgtggtgt 25560
 cgggtcgtgt ttcgtatgtg ttggggttgg aggggtccgtc gatttcgggt gatacggcgt 25620
 gttcgtcgtc gttggtggcg ttgcatttgg cgtgtcaggg gttgcggttg ggtgagtgtg 25680
 atgtggcgtt ggcgggtggg gtgacggtga ttgcggcgcc ggggttggtt gtggagtttt 25740
 ctccgcaggg tgggttgctg ggtgatgggc ggtgtcgggc gtttgcgggt ggtgcggatg 25800
 ggacggggtg gggggagggt gcgggggtgg tgggtttgga gcggttgctg gtggcgcggg 25860
 agcgtggtca tcgggtgttg gcggtggtgc ggggttctgc ggtgaatcag gatggtgggt 25920
 cgaatgggtt gacggcgccg tcgggggtgg cgcagcgtcg ggtgattggt gcggcgttgg 25980
 tggcgcgggg tttgggtgtg tcggatgtgg atgtggtgga ggcgcatggg acggggactc 26040
 ggttgggtga tccgattgag gctgaggcgt tgttggggtc gtatgggcgg ggtcgtgtgg 26100
 gtggggcgtt gttgttgggt tcgggtgaagt cgaatattgg tcatacgag gcggctgcgg 26160
 gtgtggcggg tgtgatcaag atggtgatgg cgttcggggc gggggtggtg ccggcgacgt 26220
 tgcattgtga tgtgccgtcg ccgttggtgg attggtcttc ggggtggggtg gagttggtga 26280
 cggaggcgcg ggattggccg gtggtgggtc gtgtgcgtcg tcgggtgtg tcggcgtttg 26340
 ggggtgcggg gaogaatgcg catctgattt tggagcaggc ccccgagttc gacgatcctg 26400
 ccgattccga ttccgattcc gattccgatg ccggtgtcgt ggatggcggc gaggggtggtg 26460
 ttggcaggag cttgtcggtg gttccggtgg tgggtgcggg tcgttcggtg ggggctttgc 26520
 gggcgtatgc gggtcgggtt cgtgaggtgt gcgcgggggt gtctgacggg ggtggctccg 26580
 gtggtgggtt tgggttgggt gatgtgggtt ggtcgttggg gtcgtcgcg tcgggtgttg 26640
 agcatcgggc ggtcgtgttc ggtgggggtg tggaggagggt tgttgctggt cttggtgcgg 26700

tggcttcttg	ggcgggtggc	tcgggttcgg	tgggtggggg	ttcgggtggc	tcgggtgttg	26760
ctgggtgggg	tggtcgggtg	gtgtttgtgt	ttccgggtca	gggttggcag	tgggtgggta	26820
tgggtgcggc	gctgctggac	gagtcggagg	tgttcgccga	gtcgatgggt	gagtggtgct	26880
gggcgttgct	ggggtttgtg	gattgggatt	tgttgagggt	ggtgcgcggc	ggggcgggtg	26940
aggggggtg	gggtcgggtt	gatgtgggtc	agccgggtgc	gtgggcgggt	atggtgtcgt	27000
tggcgcgggt	gtggatgtcg	gtgggtgtgg	tgccggatgc	ggtgggtggg	cattcgcagg	27060
gtgaggttgc	tgcggcgggt	gtgggggggt	tgttgagtgt	ggctgatggg	gcgcgggtgg	27120
tggcgttgcg	gtcgcgggtg	atcggtgagg	tgttgccggg	tgggtggtcg	atggtgtcgg	27180
tcggactgcc	gatcgtggat	gtgcaggaac	ggttggcggg	gtgggggtgg	cggttgggtg	27240
tggcggcggg	gaatggtccg	tcgttgacgg	tgggtgtcgg	ggatgtggat	gctgctgtgg	27300
ggtttgttgg	tgagtgtgag	cgggatgggg	tgtgggtgcg	gcgggtggcg	gtggattatg	27360
cgtcgcattc	ggcgcattgt	gaggcgggtg	aggggatgct	gtcgggggtt	ttgggtgggt	27420
tgtgtccggg	gcgggggtgt	gtgccgtttt	attcgtcggg	ggtgggtggg	gtggttgatg	27480
gggtgggttt	ggatggtggg	tattgggtat	ggaatctcgg	tgagcgggtg	ttgttttcgg	27540
atgtggtggg	gcggcttggt	ggggatgggt	tttcgggggt	tgtggagtgt	tcggggcatc	27600
cggtgttggc	gggtgggggt	ttggagtccg	tggcgggtgg	ggatccggat	gtgcggccgg	27660
tgggtgggtg	gtcgtgcgc	cgtgatgatg	gtgggtgggg	ccggtttctg	acgtcgggtg	27720
gtgaggcgtt	cgtcggcggg	atgagtgttg	actggaagg	tgtgttcggg	ggggcggggc	27780
cgcgggttgg	tgacctgccg	acgtatccgt	tccaacgccg	ccactactgg	gcaccgactc	27840
ccaccaaccc	cgcaccaaac	cccgccacca	accccgccac	caaccccgcc	acgggcgaca	27900
ccaccaccgc	cgaccgcggc	ggtgacctgc	ggtatcggat	cacctggaaa	ccgttgccga	27960
ccgacgaccc	ccgacccttc	accaaccgct	ggctgctgat	ggtgcccag	gcgctggccg	28020
gtgacggggg	ggtggcgggc	gtacggcagg	cgctggccgc	gcgtggcgcc	tccgtcgaac	28080
tgctgaccgt	cggcaccgcc	gaccggggcg	gccttgccgc	gctcctgacc	tccgccgccc	28140
ccggcgaccc	ggaggcggcc	ggcccgggcg	gcgtggtctc	cctgctggcg	ctcgccgagg	28200
gcgcggacgc	gcgccacccg	gccgtaccgc	tcggcctgac	cgcctcgctc	gccctgatcc	28260
aggcattggc	ggacgcgggg	acgcaggccc	gcctctgggc	ggtcaccocg	ggggccgtcg	28320
ccgtgtcctc	cggcgagggt	ccggacgcgc	ggcaggccca	ggtgtggggg	ctcgcccggg	28380
tcgcggccct	cgaactgccg	gaccgatggg	gcgggctggg	ggacctgcgc	gcgctcaccg	28440

gggagcgtgc cttcgcgcag ctcccgatg tcgtggcgcg ctcgaacggc gaggaccagg 28500
 tcgccgtacg ggccctccggc gtctacggtc gacgcctcgt gcgttcccgc gccaccgtca 28560
 cgtccggcga ctggccggcc cggggcacca tcctcgtcgt cggggacacc ggcccggtcg 28620
 ccgcgctcct ggccggccgc ctccctggcg acggggcgcg gcacgtggtg ctcccgggcc 28680
 cggccgcccgc gtccaccgtc gggctcaccg gcggggccga ccgggtggcc ctgatcgact 28740
 gcgacccgag cgaccgggac gcgctcgccg ggctgctcgg cgcgtaccgg cccacgacga 28800
 tcgtgggtggc tccgcccgcg gtccgctca ccgccctcgc cgagaccacg ccggaggact 28860
 tcgtcgccgc cgtcgccgcg aagacgacga cggcagtga cctcgacgcc cttgcccggg 28920
 aggcggaact ggagctcgac gcgttcgtcg tcttctcctc ggtctccggc acctggggcg 28980
 gcgcggggca cggcggtac gcggcgggca ccgccggct ggacgcgctg gtcgaggaga 29040
 ggcggggccc tggcctgccc gccacggcga tcgctggac gccgtgggccc gacgcgacca 29100
 cagccgcccg cgggcaggca cccgatgcc gcgccggcg gcacgaaccc gacacgaggg 29160
 ccggggggccc cgaccgcgaa ctgctgcgcg ggggtggcct caccocgttg gaccgggggg 29220
 ccgcgctgga cgtgctgcgc gggcggtgg cgccggcgga gggcctggtg accgtggccc 29280
 acgtcgactg ggcgcggttc gtccctcgt acaccgcggc ccggcccacc acgtctctcg 29340
 acgaactgcc cgagctgcgg gcgaccggg aggcggagca caccocggcc gaggactcgt 29400
 cggccggcg cgaactggtc cgtgccctca gcggccggcc cgcggccgat cagcaccgga 29460
 cgctgctgcg gctgggtccgt gcgcacgtcg cggcctcct ggggcaacgac gaggccgagg 29520
 cggccgatcc ggaccggcg ttccgggaac tcggcttcac ctccgtgacg gcggtggacc 29580
 tgcggaaccg gctgaacgcg gccaccgggc tgaacctgcc ggcgtccgtc gtcttcgacc 29640
 atcccagcgc ccgggtgctg gccgcgtacc tgcgtgccga gctgctcggg ccggaggccc 29700
 acgaggacac ggcggaggcc gtcccccgc cgtccgcgcc ggccggggcg ggcgacgacg 29760
 agccgatcgc ggtgatcggg atggcctgtc ggttcccggg cggggtcgac gccccgacg 29820
 acctgtggga tctgctggcg aaggcccgcg acgccatctc caggttcccc acgaaccggg 29880
 gctgggacgt cgacggcctg tacgaccgg acccgaggc gcccgccgc acctacgtcc 29940
 gcgagggcg cttcctgcac gacgcgccg acttcgatgc cgcgttcttc gggatctcgc 30000
 cccgcgaggc cctcgccatg gatccgcagc agcgcctgct gctggagacc acgtgggagt 30060
 ccctggaacg ggccgggttg gaccgaccg cgttgccgcg caccocggacc ggggtgttcg 30120

tggggaccaa cggccagcac tacatgccgc tgctgcgaga cggcgcgagc gacttcgacg 30180
gctacctcgg caccggcaac tcggccagcg tcatgtccgg ccggctctcc tacgtcttcg 30240
gcctggaggg cccggcggtg accgtggaca cggcctgctc cgcctccctc gtggcgctgc 30300
acctcgcggt gcaggcgctg cgccggggcg agtgcacgct ggccctggtc ggcggggcca 30360
cggatgatgtc gacgccggac atgctggtgg agttctcccg gcagcggcg atgtcgccgg 30420
acggccggtc gaaggcggtc gccgccgcgc ccgacggggg ggcgctcagc gagggcgccg 30480
ccatgatggt ggtgcagcgg ctgcgcgacg cggaggccgc cgggcacgag atcctggccg 30540
tggatcaagg ctcgccgctc aaccaggacg gggccagcaa cggcctcacc gccccgaacg 30600
ggccctccca ggaacgggtc atccggcagg cgctggccga cgccggcctg cggccggacc 30660
aggtggacgc ggtcgaggcg cacggcaccc gcaccgccct gggcgacccc atcgaggcg 30720
aggcgctgct cgccacgtac ggccgggacc ggccggcggg ccggccactg tggctcggct 30780
cgctgaagtc caacatcggt cacacccagg ccgccgccgg catcgccggg gtgatgaagg 30840
tgatcctggc gctgcggcac gacacgctgc cgcgcacgct gcacgtggac cggccgacgc 30900
cccgggtgga ctgggcttcc ggggcgggtg cgttgctgac cgagccggtg ccgtggccgc 30960
agggcgacga accccgccgg gcggcggtgt cctcgcttcg gatcagcggc accaacgccc 31020
acgtgatcgt cgagcaggcg ccgccgggtg tcggggaacc gatcgaccac gaggcggacg 31080
aggtcacctg cccgctgttc ctgtcggccc gggggagcgc cgcgctctgc gccaggcg 31140
cacggctgcg ggcccggttg atcgaggaac ccgacctgga catcgccgag gtcggctaca 31200
cgctggcggc caccggggcc cgcttcgagc accgggcccgt ggtgatcggg gagagccgcg 31260
cggaggtcgg cgacgcgctc gccgcgctgg cccggggcga ggagcaccgc tcgctgctgc 31320
gggggcgggc cggcgcgagc gaccgggtcg cgttcgtctt tcccgccag ggctcgcagt 31380
gggcccagat ggccgacggc ctgctcgacc gctccccggc cttccggcg agcgctcg 31440
cgtgcgacga ggcgctgcgg gcgcacctcg actggtccgt gctggacgtg ctgcgtcgcg 31500
tgccggacgc gcctgcgctg agccgggtcg acgtggtcca gccggtgctg ttcacgatga 31560
tgggtgctcg ggcggcggcc tggcgggcgc tgggcgtgca cccgtccgcc gtggctggcc 31620
actcgaggg tgagatcgcg gcggccacg tggcgggcg cctctcgctg gacgacgcgg 31680
cgcgcacgt cgccctgcgc agccaggcgt ggctgcggct ggccgggcag ggcgggatgg 31740
tggcggtgct gctccccgtc gacgcgctcc gcgcccgcct ggcgcggttc ggcgaccggc 31800
tgtccgtcgc cgcggtcaac agccccggtc cggcgggcgg gagcggtac cccgacgcgc 31860

tgcgcgaact cgtcgacgag ctgaccgccc agggcgtgca cgccaaggcg atcccggggg 31920
tggacacggc cgggcactcc gcgcaggtgg aggtgctgaa ggaccacctg atggccgccc 31980
tcgccccggg gtcgccccgc agctcgacaga tccccctcta ctcgaccgtc acgggcggcc 32040
tgctggacac cgcgctgctg gacgccgcct actggtaccg caacatgcgc gaccgggtgg 32100
agttcgagca ggcgacccgg gcgatgctcg cggacgggca cgaggggttc ctggagccca 32160
gcccgcaccc gatgctgtcg gtgtcgttgc agggcaccgc ggccgatgcc ggggtcgccg 32220
cgacgggtgt ggggacactg cggcgcgcca agggcggcgc ccgctgggtc ggcatggcgc 32280
tcgggctcgc ccacgcccac gggatcgaga tcgacgcgag tgtgctcttc ggaaccgact 32340
cgcgcggggg cgacctgccg acgtaccgt tccagcgcca gcgcttctgg tatcaccgc 32400
cggccgcgcg cggggacgtg gcctccgccc ggctcagcgg tgccgacct ccgtgctgg 32460
gcggggcggt cgagctgcct gaccggggcg gccacgtgta tccggcccgg ctcggcgtcc 32520
gacaccaccc gtggctcggc gagcatgccc tgctgggcgc ggcgatcctg cccggggccg 32580
cgtacgcgga actcgccctg tgggcccggc ggcgtgacgg ggccggccgg atcgaggagc 32640
tgaccctoga cgcgccgctg gtggtggccg acgagtcggc ggcgcaactg cggctcgtgg 32700
tgggcccggc ggacgcggag gggcgccggc agctcacgt ccactcgcg gccgacggcg 32760
cggacgcgga caccgctgg acccggcac gcgagggcac cctcgtgccg gccgacgccg 32820
acgccgcgg gagcggggac ccgggcgcgc cctggccgcc ggccggggcc gagcccgctg 32880
aggtggcggg cctgtacgac cggttcgccg accggggcta ccagtacggg ccgtcggtcc 32940
gggggggtcc ggccgcctgg cgggccggcg acacgggtgta cgccgaggtg gccctgccc 33000
tccgcagcc cgggagccc cgcttcggtg tccaccggc gctgctcgac gcggcggtcc 33060
aggcgatgag cctcggcgcg ttcttccccg aggacgggca ggtccggatg ccgttcgccc 33120
tgccgggctg gtcgtcgtcc ggggtcgggg ccgaccggct gcgggtcacc atcagcccgg 33180
ccggtgccga ggcgggtccg atgcctgctg tcgacgagcg gggcaaccgg gtcgtggtga 33240
tcgactccct ggtggcgcg gcggtgccgg tggaggcgct ccccccgcc acccccgga 33300
ccggggacgg cgcgctgcac cacgtgcct ggaccgcccg gccggaaccg ggggtcgccg 33360
ccgtgcagcg ctgggcggtc gtgggcgcgg ccgatcccgg gctggccggg ggcctggacc 33420
gggcgggcgg cctctcggg gcgtacccc atctcgccgg tctggtcgcg gcggtggccg 33480
aaggggcggc gctgcccga gtggtcgcg tgccgggtcc gtcgggcgcg ccggtcgggc 33540

ccgacgcggt gcgcgccacc gtgctcggcg ccctggacct gatccgggcc tggctcgcgg 33600
 tcgagggccg gctggggctg gccaggetgg cgttcgtcac cacctcggcg gtggcggtcg 33660
 gcgacggcac cgagcacgtg gacccggtgt cggccgccct gtgggggctg gtgcgttccg 33720
 cccagtccga ggagcccggc cggttcgtcc tcgtcgacct ggacgccgac ccggccagcg 33780
 cctcggccct gcccgcccg ctcgcgcgcg gtgagccgca actggccgtt cgcgcggggg 33840
 cgggtgcacgt gcccgggctg gttcggcacc gaccccgccc ggacggcccg ctgacgcccc 33900
 cggccggtgc cgcgtggcg ctcgcgcgcg gtgggcaggg caccctggag ggcctggcgc 33960
 tgggtcccgc ccgggacgcc ttggcgccgc tggcccccgg gcagggtccg gtgcgcgtgc 34020
 gcgcgcgcgg agtgaacttc cgggacaccc tcatcgcgct cggcatgtac ccgggcacgc 34080
 cggtgctggg tgccgagggg gccggggtga tcaccgaggt cgcgcgggac gtggccggct 34140
 tcgcccccg cgaccgggtg ctgggcatgt ggaccggcg cctggggccg gtggcggtcg 34200
 ccgacgccc gatgctcgc cgggttcgc gcggctggtc gtacgccgag gccgcgtcgg 34260
 tgccggccgt ctctctcac gcccaactac cgctcaccag gtcgcgcggg atccgcccgg 34320
 ggcagtcgt gctggtgcac gcgggggccc gcggcgctcg catggcgacc ctccaactgg 34380
 cccggcacct gggcgtggag gtctacgcca cggcgagccg gggcaagtgg gacaccctgc 34440
 gtggcctcgg cctggacgac gcgcacatcg ccgactcccg cagcctcgac ttgcgcggac 34500
 ggttcctggc cgccaccggg gggcgcgcg tcgacgtggt gctgaactcc cttgcggggg 34560
 acttcgtgga cgcgtccctg cggctgctgc cgcgcggcg ccacttcctg gaactgggca 34620
 aggcgcgacgt ccgcgacccc gaccggatcg cggccgacca cccgggggtc ggctaccggg 34680
 cgttcgacct cgtcgaggct ggtccggagc tggtcgggca gctgctcggc gagctgatgg 34740
 agctgttcgc cgccgggggtg ctcagcccgc tgccgttgac cgtgcgggac gtccggcggg 34800
 cccgggaggc gttccgcctg atcagccagg cccggcacgt cggcaagggt gtgctgacca 34860
 tgccgcccgc gttcggcgcg tacggcaccg tcttggtcac cggcggcacc gggacgctcg 34920
 gcggcgccgt cggccggcac ctggtcgccc ggcacggcgt acggcacctg gtgctcaccg 34980
 gccgcagcgg cccggcgggc gacggggcgt ccgcgctcgt cgacgagctg accgcgtccg 35040
 gcgcgtcggg gaccgtcgtc gcctgcgacg ccgcgcgacc ggtcgcgctg cgcgggtgc 35100
 tcgacggcat tccggcccg caccgctca ccgcgctcgt gcacgctgcc ggcgtcctcg 35160
 acgacgccac catcaccgcg ctgaccgcg ggcagggtga cgcggtgctg cggcccaagg 35220
 ccgacgcggt gatcaacctg cacgagttga cccgggaccg ggagctgtcc gcgttcgtgc 35280

. tgttctcctc ggcggcggcc ctgttcggca gcccggggca gggcaactac tcggcgggcca 35340
 acgggttcgt cgacgcgttc gcccagtacc gccgcgcgca ggggctccac gcgggtgtcgc 35400
 tggcctgggg cctgtgggcc gacagcagcc ggatggccgg gcacctcgac caggagggga 35460
 tgcggcgccg gatggcgcg cggggcgctc tgccgctcac caccgaccag ggcctcgccc 35520
 tgttcgacgc cgcgcagctg gtggacgagg cgctccagggt gccgatccgg ctcaacgtcg 35580
 gcgcgttcgc ggccgcgggg agggctcccc cgctcctcgc cgacctggtg ccggcgggcg 35640
 cgtcgggggc cccggccgcc accccgaccc gggacgacgc ggaccgcacg ctcgccgacc 35700
 ggctcgccgg gctgaccgtg gccgaacagc gggagctggt gctggagagc gtgcgcggac 35760
 acgcggcggc cgtcctcgga cacgccgacc cgcaggccgt cgacgccgac cgggccttcc 35820
 ggggaactcg cttcgactcg ctgacggcgg tggagctgcg caatcggtg gccaccgcgt 35880
 ccgggctgcg cctgccggcg acgctggtct tcgaccacc caccgccgaa gcgttggcgg 35940
 agcacctgct cgccgggctc gcgcccagac agggccgggc cgagttgccg ttgctggccg 36000
 agctggggcg gctggaggcg gccctggccg ccaccgacgg ggccgccctc gacgggctgg 36060
 acgacctggt gcgcggggag gtcggcgctc ggatcgcggc gctggccgcc aggtggggcg 36120
 cggccggcga cgacgtggcc ggcagcgacg gcggcgggac ggccgacgcg ctcgagtccg 36180
 ctgacgacga cgagatcttc gcgttcacgc acgagcgggt ccgcgcctga cgaccccgcg 36240
 , tacgcgaggg acggggtgga cgggaccgac ggtcaggagg gacgaggcgg catgtcgaac 36300
 gagcagaagc tccgcgagta cctgcggttg accaccaccg agctggccag ggccaccgac 36360
 cggctgcgcg cggctcaggc gcgggcgcac gagccgatcg cgatcgtcgg catggcctgc 36420
 cggtagcccg gcggggctcg ctcaccggag gaactgtggg agctggtcgc ctcgggcacg 36480
 gacgcgatct cccggttccc cgacgaccac ggctgggacg gcgacgcgct gtacgacccg 36540
 gacccgagg cggcgggccg cacctactgc cgcgaggcg ggttcctcgc cggggtcggc 36600
 gacttcgacg ccgcgttctt cggcatctcg ccccgcgagg cgctggccat ggaccgcgag 36660
 cagcgctgc tgctggagac gtctgggag gcgctggagc gggccgggat ccccccggac 36720
 tcgctgcgcg gcagccgtac cggggtgtgc gtcggggcgt ggcacggcgg ctacaccgac 36780
 gtcgtcgggc agcccccggc ggaactggag ggccacctgc tgaccggcgg ggtggtcagc 36840
 ttacactcgg ggcggtatct gtacgcgctg ggcctggagg ggcccgctt gacggtggac 36900
 accgcctgct cgtcctcgct ggtggccctg cacctggcgg tcggggccct gcggcagggc 36960

gagtgcgacc tggcgttggc cggcggggcg acggtgctgg ccagcccggc ggtgttcgtg 37020
cagttctcgc ggcagcgggg gctggccccc gacggccggg gcaaggcgtt cgccgactcg 37080
gcggacgggt tcgggccggc cgaggggggtc ggcatgctgg tcgtggagcg gctgtcggac 37140
gccgtccgcc acgggcgccc ggtgctggcc ctggtcaccg gcacggcggg caaccaggac 37200
ggggcgagca acggcctcac cgcggccagc ggcccggcgc aggagaagggt gctgcgccag 37260
gcgctcgtgg acgcccgggt gacggccgcc gacgtcgacg cggtcgaggc gcacggcacc 37320
ggcaccggc tcggcgaccc gatcgagggt cggggccctga tgaacgtgta cggtgccggc 37380
cggcccggcg accgtccgct ctggctcggg tcgctgaagt ccaacatcgg ccacacccag 37440
gcggcggccg gggtcggcgg ggtcatcaag acggtgctgg cgatgcggca cggcgtcctg 37500
ccggccaccc tgcacgtgga cgcggcgacc accgaggtcg actggtcgc cggccagggtg 37560
gccctgctgc gggcagagac accgtggccg gacacgggtc gcccgcccg cgccgggggtc 37620
tcctccttcg gggtgagcgg caccaacgcg cacgtggtgc tggagcaggc ccctgggccc 37680
gccgcccggc cggcgggtga cgcggcgccc gccgagaccc ggcccgtcgg cgaccgccc 37740
ccggtcgtac cgctggtgtt gtccgcccagg tcgcagccgg cgtggccgg gcaggcccg 37800
cggctgcgcg acctgctggc cgcagcggc gagaccgacc tcgccagcg cggaactcgc 37860
ctggccaccg cgcggtcggg gtccgaccac cgggcggtgg tgacggccgc cgggcgaccg 37920
caggcgctcg acgcgtcga cctgctggcc ggcggcgaac ccggaccggc ggtcacgacc 37980
ggcgtcgccg ccccgaccgg gcgcaccgtg ttcgtcttcc cggggcaggg gacgcactgg 38040
gccggcatgg gtgcccacct gctcgaccag tcaccggtgt tcgccagtc gatgcgacgg 38100
tgcgagcagg cgctgtcggc gcacaccgac tggaaagctc gcgaggtgat ccggggcgcg 38160
gccggcagcc cgcgctgga ccgctggac gtgctccagc ccgtctcctg ggcgggtgatg 38220
gtgtcgctgg cgcaggtgtg gcggctcgtc ggcgtcgagc cggacgcggg ggtcggccat 38280
tcccagggcg agatcgccgc cgcggtggtc tgcggcgccg tgaccctgcc ggacgcggcc 38340
cgggtggtcg cgctgcggtc ccaggctatc ggtcgggtgc tctccggtcg cggcggcatg 38400
gcgtccgtcc agctgccggc ccgggaggtc gcggggcgcc tggccgcctg ggcgggcccg 38460
ctcgacgtcg cggccgtcaa cgggccacag tcgaccgtcg tgtccggtgc cgccgacgcg 38520
gtcaccgaac tggtcgaggc gtccgcggcc gaggacgtcc gggtcggcg gatcccgggtg 38580
gactacgcgt ccactcgac gcaggtggac cggctgcgcg ccgagctgct caccgtcctg 38640
ggcccggtcg acgcccgtcc ggcgcaggtg cccttctact cgacggtgca ggcggggcgc 38700

gtcgacactg ccggcctgga cgcgggctac tggtagcgca acctgcgggg gcaggtccgc 38760
 ttcgaggaga ccgtgcgggt gctgctcgac gacgggcacc gcgccttcgt cgaggccgcc 38820
 gcgcacgccg tcctcgtacc cgcgatccag gagctggggg acagcgccgg cgtccgggtg 38880
 gtggccgtgg ggtcgtcgcg ccgggaggcg ggcggcctgg accggctcct ggcctcggcg 38940
 gccgaggcgt tcacccaggg ggtggccgtg gactggtccc gggctctggc cggggccgcg 39000
 cgcgtcgccg tggacctgcc cacgtacgcg ttccagcggc aacgctactg gctggagccc 39060
 gccgcgcagg cggactccgg cccggccggg gacggctggc gctaccgggt cggctggcgg 39120
 cggcttcagc gcaccggcgc cgcgccggcc gaccggtggc tgctggtgac cggcccggag 39180
 cagccggcgg agctggtcga ggcggtgcgc gacgcgctca ccgcgcgggg cgcgaggtg 39240
 cgcctggtga ccgtcgagcc gaccagcacc gaccgggccg cgtgcgcggc gttgctcacc 39300
 gcggccgggtg cgggcggggc gaccgggtg ctgtcgtgc tcggcaccga tcgtcgcccg 39360
 caccgccacc acccggccgt gtccgtcggc gccgccgca cgttgctgct gaccaggcc 39420
 gtcgccgacg ccctgccggc cgcgggctg tgggtcgtca ccgggggcgc ggtctccgtc 39480
 gggcccgccg agaccgccga cgagcgccag gcgcaggtct gggggttcgg cgggtcgcg 39540
 gccctcgaa tgccccgcac gtggggcggg ctctcgacc tgcccgccga cgcggacggc 39600
 ccggtgtggg aggcgttcgt ggacgtgctg gccggggacg aggaccaggt cgcgctgcgc 39660
 ggcccggtcg ggtacggtcg ccggctccgg cgcgccccg cgtaccgc gaagcggcgg 39720
 taccggccca ggggcaccgt cctggtcacc ggcggcaccg gcgcgctcgg cgcgcacgtg 39780
 gcccgccggt tggccgccgg cggggccgcg cacctcgtgc tcaccagccg gcgcggggcc 39840
 gacgcccccg gtgcggccgg gctggtcggg gaactccggg cgtggggcgc cgaggtgacc 39900
 gtcgcggctc gcgacgtcgc cgaccgggcc gccgtggcgg cgtgctcgc cgggctgccc 39960
 gccgacgcgc cgtgagcgc ggtcttcac acccgggcg tggcgactc gatgccgatc 40020
 ggcgagaccg ggctcaccga cgtcgccgag gtgttcgccg ggaaggcgc cggagccgc 40080
 cacctcgacg aactcaccg ggggcacgac ctggacgcgt tcgtcctgta ctgtcgaac 40140
 gcgggcgtgt ggggcagcag cgggcagagc gcgtacgggg cggccaacgc ggccctcgac 40200
 gcgctcgccg aacggcggcg cgcgccggg ctgaccgcca cctccgtcgc ctggggcctg 40260
 tggggctccg ggggcattgg cgaggcgac gccgaggagt acctgagccg cggggcctg 40320
 cggccgatgc ctccgagcg tggcgtggac gccctcctgg ccgcctgga cggggacgag 40380

accttcgtcg ccgtcgccga cgtggactgg acgctgttca cggccggggtt caccgcgttc 40440
 cggcccagcc cgctgctcgg cgacctcccg gagggccgcg cgacgctggc cgacgccgga 40500
 cccgcggggt ccgacctgcc ggcttggcac gccgccgcga gcccgcacga acgccgccgg 40560
 ggcttgcctc acctgggtacg ccggcaggtc gccgccgtcc tcggccaccc gggggccgag 40620
 cacgtcggcc ccgacgccgc gttccgggag atcggattcg actcgtgac cgccgtcgac 40680
 ctggccaagc ggctcagggc ggcggtcggc gtgccgctgt ccgccaccct cgtcttcgac 40740
 caccaccg cgacggcggc cgccgagcac ctggccgggc tgctcggtc ccgccggcc 40800
 ggccgagacc cgcgcgaggc cgaggtgcgc cgggccctgg ccgacctgcc gctggcccg 40860
 ctgcgggacg ccggcctact ggacggcctg cttgcgcttg cggggtgga ccgcgacgcg 40920
 gtgccggacg ggcccagacc ggctcccggc gacgccatcg acgaactcga tccagaggag 40980
 ctggtgcgcc ggggtgctgga caacgccagc tcctgaccgc ttccctcttc ccccgagga 41040
 gccgcgccat ggtcatgccc cccgacaagg tgatcgaggc gctgcgtgtc tccgtcaagg 41100
 agacggagcg gctgcgccgg cagaaccacg agctgctcgc cgccctgcac gggccgatcg 41160
 ccgtcgtggg catggcctgc cgctaccgcg gcgggggtgt ctctccggag gacctgtggc 41220
 ggctggctga gacgggcacg gacgcgatcg gcggcttccc caccgaccgt ggctgggacg 41280
 tcgacgccgt gtacgaccgc gatcctgagt cgcggaacac cacctactgc cgggagggcg 41340
 ggttcctggc cggggcagga gacttcgacg ccgcgttctt cgggggtgtc ccgcacgagg 41400
 ccgtggcat ggacccccag cagcggctgc ttctggaggc gtccctgggag gcgctggagc 41460
 ggtccgggac cgacccgcac agcctgcgcg gctcgcgcac cgggggtctac gtcgggtcgg 41520
 cccaccaggg gtacgcggtc gacgccggtc aggtgccgga gggcgaggag gggttccggc 41580
 tgaccggcag ccgcgacgcc gtccgtgccc gacggatctc gtacctgctc gggctggagg 41640
 gtccggccct gaccgtcgag acggcctgct cgtcctcgtc ggtggcggcg cacctcgcgg 41700
 tgcaggcgt gcgccggggc gaggcgggc tggcactggc cggcggggtc gccgtgatgc 41760
 ccgaccggc ggcatcgtg gaggttctcc gccagcgggg cctcgcggcg gacgggcgct 41820
 gccgggcgtt cggggcgggc gcggacggca ccggctgggc ggaggcgctc ggtgtgctgg 41880
 tcctgcaacg gctctccgac gcggtgccgc acggccgctg ggtgctgggc gtgatccggg 41940
 gttcggccgt caaccaggac ggggccagca acgggctgac cggcccgagc gggcccgcc 42000
 agcagcgggt catccggcag gcgctgaccg acggccggct cggcgccgac cagatcgacg 42060
 cggctcaggc gcacggcacg ggcaaccggc tcggcgacc gatcgaggcg caggcgctga 42120

tcgccgccta cggcgccgac cggaccccg accggccgct ctggctcggc tcgttgaagt 42180
 cgaacatcgg gcacgcccag gcggcgccg gcgtcggcgg cctgatcaag atgctcctgg 42240
 cgatgcgggc cgggacgctc ccacccaccc tgcacgccga cgtcccgacc ccgctggctg 42300
 actggtccgc cgggtgctgc cggctgtcga cgggggtggt gccctggccc gcgttgcccg 42360
 gggcgccccg cagggccggg atctccgct tcggggtgag cggcaccaac gcgcacgtga 42420
 tcgtcgagca gccgccgccg gtcccggctg acgaccggc gccaccacg aggaccctgc 42480
 cgctggtgcc gtgggtgctc tccggccgga cggaggcggc gctgcgcgcc caggcggacc 42540
 ggttgcttac gcacctggcg gcgcacccc acgcggacc cctggacgtg ggattctccc 42600
 tggccaccag ccgggcccg ctggagcacc gggccgtgct ggtggccgcc gaccgcgacg 42660
 gcctgctccg cctcgctgac gcgctggccg ccggcgagcc ggccggcgcc ctgatccggg 42720
 gcacggtacg tcacgatcgc cggaccgggt tctctctcgc cgggcagggc ggccagcgcg 42780
 tcgggatggc gcgcgaactg tacgaggcgt tccccgcctt cggcgacgcc ctggaccagc 42840
 tcgccgcccg gctggaccgg cacctcgatc gtccgctgct gcgggtgctg ttccgccgagc 42900
 cggggtcggc gcacgcccgg ctgctcgacg gcaccggta cgcgcaggcc gccctcttcg 42960
 ccgtcgaggt ggcgttgctc cgaactggtc acggctgggg ggtccggccc gacgtgctgc 43020
 tcggccactc ggtgggcgag ctggcggccg cgcacgtggc cggcgtactc gacgtggacg 43080
 acgcgtgca gctggtcgcg gcgcggggcc ggctgatggg ggagctgccg tcgggcggcg 43140
 cgatggtggc ggtccgggcc accgaggagg aggtcgggcc cctgctcgac gggcagcggg 43200
 tcgcggtggc ggcggtcaac ggcccgcgt cggctcgtgt ctccggcgac gaggaggcgg 43260
 tgctggccgt ggccgcccg tgccgcgcc tcggccaccg gacgcgacgc ctcaacgtca 43320
 gccacgcgtt ccaactcccc cactggagg cgatgctgga gccgttccgg cgggtggcgc 43380
 ggggcctgac gtaccatgcc ccgacgatcc cgggtggtgc gaacgcgacg ggccggctcg 43440
 ccaccgccga cgcgtgcgc gaccccggtt actgggtccg gcacgtccgc cagcccgctc 43500
 ggttccggga cggggtgcgg gcgcgcccg accagggggc caccgccttc gtcgggctcg 43560
 gcccgacgg ggtgctgtgc gcgttgccg aggagtgcct cgggcccacc ggcgacgtgc 43620
 tgctgctgcc ggtgctgcgc cccggtcggc cggagcccgc caccctgctg gccgccctgg 43680
 ccggggcgta cgcggcggc gcggaaatgg actggtcccg ggtgttcgcg ggcaccggcg 43740
 cgcgcagggt cgagctgcc acgtacgct tccagcacg gcgctactgg ctggcgccgg 43800

gccccgccgtc ggccccgccgc gacgacgcct ggcgggtaccg gatcgccctgg cggccccctgc 43860
cgaccgtgcc cgccgccgcc gggaccgaga cgggtggccgg ggcgtggttg ctggtggtcc 43920
ccgcccacga cggcgtcgcg tcgctcgccg acgccgccga gcgggcccgtg caccggggcg 43980
gggccacggt caccgggctg acggtggacg ccgccgacgt ggaccgggac accctcgccg 44040
ccgtgctgac cgaggccgcc gccgacgcgg acggcgggcc ggacggggtg ctctgcctgc 44100
tgggcctcga cgaccgggca catccccggt ccgcctcggg gccccgcggg gtgctggcga 44160
ccctgtccct cgcccaggcc ctgaccgacc tgggggcctc cgcgcggtg tggtgcgta 44220
ccgggggggc ggtcgccgtg acgcccggcg agtccccgtc ggtcgccgga gccagttgt 44280
ggggcttcgg ccgctggcc gcgctgaac tccccggtc ctggggcggc ctggtggacc 44340
tgccggtcga ccgggacgac cgggactggg acctgctgcg gcgcgcgctg cgcggccccg 44400
aggaccaggt cgcggtccgg ggggcggctg ggtacgcccg gcggctggtc ccgcgccccg 44460
cgccccgggc cgagcgggcc tggcgtccgc gcggcacggt cctggtgacc ggcggtacgg 44520
gcgcgctcgg cgcgcacacg gccgctggc tggcgcgcaa cggcgccacg caccctcgcc 44580
tcaccagccg ccggggcggg aacgccccg gggtcgcccgc gctgcgggcg gaactggtca 44640
cgctcggtc cgaggtgacc gtggtcgct ggcacgtcgc cgaccgggag gccgtggccg 44700
gcctgctcgc cgggattccc cgcgcgcctc cgctcaccgc cgtgttccac gcggcgggcg 44760
tgccccaggt gacgccgctt caccgagcga ccccgagtt gttcgcgag gtctgcgcag 44820
gcaaggctgc cggggcggtg caccgacg agttggccgg tgacctggac gccttcgtca 44880
ccttcgcctc cgccgccggg gtgtggggca gcggcgggca gtgcgcgtac gctgcggcca 44940
acgccgccct cgacgcgctc gccgagcgtc gtgcgcgcgc agggctgccc gcgacctccg 45000
tcgcctgggg ggtctggggc gggcccggca tggggggcg cgcgggggag gagtacctgc 45060
gccgcgggg cgtccgggcg atgcccccg cagccgccct cgccgccctc gggcggatcc 45120
tggaacccga cgagaccggg gtgacggtct ccgacaccga gtggggccgg ttcgcgtccg 45180
gcttcgcgc cgcgcgtccc gccccgctgc tcgcgagct gccggggcg gacgtcgatc 45240
cggccggccc ggcgcaccgg gcgcagccgc ccgtgccccg accggccccg gcagccaccg 45300
accgccccg gctgctggcg ctggtccgcg ccgagggcgc cggggtgctg gggcacgacg 45360
gtgcgacga cgttcgggcc gacgcggagt ttccgcctc cggcttcgac tcgctcgccg 45420
ccgtccagct gcgcgcggc ctgcgcgagg ccaccggcct gagcctctcg gccccggttc 45480
tgttcgacca ccgcaccct gacgcgctc cgcgccacct gcacggcctg ctcaccggcg 45540

cgggcgggcgg gccacccgcg ccggccgcgg ggagcgccct ggtcgagatg taccggcggg 45600
ccgtcgccac cggccgcgcc gccgagggcg tggaggtgct cggcaccgtc gccacgttcc 45660
ggccgggtgtt ccggtccccg gacgaactgg gcgagccacc ggccctcgtc ccgctcggca 45720
ccggggcgggg gggaccgcg ctggtctgct gcgcgggcac ggccgcggcg tccggccccc 45780
gcgagttcac ggcgttcgcc gccgcgctgg ccggtctccg ggacgtcacc gtccttccgc 45840
agaccggctt cctgcccggc gagccgctgc ccgcccggct ggacgtgctg ctgcacgccc 45900
aggccgacgc cgtcctggcc cactgcgcgg ggggaccctt cgtcctggtc ggccactcgg 45960
ccggggcgaa catggcgcac gcgtgacgg tccgcctgga ggcgcggggc gcggaccccg 46020
ccgcgctggt gctgatggac atctacacgc ccgccgccc gggggcgatg ggggtgtggc 46080
gcgaggagat gctggcctgg gtcgccgagc ggtccgtcgt ccccgtcgac gacacgcggc 46140
tgaccgcgat gggcgccat caccggctgc tccggactg ggcgccccgg ccgacccggg 46200
caccogtget gcacctgtat gccggtgaac cggcgggcgc ctggccggat ccccggcagg 46260
actggcgttc gcgcttcgac ggcgcgaca ccagcgccga ggtgcccggc acccacttct 46320
cgatgatgac cgagcacgcc cccgtcaccg ccgcgaccgt gcacaagtgg ctgcacgagg 46380
tgtgcccgcc ccgcgttccg tgaccgtac gccgggtccg tcccggcgag tccgacgaca 46440
gcaggagagg aagcgcatga tcacagtccc gcccgacggg gatcccgcga cctgggcccg 46500
ccggctgcaa ctgaccgcg ccgcgcagt gttcgccggc aaccacggcg acccgctacgc 46560
gctgacctg cgcgcgga cgcgacccc gaccccgta gagcagcggg tggccgccc 46620
gccgctgttc cgcagcgagc agttggacac ctgggtgacc ggggacgccg cgtggcccg 46680
ggaggtgttg accgacgacc ggttcggctg gctgacccgg gctgggcagc ggcccgccga 46740
gcggaccctg ccgctggccg gcacggcact ggaccacggg ccggaggccc ggcgtcggct 46800
ggacgcgctc gccgggttcg gcgggcccgt cctgcgggccc gacgcccag gggcgctac 46860
ccgggtcgtg gagaccaccg cggctcctgt cgacgggatc ggggagcggg tcgacctggc 46920
cgtgctcgcc cggcggtgg tcgctgcggg gctggccgac ctgctggggg tgcccggcg 46980
gcggcggggc cgttcgccg aggcactcgc cgcgcgggc cgtacgctgg acagccggct 47040
gtgcccgcag accgtggcga ccgctctcgc caccgtcgcc gccaccgcc agctgaccga 47100
cctgctgggc gaggtgccgc cccgcgcgtc gctgtcccgc tccgcgcgcg gctccgggccc 47160
gccgcgtccg tccgcagccg gttcctggcc gccgctgccg gctgacgacc ggacggccgc 47220

cgcgctcgcg ctggcggtcg gcacggccga accggcgatc accctgctct gcaacgcggt 47280
cgggtgcgctg ctcgaccgcc ccgggcagtg ggccctgctc ggtggggacc tcgaccggtc 47340
cgccgccgtc gtcgaggaga ccctgcgctg ccttccgccg gtgcgcctgg agagccgcgt 47400
cgcgcagcag gacgtcacc c tggggcggca gttcctcccg gcggacagcc acctggctgt 47460
gctggtcgcc atggcgaaacc ggggtccgcg cgcggcgacc gccccgagcc cggacgcgtt 47520
cgaccctggc gggtcgcgcg tcccggcccc cgacgtggtg ggctgcgcg agcttgccgg 47580
cgccgggccg ctgatcagac tcgtcgtcac gaccgccctg cggaccctcg ccgaggcgct 47640
gcccacgctg cggcgggcgt ccggcggcgt ccggtggcga cgctgcgccg tcctgctcgg 47700
ccacgcccgc tttcccgctc caccggcgga gagcggcgaa cagcggctcc acgaccgccc 47760
ggcgctggag gaggcgatcc gatgcgcgtc ctgatgacgt ccttcgcgca caacaccac 47820
tactacagcc tgggtgccgtt ggctggggcg ctgcgcgcgg ccggccacga ggtacgggtg 47880
gcgagccagc cctcgctcac cgacaccatc gtgcggtcgg ggctgaccgc ggtgccggtc 47940
ggcgacgacc aggcgatcat cgacctgctc gccgaggtcg gcggcgacct ggtgccgtac 48000
cagcggggac tggacttcac cgaggccctg cccgaagtgc tgacctggga gtatctgctc 48060
gggcagcaga ccatgctcac cgcgctgtgc ttgcgcgcgc tcaacggcgt ctccacgatg 48120
gacgacatgg tcgccctggc ccggctcctg cagcccagac tgggtgatctg ggagccgttc 48180
acctacgccg ggccggtcgc ggcgccggtc gtccgtgcga cgcacgcccg gctgctctgg 48240
gggcccagc tggtcggcaa cggccggcg ctgttcaccg agagcctggc gcggcagccg 48300
gatgagcagc gcgaggacct gatggccgag tgggtgcgct gcacctgca ccggtacggc 48360
tgcgagctcg gcgacgacga ggtggagacc ctggtcaccg gcgggtggac catcgatccc 48420
accgccgaca gcacctggct tcccgctccc gggcgctcggg tggccatgcg gtacaccccg 48480
tacaacagcc cgtccgtggt gccggagtgg gtggccaagg ccgaccggcc ccgcgtctgc 48540
ctcaccctcg gcgtgtcgag ccgggagacg tacggcaggg acgtggctct cttccaggag 48600
ctgctcggcg ccctcggcga cctggacgtc gaggtcgtcg cgacgctcag cgacgcccag 48660
cgcgaggacc tgggtgacct gccggacaac gtccgggtgt gcgacttcgt gccgtggac 48720
gtgctgctgc cgacctgtgc cgcgatcatc caccacggcg gggcgggcac gtggtcgacg 48780
gccatgctct acgggggtgcc gcagatcatg atcgcgtcgc tgtgggacgc ccgctcaag 48840
gcgcagcagg cggagcgact cggcacgggg atctcgatcc cgccggagcg gctcgacgcc 48900
ccgacgctgc gggcggccgt cgtccggtat ctcgacgacc cgtcgatcgc cgccgccgcc 48960

cgccgtcagc gcgacgagct gcgtgccgcg ccgtcgccgg ccgaggtggt ccgcacacctg 49020
gaacgcctcg tcgcggacga ccggccccggc cggccggccg gaaccgccac cgaccactcc 49080
tgaaaggaac gatgtccatg atgtacgcgg acgccatcgc cgaggtctac gacctgatct 49140
accagggcaa gggcaaggac tacgcggcgg aggcggcgga gctggaggcg ctggcccggg 49200
cccgtcggcc gcacgcccgg acgtgctgg acgtggcgtg cggcacgggg ctgcacctgc 49260
ggcacctggc ggggctcttc gacgacgtgg gcggcatcga gctggcaccg gacatgctga 49320
gcatcgccca gcagcgaaac cccggggcgg ccctgcacct cggcgacatg cggaccttcg 49380
acctggggca ccgctacgac gtcacacct gcatgttcag ttcggtgggc cacctggcca 49440
ccacggccga gctggacgcg acgttgggcc ggctcgccgc gcacctgtcc cccgggggag 49500
tggcgatcgt cgagccgtgg tggttcccgg agaccttcac ccccggttac gtgggcgcga 49560
gcctggtgga ggtcgacggc cgtaccatct cgcgggtctc ccattcggtg cgcgaggcg 49620
gcgcgaccgc gatcacctg cactacctcg tggccagccc cggcggggga gtccggcact 49680
tcgacgagag ccacctgatc accctcttcg aacggtccga ctacgaacgt gccttcgccc 49740
gggcggggtt caccgacggag tacctgacgc ccggcccgtc cggccgcggt ctgttcgtcg 49800
gcgtccaccc ctgacgacc gttgccggtg cgcctcgacc cgcgcccccg acccgctgga 49860
ggaacagatg ccagacaccc ccgagctgaa ccggatactc gacgcgatcc tcgcccagga 49920
gaccgacgcg cgggagctgg cggccctgcc gctgccctcc tcctaccggg ccgtgacggt 49980
gcacaaggac gagacgggga tgttcctggg ccttccccgc caggagaagg acccgcgcaa 50040
gtcgtgcac acggaggagg tgccggtgcc cgagctgggc cccggggagg ccctcgtcgc 50100
ggtcctggcc agctcgtca actacaacac ggtctggtcg tcgttggtcg agccgtgcc 50160
caccttcggc ttcttgaggc gctacggccg gctctccgag ctggcccggc ggcacgacct 50220
gccgtaccac atcctcggct cggacctggc cggcgtggtg ctgagggtcg ggcccggcgt 50280
caaccgctgg cggccgggtg acgaggtcgt ggcgcactgc ctctcgggtg agctggagtc 50340
cgccgacggc cacggcgaca ccatgctcga cccggaacag cggatctggg gcttcgagac 50400
caacttcggc ggcctcgccg agatcgctt ggtcaaggcg aaccagctga tgcccaaacc 50460
cgaccacctg acctgggagg aggcgcgcgc gccgggactg gtcaactcca ccgcctaccg 50520
ccagctggtc tccggcaacg gggcccgat gaagcagggc gacaacgtcc tcgtctgggg 50580
ggccagcggc ggtctcggcg cgttcgccac ccagctcgtg ctggccggcg gggccaatcc 50640

cgtctgcgtg gtctccagcc cgcgcaaggc cgacatctgc cgtcggatgg gcgccgaggg 50700
 cgtcatcgac cgggtcgccg aggactaccg cttctgggtcc gacgagcgca cccagaatcc 50760
 ccgggagtggt aagcgcttcg gcgcacgcat tcgggagctg accggaggcg aggacgtcga 50820
 catcgtcttc gagcaccctg gccgggagac gttcggcgcc tcgggtctacg tgaccgcaa 50880
 aggaggcacc gtggtcacct gcgcctcgac gagcgggttc gagcacgtct acgacaaccg 50940
 ttacctgtgg atgtccctga agcgcatcgt cggcacgcac ttcgccaatt accgggaggc 51000
 gtgggaagcc aaccgggttg tggtaagggt caagatccac ccgacgctgt cgcgctgcta 51060
 cccgctggag gaggtcgccg aggcgggtcta cgacgtccat cacaacctgc accagggcaa 51120
 ggtcggcgtg ctgcgctcg cgcgcgcga ggggctcggg gtccggaacc cggagctgcg 51180
 ggaatgccat cttgccgcga tcaaccgctt ccgggtgccg gcctgacggg ccgcctttga 51240
 cgcccggggg cgcgcggtt ggcatgcggg cgaaccgggt gttaccgggc ggaagcaatt 51300
 ctcaactgca gtagttgcag ggtgcaccgg ctactgtgaa catatcgata gtcttatgta 51360
 gccatcgacc cccctgaatc ctctattcgt tgtgtgcgag gtggttggtg gcatgactgg 51420
 taccagcatt ccccgcggg accacgaact ccgattcttc gaacttctgg ccagggaggc 51480
 acccttaccg cagtacgagg aactggtgca ccaggcgac cgggacggag tggaccaggc 51540
 cagctcgac cgggtgatga tcgccaagcg actcgcgttg gagcttcgag aggtcatcgg 51600
 gaggcggtgt cagcggcagg cggagctggc cgccctcgtc gacaccgccc gtgacctcgc 51660
 cggggcgacg aacctggagg ccgggctgca gctggtggtg cggcggaacc aactgctgct 51720
 cgccggggac gtggcggttc tcagcctcgt cgacgacgag accggcgaat cctacgtcgc 51780
 ctcgccgctc gggcgggcca ccgcgctgac cagcggctac cggctgccct ggcgcgacgg 51840
 gctggtcgtg gccgcccac cgcgcgagcc actctcctgg acggcggaacc acctcgccga 51900
 cgagcgctc gaacgacacc cggccgcca cggcctggtc cgcgcggaag ggctgcacgc 51960
 ggtgctgtcc gtggttctga gcgtcgagg cggcacctc ggcaacctgc acgtcgcca 52020
 ccggcaggtc cgccacttcg ccccgagca ggtcgcgtcg ctgcgcctgc tcgccgatct 52080
 cgcggcgacg gcagtggagc ggatcatgct gctcgacgac acgtgggccc aactcaagca 52140
 ggcccagcag gaggcggcca gggcccagc cgagctgaac gcggtccgca tggccgaccg 52200
 cctgcaacct gaactcgtcc agctcatcct cgacggcggc gaactcgacg acctggtggg 52260
 cagcgccgtg cggcgactgg gcggcgccct gcacgtgcgt gaccgggcca acggcgtgct 52320
 ggcggcgcc ggtgaaatcc ctgtcccga cgagcgggaa ctggcccgag tgcggctgaa 52380

cgccacgcc accggccgac ccggccgcct gaccaccggt tcctgggtgg tgcccctggc 52440
 ggcccgcgcc ggtgacctcg gctgtgtgtt gttccacgcc gacgagccgt ccgacgacga 52500
 gcggatggcg gccctgccgg cggtcgcgca gaccgtggcg ctgctgatga ccaggaacgg 52560
 cgggagccac ggccagccgg gcgacgggct cctggaggac ctgctcggcc cgtggccgga 52620
 cctggagcgg ggccggaagc gccgtcggta cacacctgtc gagttcgacc ggccctacgt 52680
 cgtcgtggtc gcccgcgccg agggcgccac ctgcgcccg gtgttcgaac gggcgggtctc 52740
 cgtcgccac ggccgaacg gcatgaaggc catccgggac ggccaggcgg tgctgctgct 52800
 gcccggtgac gaccgggggg ccggggcccg ggacgtgacg cgggaactga gcgggctgct 52860
 cggcctaccg gtcacggccg gaggcgcccg accggtgcgc accggcgact cggtcagccg 52920
 cacctaccag gaggcgggcc ggtgcgtcga cgccctggcc gcgctggacg cgaaggggcg 52980
 ggccgctgc tcacgggacc tgggcttctt cgggctgctg gtcgccggcg gccacgacgt 53040
 caccggtttc gtcgaccggg tcatcggacc cgtgctgagc tacgacgcgc gccggctcac 53100
 gaatctcagg gagacctctc agacctactt cgaactggcg ggacgacgta cccgggcggc 53160
 ggagatgctg catctgcatc cgaacaccgt gtcccgcgg ctggaccgca tctcccagct 53220
 gctcgccgg gactggcggc agccggaccg ggccctcgac acgcagctcg ctctgcgcct 53280
 gcaccggatc cgtggcctgc tctgccagga acggggctac cggggcccat cgcaggagcc 53340
 ggaccaaccc gcgcggccta tccggcgga ccgccctcca gcatccgag ggcggtgcgc 53400
 acggacgcca aggtgacgtg ccggtgccag ccttggtatg accgaccctc gaagtcttgg 53460
 atgccgacgt cgaggctgac ctgggagaag tcggtctcga cccgccgggt gagcttgtctc 53520
 agccgcagca gtgggccgta cccggcgctg gtcatgttgg tcagccacat ctgccgtacg 53580
 ccgcgctcgt aggtctgcca cttgccgagc agtgtcaggg gcagcccggg cgcggcgggc 53640
 cgcgcgcgcc ccggcggggc cggggcgga ggaccggcg ggccggcacc ggacaggccc 53700
 ggccaataga cctgtagcgg tgcgaccagg ctgctgcgcc gtgcgccggg gctggccggg 53760
 tcgatccact ccaccggacg gcgctgggcc cgcgtcaggc tgagcagggt ctccggcgag 53820
 gccgccgca cccggttctc gcgcgggccc ggcccggcg ccagcagggt gcagccgctg 53880
 ttgatccgta gcaggaaggg cagaccgcgc gtggtgaacg cctcgatcag cgggggcagc 53940
 gccgagtgc gggcgtccat taccaccggg cgagggccga ttcccaggc cgcggccttc 54000
 agcaccgcct gcaccgccg gccgtcgtg gtcgtgccgt cctcgtccgc cggtagctc 54060

gcgcggggcgc ggtgtgcctg gagccaaccc ttaccgatgg acaactgcca gttgatgggc 54120
 gcggcgacgg tctccgaggg cagccagagg ccgtagctct gctggctgtt gaccgtctcg 54180
 cccagcgcgg gcacgtaccg gcgttccacg ccgaccgagt gccggccggg cttcggcacc 54240
 agcatcgacc gcaccaccca gggccggggc gacagcgtcc ggtccagggt gccggcgagc 54300
 gcggcacgga cggctctcca gtcccagggt gagcaactga tgaagtgggt catgctctgt 54360
 gccgcgcgcg gatcgctcggc gatggcggcc aggttgcgca tggctcttgc gccggaggcg 54420
 gtcagcagcc cccggacgta cagttcggcc ttgcgtcgct ggtcggcgcg gggcagcgag 54480
 gccagcaggg cggcgcgaga ccgggacacg gtgtccggat cggacottgc cgcggtcacc 54540
 tcctcgcgga cgtcgagcgt cggcaccatc actcccctcc tgcggcggga cgatgtgctg 54600
 atcacggcag acccgggccc ccggtccac catcgcccgg cgacgcctgc cttgcccagg 54660
 tgcgtcgga acacacttgg cgacgacggc gatcccgcac ccaccgcagc cccgccggtg 54720
 cgtgtccgtg gcgggccccg gcgggccccg gaccgggtga cgcccgaca catcgcggt 54780
 tcggcccgcg cgaagtgtgt gaccggcgaa cctcgcttcc cgcccgcca tccggaagcc 54840
 tgcaagggac cggaagcctt ccaacgagat tggcatcccc ccggcaaagg acccagatga 54900
 cctccgcagc gcaccattcc ccgcatccgg cgaaggccga cgccctgatg gacgacgccc 54960
 acgccgacat cggggccgat gccgaggccg acggtcgacg gctcgaccgg gccgccctgc 55020
 ggccgggtcgc cgggctgtcg accgagaggg ccgacgtcac ggaggctcag taccggcagg 55080
 tgcggctgga gcgcgtcgtc ctggctggcg tgtggacctc gggcacccgc gacgaggccg 55140
 aacggctcct cgccgagctg gcggcactcg ccgagaccgc gggagccgtg gtgctcgacg 55200
 ggggatcca gcgccgcgac cggcccgacc cggcgacgta catcggtcc ggcaaggcgc 55260
 gggagtgtcg ggacatcgtc caggagggtg gggccgacac ggtgatctgc gacggtgagc 55320
 tgagcccggc ccaactggta cgctcgaag aggtcgtcga cgccaagggt gtggaccgca 55380
 ccgcgctgat cctcgacatc ttgcgccagc acgccacgtc ccgcgagggg aaggcgcagg 55440
 tggccctggc acagatgcaa tacatgctgc cgcggctcgc cggctggggc cagtgcgtct 55500
 cccggcagat gggcggagggt gccggcgggc gtggcatggc caccggggg cccggcgaga 55560
 ccaagatcga gaccgaccgg cggcgcatcc acgagaggat ggcccggtc cgacgggaga 55620
 tcgcggagat gaagtccggc cgcgaactca agcgcgcga tcggcgggcg aacagcgtcc 55680
 cgtcggtcgc gatcgccggg tacaccaacg ccggcaagtc ctgctgctc aaccggctca 55740
 ctggcgcgag cgtgctgggt cagaacgcgc tgttcgccac cctcgaccgc acggtgcgc 55800

gggccaccac cccgagcggg cgcagctaca cgatcaccga caccgtcgga ttctgcccgc 55860
acctgccgca ccacctggtg gaggcgttcc gctccaccct ggaagaggtg gccgagggcg 55920
acctctgct gcacgtggtg gacggcgccc accccgcccc gctggagcag ctgcctcgg 55980
tgcgcgcggt catccgggac gtggacggcg cgggagtgcc cgaactcgtc gtgatcaaca 56040
aggccgacgc cgccaccccg gccgccttgg ccgcgttggc ggaggccgag ccgcaccacg 56100
tcgtcgtctc ggcccgacc ggtcagggca tcgacacgct tcggcagttg ctggaggccg 56160
cgctgccgca ccgggaggtc cgggtogaag tcttgatccc gtacgtcgcg ggcagcctcg 56220
tgggccgggt gcacgccgac ggcgaggtgc tggccgagga gcacacggcc gacggcacc 56280
tgctgcaggc gcgggtggcc cccgacctgg ctgccgagct cagcgcgtac gccaggacct 56340
gagcgtcgcc gcccccggg cggcatccgg agctggcgaa gctgtggccc gtagaggag 56400
gcaggcgatg aagcgagatc tcggggatct ggactcttc ggaggacacg ccagcttct 56460
ccagcagatc cacgtcgggc gcccacccg gatcgatcgg gccaggctgt tcgaccggct 56520
gtcctgggcg ctcgacaacg agtggttgac caacaacggg ccgctggcac gggagttcga 56580
ggagcgggtc gccgacatgg tcggggtcgg caactgcgtg gcgacgtgca acgccacgg 56640
ggccctccag ctgctcgcg acgccaccga gctgaccggg gaggtgatca tgccatcgct 56700
caccttcgcc gcgaccgcac acgcggtgcg ctggctcggg ctggagccgg tcttctcgca 56760
catcgaccgg cgcaccggat gcctcgacca cgtggcggtc gccgcggcca tcacgcccg 56820
cacgtcggcg gtcttcggcg tcacacctg gggccgcccc tgcgacgtca acgcgtgga 56880
gaaggtgacc gccgacggcg gcctgcgcct gttcttcgac gccgcccacg ccatcgggtg 56940
cacctcacag ggccgcccgg tggggcggtt cggccacgcc gaggtgttca gcttccacgc 57000
gacgaaggtc gtcaacgcct tcgagggcg ggcgatcgtc accgacgacg acgacctcg 57060
ccaccgcgtc cgctccctgg cgaacttcgg cttcggcctg cacagcccca gcgggcccg 57120
cggcaccaac gcgaagatga gcgagggcgt cgcgcctatg gggctcacct cgctcgacgc 57180
gttccccgag gtggcccgcc acaaccaggc caactacgag cagtactgcg gtgagctggc 57240
ccggtattcc ggctcagcg tgatcgactt cggcccgac gagcggcaca actaccagta 57300
cgtgatcgtc gagatcgacc cggacgtcac cgggttgac cgcgacctgc tcgtcgacct 57360
gctccgggcc gagaacgtcg tggcgagcg ctacttctcg ccggcctgtc accaattgga 57420
gccctaccgg tcccggcagc agttccagct gccgcacacc gagcggctct cggcgcgcg 57480

cctggcgctg ccgaccggct ccgccatctc ccgggaagac atccgcaggg tgtgcaacat 57540
 cgtgcggttg gcggtctccc ggggattcga attgaccgct cgggtggcagc agcagcccgg 57600
 gcccgcagga cagagcgtgg tggcaccggg ttgaccgaac ggcaccggac ggacgtgtgg 57660
 gagggcccgt gaccatggag atctccgcct cgaatcccgt ggcgacctgc gctgtccccg 57720
 gcagcgaccc gaccgcggcg gcgcgcgtgc tgtacgacga ggtcgccggg tcaggaatcg 57780
 tgccgcggc agagatcggg gcgcgcgcc aggggttggg ggcatggca cgcactacg 57840
 ggaccacacc ttttctgccg cttgagcagg ccgcgcgca aatcgccctg gaccggggcg 57900
 ggttcggcg gctgctggac ctgttcgccc ggattcccgg gttgcgcacc gcagtggaga 57960
 acggaccgtc cggtcgctac tggaccaaca cggtgctcgg cctcgaaagg gccggcgtct 58020
 tcgacgccgt gctcgaccgg aggcggcggt ttccgcactc cgtcgggctc taccggggcc 58080
 ccacgtgcat gttccgctgt cacttctgcg taagggtcac cggggcccgc taccaggcct 58140
 cggcgctgga cgacgggaac gccatgttcg cctctgtcat cgacgaggtc ccgcgcaca 58200
 accgcgacgc ggtgtacgtc tccggtggcc tcgagccact caccaacccc gggctcgggtg 58260
 cactggtcag ccgggcggcc gagcggggat ttccgatcat cctctacacc aactcgttcg 58320
 ccctcacgga gcagaagctc aagggtgagc ggggattgtg gagcctgcac gccatccgca 58380
 cgtcgctgta cgggttgaac gacgaggaat accgggcgac caccggcaag cagggggcct 58440
 tcaccgggt acgggcgaac ctcacgggt tccagcagct gcgtgccgag cggggcgagc 58500
 cggtgccgct cggcctcagc tacatcgtcc tgcccggccg cgccgggcgg ctgagcgcgc 58560
 tgatcgactt cgtcgccgag ctcaacgagg cggcaccgga ccgcccgctg gactacatca 58620
 acctgcggga ggactacagc gggcgccggg acgggaagct ctccctggac gagcgcgccg 58680
 agctccaggc cgagctgcac cggttccggg agagggcaat gcagcggacg ccgaccctgc 58740
 acatcgacta cggctacgcc ctgcacagcc tgatgacggg aagcgacgtg gagctcgtgc 58800
 gtatccggcc ggagacgatg cgccctgcgg cccaccgca ggtgtcgggtg caggtggata 58860
 tcctcgggtga tgtctacctc tatcgggagg cggcgtttcc gggcctggcc ggtgccgacc 58920
 gctatcgcat cggcacggta tctcccggca cgacgttggc gcaggtgggtg gagacgttcg 58980
 tgaccagcgg cggatcgggtg gtcgcgaagc ctggcgacga atacttctg gacggattcg 59040
 accaggcggg gaccgcgcgg ctgaaccaga tggagaccga cgtcgccgat ggctggggag 59100
 accgacgggg tttcctccgc tgatggagat cgactggtga gagcgggtgg ccaacgccga 59160
 agaaagccag ttgccgggtg ccgcaccgc cgtttcagtc gtcgggtata gtgcccgtca 59220

tggctgttgt gtgcttcatg aggctccgcc gcgcatacg gcggaccatc gcttctcttg 59280
 atgagtgtcg ccgcccacg ggtcactgcc ggtgcggcgt tccctgccga ccggtccga 59340
 acgatattcg cggagcacgc acatgcccta catccagcac gccgggcgac atgaattcgg 59400
 ccagaatttc ctggtcgacc gctcgggtgat cgacgatttc gtcgaactcg tcgcccggac 59460
 cgacggccct atcgtggaga tcggcgccgg cgacggtgcg ctgaccctac ccctgagccg 59520
 gcagggaagg gaggtagccg cagtggagat cgactccaag cgttccaagc gggtcagccg 59580
 gcagacaccc gacaacgtca ccgtggtctg cgcggatgtc ctgagcttcc ggttccccca 59640
 gcatccgcac gtggtcgtcg ggaacatccc cttccacgtg accacccccca tcgtgcgggc 59700
 tctcctcgcc gcggaccact ggcacacggc ggtgctgctg gtgcagtggg aggtggcccc 59760
 caggcgggcc ggctcggcg gcgcgacgt gctgaccgcg agctggtggc cctggtacga 59820
 cttcgaactg cactcccggg ttccggcccc cgcttccgg cctgtccctt ccgtcgacgg 59880
 cgggctgttc tccatgggtcc gtcgcgggac cccgctggtc gacgaccgga ggggttacca 59940
 ggaattcgtc cggctggtgt tcaccggcaa ggggcacgga ttgccggaga tccttcagcg 60000
 gaccgggcgg atcgcccgca aggaccagca ggactggcaa cgggccaacc ggggtggggcc 60060
 gcagcacctg cccaaggacc tgaccgcccc ccagtgggcc tccctgtggc acctggtggc 60120
 acccgcccg cggcgccggc cccgccgtcc ggcaccgcgc cggccaggaa gccccgttc 60180
 ggcgcgccgg cgctga 60196

<210> 2
 <211> 560
 <212> PRT
 <213> micromonospora carbonacea subspecies aurantiaca

<400> 2

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

Ala Pro Gly Gly Ile Gly Tyr Leu Ala Gln Arg Leu Arg Leu Pro Ala
 85 90 95
 Gly Gly Ser Thr Val Arg Asp Val Val Asp His Thr Leu Ala Asp Leu
 100 105 110
 Arg Asp Leu Glu Ala Arg Leu Arg Ala Ala Glu Ala Asp Leu Ala Thr
 115 120 125
 Ala Thr Pro Glu Gln Leu Asp Ala Tyr Gly Thr Leu Leu Thr Val Phe
 130 135 140
 Glu Ala Arg Gly Gly Tyr Gln Ala Asp Ala Arg Val Asp Ala Ala Leu
 145 150 155 160
 His Gly Leu Gly Leu Ala Glu Leu Asp Arg Asp Arg Asp Val Asp Thr
 165 170 175
 Leu Ser Gly Gly Glu Arg Ser Arg Leu Ala Leu Ala Ala Thr Leu Ala
 180 185 190
 Ala Ala Pro Glu Leu Leu Leu Leu Asp Glu Pro Thr Asn Asp Leu Asp
 195 200 205
 Ile Glu Ala Val Glu Trp Leu Glu Asp His Leu Arg Ser His Arg Gly
 210 215 220
 Thr Val Val Val Val Thr His Asp Arg Val Phe Leu Glu Ser Val Thr
 225 230 235 240
 Ser Thr Ile Leu Glu Val Asp Thr Asp Thr Arg Ala Val His Arg Tyr
 245 250 255
 Gly Asp Gly Tyr Ala Ser Tyr Leu Arg Ala Lys Ala Ala Leu Arg Glu
 260 265 270
 Ser Arg Glu Arg Ala Tyr Ala Glu Trp Val Ala Glu Val Glu Arg Gln
 275 280 285
 Ser Gln Leu Ala Glu Arg Ala Gly Thr Met Leu Arg Ser Ile Ser Arg
 290 295 300
 Lys Gly Pro Ala Ala Phe Ser Gly Ala Gly Ala His Arg Ser Arg Ser
 305 310 315 320
 Ser Ser Thr Ala Thr Ser Arg Lys Ala Arg Asn Ala Asn Glu Arg Leu
 325 330 335
 Arg Arg Leu Arg Glu Asn Pro Val Pro Arg Pro Ala Asp Pro Leu Arg
 340 345 350
 Phe Thr Ala Ser Val Ala Pro Asp Ala Thr Asp Ala Asp Thr Arg Arg
 355 360 365
 Val Glu Leu Thr Asp Val Arg Val Gly Arg Arg Leu His Val Pro Glu
 370 375 380

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

<210> 3

<211> 1683

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 3

gtgccagttc cgacacagga ggcccccttg cggaacagcc cgccgccagc ccattcgcag 60
 ctcgtcctga gcgaggtcac gaagcactac gccgagcggg tcgtcctgga ccgcgtttcg 120
 ctcaccgtca agccggggga gggggtcggc gtcacgcggc agaacggggtc ggggaagtcg 180
 accctgctgc ggctcgtcgc ggggctggag acgccggaca acggcgagtt gaccgtctcg 240
 gcgccccggg gcacgcgcta tctcgcccag cggcttcggc tgccggccgg cggcagcacc 300
 gtacgggatg tggaggacca cacgctcgcc gacctgcgag acctggaggc gcggttgccg 360
 gccgccgagg cggacctggc caccgccacg cccgagcagt tggacgccta cggcacgctg 420
 ctcactgtgt tcgaggcccc gggcggctac caggccgacg cccgggtgga cgccgccctg 480
 cacggtctcg gcctggccga gctcgaccgc gatcgcgacg tcgacacgct ctccggcggg 540

```

gaacgggtccc ggctcgcgct cgccgcgacc ctggccgccc cgccggaact gctgctgctc 600
gacgagccca ccaacgacct cgacatcgag gccgtggagt ggctggagga tcacctgcgg 660
tcgcaccggg gcaccgtcgt cgtggtcact caccgaccggg tgttcctgga gtcggtcacg 720
tccaccatcc tcgaggtcga caccgacacc cgggccgtgc accggtacgg cgacggctat 780
gccagctacc tcggggcca ggcgcacctc cgggagagcc gggagcgcgc gtacgcggaa 840
tggggtggccg aggtcgagcg gcagtcccaa ctgcgcggagc gggccgggac gatgctccgg 900
tcgatctccc gcaagggacc ggctgcggtc agcggggccg gtgcccaccg ctcccggctc 960
tcgtcgacgg cgacgtcacg caaggcccg c aacgccaacg agcggcttcg ccggctgcgg 1020
gagaatccgg taccgcgacc cgccgacctg ttgcgcttca ccgcgtcggc cgccccggat 1080
gccacggacg ccgatacccg ccgcgtcgag ttgaccgacg tccgggtggg ccgccgcctg 1140
cacgtgcccc agctgacct cggaccgcgc gaacgggttc tggtgaccgg acccaacggc 1200
gcgggtaaga gcacctgat gcgggtgctc gccggggaac tcgtgcccg cggcggaacg 1260
gtgcggctgc cggctcggat cggccacctg cgtcaggacg tgacggtcgg gcagcccggg 1320
cgctctctgc tggagacgta ccgcgtcggg cggccggggc atcccagga gtacgcggag 1380
gagttgctcg cccgcggtct gttccggccc gatgacctgc gcatgccggt cgggacgctc 1440
tccgtcgggc agcgcgcgc gatcgacctg gcccggtgg tcgcccgcgc ggccgacctg 1500
ctgctgttgg acgagcccac caaccacttc gcgccctgc tcgtggagga gctggaacag 1560
gcgctggacg gctacgccgg agcgtcggtc gtggtgacgc acgaccggcg gatgcggagc 1620
accttcaccg gggctcggct ggaactgcac cagggcgtgg ccaccggggc gagccggggc 1680
tga 1683

```

<210> 4

<211> 264

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 4

```

Met Ser Pro Ser Ala Asp Pro Ser Glu Leu Trp Leu Arg Arg Tyr Arg
1           5           10           15

```

```

Pro Val Asn Asp Pro Ala Val Arg Leu Phe Cys Phe Pro His Ala Gly
          20           25           30

```

```

Gly Ala Ala Ser Ala Tyr Leu Pro Phe Ala Arg Arg Leu Ala Ala Asp
          35           40           45

```

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

<210> 5

<211> 795

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 5

atgtccccgt ccgccgatcc gtccgagctg tggctacgcc gctaccggcc cgtcaacgac 60
 cccgccgtcc ggctgttctg cttccgcac gccggggcg cgccagcgc gtacctgccg 120
 ttcgcccgcc ggctcgccgc cgacgtggac gtgctggcg tccagtacc gggccggcag 180
 gaccgccgc gcgaaccctt gatcgagtcc gtcgacgcc tggtagcgg gctcctgcc 240

gcactgctcg cctgggcgga ccgaccggtg gccttcttcg gtcacagcat gggcgccacg 300
gtggccttcg aggccgcccg ccggctccca ccggccgacg ccgatcggct cgtgcacctc 360
ttcgcctccg gccgccgtag cccgtccgtc gggcgggcggg accggttcta ccggtttgac 420
gacgagctga tcgacgagat ccgccggctc cagggcaccg attccagcct cctggacgac 480
agggaaactgc tggacatgct cctccccgcc atccgcaacg actaccgggc cgccgccgcc 540
tacgaatacc ggccagggcc caggctgcgt tgcccgggtca ccgtactcgc cggggccgcc 600
gacacccacg tcaccaccga cgaggccgcg gcgtggggccg aggtgaccgc agcggccacg 660
atgggtccga cgttcccggg cgggcacttc tatctcaacg atcagctcga cgtgtgtgtgc 720
gccgaggtca cgaccaccct cgcagcgggtg tccacgaccg ccctcacggc ggtgccgggt 780
gccgaccccg gctga 795

<210> 6

<211> 410

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 6

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

Glu Lys Phe Arg Glu Trp Val Asp Ile Ile Leu Ser Asn Glu Gly Leu
 165 170 175
 Glu Tyr Pro Asn Leu Pro Asp Asp Phe Thr Glu Thr Val Gly Pro Ala
 180 185 190
 Ile Glu Glu Trp Ser Glu Phe Leu Tyr Ala Gln Ile Ala His Lys Arg
 195 200 205
 Ala Glu Pro Lys Asp Asp Leu Ile Ser Gly Leu Cys Ala Ala Glu Val
 210 215 220
 Asp Gly Arg Lys Leu Thr Asp Glu Glu Val Val Asn Ile Val Ala Leu
 225 230 235 240
 Leu Leu Thr Ala Gly His Ile Ser Ser Ala Thr Leu Leu Ser Asn Leu
 245 250 255
 Phe Leu Val Leu Glu Glu His Pro Gln Ala Gln Ala Ala Val Arg Ala
 260 265 270
 Asp Arg Ser Leu Val Pro Gly Val Ile Glu Glu Thr Leu Arg Tyr Arg
 275 280 285
 Ser Pro Phe Asn Cys Ile Phe Arg Ile Leu Asn Glu Asp Thr Asp Ile
 290 295 300
 Leu Gly His Pro Met Arg Lys Gly Gln Met Val Ile Ala Trp Ile Ala
 305 310 315 320
 Ser Ala Asn Arg Asp Thr Glu Val Phe Thr Asp Pro Asp Thr Phe Asp
 325 330 335
 Ile Arg Arg Glu Ser Asn Lys His Leu Ala Phe Gly His Gly Ile His
 340 345 350
 His Cys Leu Gly Ala Phe Leu Ala Arg Leu Glu Ala Lys Val Phe Leu
 355 360 365
 Asn Gln Thr Leu Asp Gln Phe Thr Glu Phe Arg Ile Asp His Val Gly
 370 375 380
 Val Glu Phe Tyr Asp Ala Asp Gln Leu Thr Ala Arg Arg Leu Pro Val
 385 390 395 400
 Gln Val Val Arg Asp Gly Arg His Pro Lys
 405 410

<210> 7

<211> 1233

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 7

atgacgcaga ccccgaaacgc cccggcgagg cccgatcgacc tgcccaaggg cgccgacgcc

60

```

caggggctgc tggactgggt cgcgtacatg cggaagaact ggcccgtctc ctgggacgag      120
acccgtcagg cctggcacgt gttctcctac cgggactacc agaccgtgac caccaacccg      180
ctgatcttct cgtcggactt cacctcggtc tttcccgta cgtcggagct ggccctgctg      240
atggggcccc gcaccatcgg cggcatcgac ccgcccgggc acgcgcgcgt gcgcaagctg      300
gtgagccagg cgttcacccc ccgccggatc gccagatgg agctgcggat cgggcagatc      360
accgccgacg tgctcgacca ggtacgcgac caggaccgga tcgacatcgc cagcgacctc      420
gcgtacccgc tgccgggtgac ggtcatcgcc gagctgctcg gcattccac caaggatcac      480
gagaagttcc gcgagtgggt ggacatcatc ctacgaacg aagggtgga gtatcccaac      540
ctcccggacg acttcaccga gacggtgggc ccgccatcg aggagtggtc cgaattcctg      600
tacgcccaga tcgccacaa gcgcgccgaa ccgaaggacg acctgatcag cggcctctgt      660
gcggcggagg tcgacgggcg caagctgacc gacgaggaag tcgtcaacat cgtcgcgctg      720
ctgctcaccg ccgggcacat ctccagcgcc acgctgctca gcaacctgtt cctggtgctg      780
gaggagcacc cgcaggcaca ggccgcggtc cgcgccgacc gcagcctcgt gccgggcgtg      840
atcgaggaga cgtcgcgcta ccggtccccg ttcaactgca tcttcggat cctgaacgag      900
gacaccgaca tctcggcca ccccatcgcc aagggccaga tggatgatcg ctggatcgcc      960
tccgcgaacc gcgacaccga ggtgttcacg gaccgggaca ccttcgacat ccgacgcgag     1020
tcgaacaagc acctggcggt cggccaaggc atccaccact gcctgggcgc gttcctggcc     1080
aggctggagg cgaaggtctt cctcaaccag acgctcgacc agttcaccga gttccggatc     1140
gaccacgtcg gggtcgagtt ctacgacgcc gaccagctca ccgcgcgacg cctccccgtc     1200
caggtggtac gcgacggacg gcacccgaag taa                                     1233

```

<210> 8

<211> 402

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 8

```

Met Glu His Pro Val Thr Ala Gly Ser Cys Arg Phe Tyr Pro Phe Ser
1           5           10           15

Asp Arg Thr Asp Leu Asn Ile Asp Pro Thr Tyr Gly Glu Leu Arg Ser
          20           25           30

Lys Glu Pro Val Ala Arg Val Arg Met Pro Tyr Gly Gly Asp Ala Trp
          35           40           45

```

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

Gly Ala Ser Leu Ala Arg Leu Glu Ile Gln Ala Ala Ile Gly Ala Leu
 355 360 365

Ala Arg Arg Leu Pro Gly Leu Arg Leu Ala Val Glu Glu Asn Glu Leu
 370 375 380

Asp Trp Asn Lys Gly Met Met Val Arg Ser Leu Val Ser Leu Pro Val
 385 390 395 400

Thr Trp

<210> 9

<211> 1209

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 9

atggagcatc cagtaacggc cgggtcctgc aggttctacc ccttcagtga ccgtaccgac	60
ctgaatatcg atcccacgta cggcgaaactg cgctcgaaag agccggtcgc ccgcgtccgc	120
atgccctacg gcggggacgc ctggctggtc acccggcacg ccgacgcca gaaggccctc	180
tctgaccccc gactcagcat tgcagccgga gccgggacgg acgtgccgcg cgcctcccc	240
cgtctccagg aaccgcacgg tctgatgggt cttccccccg acgcgcacgc ccgactgcgc	300
aggctcgtcg ccacggcggt caagccgaag cgcgtacggg acatcgcccc gcgcgtcgtc	360
cagctcgccg acaagcttct cgacgacgtg gtcgaaaccg ggccgcccgc cgacctcgtg	420
cagcagctcg cgcttccctt gccgggtgatg atcatctgcg agatgatggg catcgggtac	480
gacgagcagc acctgttccg tgccctcagc gatgccctga tgtcctccac ccgatacagc	540
gccgaccagg tcgaccgcgc ggtagaggac ttcgtcgagt acctcgggcg cctcctcgcg	600
cagcgccgtg cacaccgcac cgacgacctc ctccggcgccc tggtcgaggc gcgagacgac	660
ggcgatcggc tgaccgagga cgaactcgtc atgctcaccg gcggcctgct cgtcggcggc	720
cacgagacga ccgccagcca gatcgccctg cagatcttcc tcttgetgcg cgaccggacc	780
aggtacgagc aactccatgc ccgtccggag ttgatcccca cggcagtcga ggaactgctg	840
cgggtggccc cgctctgggc ctcggtcggc cccacccgca tcgccaccga ggacctggaa	900
ctcaacggga cgaccatccg ggccggcgac gccgtcgtct tctcgtggc gtccgccaat	960
caggacgacg acgtcttcgc gaatgccgca gacgtcgtgc tcgaccgga cccgaatccg	1020
cacatcgctt tcgggcacgg gccccattac tgcacgggg cgtaactggc cagactggaa	1080
atacaggccg ccatcggcgc cttggccagg cggcttcccg gtctccgcct ggccgtcgag	1140
gaaaacgaac ttgattggaa caagggaatg atggtacgca gcctcgtgtc ccttcgggtg	1200

acgtggtga

1209

<210> 10
 <211> 4471
 <212> PRT
 <213> micromonospora carbonacea subspecies aurantiaca
 <400> 10

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

245					250					255					
Glu	Gly	Gly	Val	Ala	Met	Val	Leu	Lys	Pro	Leu	Ile	Arg	Ala	Ile	Glu
			260					265					270		
Asp	Gly	Asp	Gln	Val	His	Cys	Val	Ile	Arg	Gly	Cys	Ala	Val	Asn	Asn
			275					280					285		
Asp	Gly	Gly	Gly	Pro	Ser	Leu	Thr	His	Pro	Asp	Arg	Glu	Ala	Gln	Glu
			290					295					300		
Ala	Leu	Leu	Arg	Arg	Ala	Tyr	Glu	Arg	Ala	Gly	Val	Ala	Pro	Glu	His
			305					310					315		320
Val	Asp	Tyr	Val	Glu	Leu	His	Gly	Thr	Gly	Thr	Lys	Ala	Gly	Asp	Pro
				325										335	
Val	Glu	Ala	Ala	Ala	Leu	Gly	Ala	Val	Leu	Gly	Val	Ala	Arg	Gly	Cys
				340										350	
Asp	Asn	Pro	Leu	Ala	Val	Gly	Ser	Val	Lys	Thr	Asn	Val	Gly	His	Leu
				355										365	
Glu	Gly	Ala	Ala	Gly	Ile	Thr	Gly	Leu	Leu	Lys	Ala	Val	Leu	Cys	Val
				370										380	
Arg	Glu	Gly	Val	Leu	Pro	Pro	Ser	Leu	Asn	Phe	Arg	Thr	Pro	Asn	Pro
								390							400
Asp	Ile	Arg	Leu	Asp	Glu	Leu	Asn	Leu	Arg	Val	Gln	Thr	Glu	Leu	Gln
				405											415
Pro	Trp	Pro	Gly	Asp	Gly	Thr	Gly	Arg	Pro	Arg	Val	Ala	Gly	Val	Ser
				420										430	
Ser	Phe	Gly	Met	Gly	Gly	Thr	Asn	Ala	His	Leu	Ile	Leu	Glu	Gln	Ala
				435										445	
Pro	Val	Ala	Ala	Glu	Glu	Thr	Ala	Val	Thr	Asp	Ala	Gly	Val	Gly	Ser
				450										460	
Val	Arg	Val	Val	Pro	Val	Val	Val	Ser	Gly	Arg	Ser	Val	Gly	Ala	Leu
														480	
Arg	Ala	Tyr	Ala	Gly	Arg	Leu	Arg	Glu	Val	Cys	Ala	Gly	Leu	Ser	Asp
				485										495	
Gly	Gly	Gly	Ser	Gly	Gly	Gly	Ser	Gly	Leu	Val	Asp	Val	Gly	Trp	Ser
				500										510	
Leu	Val	Ser	Ser	Arg	Ser	Val	Phe	Glu	His	Arg	Ala	Val	Val	Phe	Gly
				515										525	
Gly	Gly	Val	Ala	Glu	Val	Val	Ala	Gly	Leu	Asp	Ala	Val	Ala	Ser	Gly
				530										540	
Ala	Val	Ser	Ser	Gly	Ser	Val	Val	Val	Gly	Ser	Val	Ala	Ser	Gly	Val

545		550		555		560
Ala Gly Gly Gly Gly Arg Val Val Phe Val Phe Pro Gly Gln Gly Trp						
	565			570		575
Gln Trp Val Gly Met Gly Ala Ala Leu Leu Asp Glu Ser Glu Val Phe						
	580		585			590
Ala Glu Ser Met Val Glu Cys Gly Arg Ala Leu Ser Gly Phe Val Asp						
	595		600			605
Trp Asp Leu Leu Glu Val Val Arg Gly Gly Gly Gly Asp Gly Ser Phe						
	610		615			620
Gly Arg Val Asp Val Val Gln Pro Val Ser Trp Ala Val Met Val Ser						
	625		630		635	640
Leu Ala Arg Leu Trp Met Ser Val Gly Val Val Pro Asp Ala Val Val						
	645		650			655
Gly His Ser Gln Gly Glu Val Ala Ala Pro Val Val Gly Gly Val Leu						
	660		665			670
Ser Val Ala Asp Gly Ala Arg Val Val Ala Leu Arg Ser Arg Val Ile						
	675		680			685
Gly Glu Val Leu Ala Gly Gly Gly Ala Met Val Ser Val Gly Leu Pro						
	690		695			700
Val Ala Val Val Leu Asp Arg Leu Ala Gly Trp Gly Gly Arg Leu Gly						
	705		710		715	720
Val Ala Ala Val Asn Gly Pro Ser Leu Thr Val Val Ser Gly Asp Val						
	725		730			735
Asp Ala Ala Val Gly Phe Val Gly Glu Cys Glu Arg Asp Gly Val Trp						
	740		745			750
Val Arg Arg Val Ala Val Asp Tyr Ala Ser His Ser Ala His Val Glu						
	755		760			765
Ala Val Glu Gly Met Leu Ser Gly Leu Leu Gly Gly Leu Cys Pro Gly						
	770		775			780
Arg Gly Val Val Pro Phe Tyr Ser Ser Val Val Gly Gly Val Val Asp						
	785		790		795	800
Gly Val Gly Leu Asp Gly Gly Tyr Trp Tyr Arg Asn Leu Arg Glu Arg						
	805		810			815
Val Leu Phe Ser Asp Val Val Gly Arg Leu Val Gly Asp Gly Phe Ser						
	820		825			830
Gly Phe Val Glu Cys Ser Gly His Pro Val Leu Ala Gly Gly Val Leu						
	835		840			845
Glu Ser Val Ala Val Val Asp Pro Asp Val Arg Pro Val Val Val Gly						

850	855	860
Ser Leu Arg Arg Asp Asp Gly Gly Trp Gly Arg Phe Leu Thr Ser Val		
865	870	875 880
Gly Glu Ala Phe Val Gly Gly Met Ser Val Asp Trp Lys Gly Val Phe		
	885	890 895
Ala Gly Ala Gly Ala Arg Leu Val Asp Leu Pro Thr Tyr Pro Phe Gln		
	900	905 910
Arg Arg His Tyr Trp Ala Pro Asn Thr Asp Gly Ala Pro Ala Pro Ile		
	915	920 925
Leu Asp Asp His Ala Glu Ala Glu Asn Glu Pro Ala Glu Ser Glu Pro		
	930	935 940
Gly Ile Arg Ala Glu Leu Leu Thr Leu Ala Glu Pro Glu Gln Leu Asn		
945	950	955 960
Arg Leu Leu Ala Thr Val Arg Ala Ser Thr Ala Val Val Leu Gly Leu		
	965	970 975
Asp Ser Ala Gln Ala Val Asp Pro Glu Arg Thr Phe Lys Glu His Gly		
	980	985 990
Phe Glu Ser Val Thr Ala Val Glu Leu Cys Asn His Leu Gln Arg Gly		
	995	1000 1005
Thr Gly Leu Arg Val Pro Ala Ser Leu Val Tyr Asn His Pro Thr		
1010	1015	1020
Pro Met Ala Ala Ala Arg Lys Leu Gln Glu Glu Ile Gln Gly Arg		
1025	1030	1035
Gln Pro Glu Asn Val Arg Gln Val Thr Ser Ala Ala Ala Val Asp		
1040	1045	1050
Asp Pro Val Val Val Val Gly Met Gly Cys Arg Phe Pro Gly Gly		
1055	1060	1065
Val Val Cys Ala Glu Gly Leu Trp Asp Leu Val Leu Gly Gly Gly		
1070	1075	1080
Asp Ala Val Ser Gly Phe Pro Val Asp Arg Gly Trp Asp Val Glu		
1085	1090	1095
Gly Leu Phe Asp Pro Val Arg Gly Val Val Gly Lys Ser Tyr Val		
1100	1105	1110
Arg Glu Gly Gly Phe Val Tyr Asp Ala Gly Met Phe Asp Ala Glu		
1115	1120	1125
Phe Phe Gly Val Ser Pro Arg Glu Ala Val Ala Met Asp Pro Gln		
1130	1135	1140
Gln Arg Leu Phe Leu Glu Val Ser Trp Glu Ala Leu Glu Arg Ala		

1145		1150		1155
Gly Ile Asp Pro Leu Gly Leu Arg Gly Ser Arg Thr Gly Val Tyr				
1160		1165		1170
Val Gly Val Met Gly Gln Glu Tyr Gly Pro Arg Leu Val Glu Ser				
1175		1180		1185
Gly Gly Gly Phe Glu Gly Tyr Leu Leu Thr Gly Thr Ser Pro Ser				
1190		1195		1200
Val Val Ser Gly Arg Val Ser Tyr Val Leu Gly Leu Glu Gly Pro				
1205		1210		1215
Ser Ile Ser Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Leu				
1220		1225		1230
His Leu Ala Cys Gln Gly Leu Arg Leu Gly Glu Cys Asp Val Ala				
1235		1240		1245
Leu Ala Gly Gly Val Thr Val Ile Ala Ala Pro Gly Leu Phe Val				
1250		1255		1260
Glu Phe Ser Arg Gln Gly Gly Leu Ser Gly Asp Gly Arg Cys Arg				
1265		1270		1275
Ala Phe Ala Gly Gly Ala Asp Gly Thr Gly Trp Gly Glu Gly Ala				
1280		1285		1290
Gly Val Val Val Leu Glu Arg Leu Ser Val Ala Arg Glu Arg Gly				
1295		1300		1305
His Arg Val Leu Ala Val Val Arg Gly Ser Ala Val Asn Gln Asp				
1310		1315		1320
Gly Gly Ser Asn Gly Leu Thr Ala Pro Ser Gly Val Ala Gln Arg				
1325		1330		1335
Arg Val Ile Gly Ala Ala Leu Val Ala Ala Gly Leu Gly Val Ser				
1340		1345		1350
Asp Val Asp Val Val Glu Ala His Gly Thr Gly Thr Arg Leu Gly				
1355		1360		1365
Asp Pro Ile Glu Ala Glu Ala Leu Leu Gly Ser Tyr Gly Arg Gly				
1370		1375		1380
Arg Val Gly Gly Ala Leu Leu Leu Gly Ser Val Lys Ser Asn Ile				
1385		1390		1395
Gly His Thr Gln Ala Ala Ala Gly Val Ala Gly Val Ile Lys Met				
1400		1405		1410
Val Met Ala Leu Arg Ala Gly Val Val Pro Ala Thr Leu His Val				
1415		1420		1425
Asp Val Pro Ser Pro Leu Val Asp Trp Ser Ser Gly Gly Val Glu				

1430	1435	1440
Leu Val Thr Glu Ala Arg Asp	Trp Pro Val Val Gly	Arg Val Arg
1445	1450	1455
Arg Ala Gly Val Ser Ala Phe	Gly Val Ser Gly Thr	Asn Ala His
1460	1465	1470
Leu Ile Leu Glu Gln Ala Pro	Glu Phe Asp Asp Pro	Val Val Thr
1475	1480	1485
Asp Thr Asp Thr Asp Ala Gly	Val Gly Arg Gly Leu	Ser Val Val
1490	1495	1500
Pro Val Val Val Ser Gly Arg	Ser Thr Ala Ala Leu	Arg Ala Tyr
1505	1510	1515
Ala Gly Arg Leu Arg Glu Val	Cys Ala Gly Leu Ser	Asp Gly Ala
1520	1525	1530
Gly Leu Val Asn Val Gly Trp	Ser Leu Val Ser Ser	Arg Ser Val
1535	1540	1545
Phe Glu His Arg Ala Val Val	Phe Gly Gly Gly Val	Ala Glu Val
1550	1555	1560
Val Ala Gly Leu Asp Ala Val	Val Ser Gly Ala Val	Ala Ser Gly
1565	1570	1575
Ser Val Val Val Gly Ser Val	Ala Ser Gly Val Ala	Gly Gly Gly
1580	1585	1590
Gly Arg Val Val Phe Val Phe	Pro Gly Gln Gly Trp	Gln Trp Val
1595	1600	1605
Gly Met Gly Ala Ala Leu Leu	Asp Glu Ser Glu Val	Phe Ala Glu
1610	1615	1620
Ser Met Val Glu Cys Gly Arg	Ala Leu Ser Gly Phe	Val Asp Trp
1625	1630	1635
Asp Leu Leu Glu Val Val Arg	Gly Gly Ala Gly Glu	Gly Val Trp
1640	1645	1650
Gly Arg Val Asp Val Val Gln	Pro Val Ser Trp Ala	Val Met Val
1655	1660	1665
Ser Leu Ala Arg Leu Trp Met	Ser Val Gly Val Val	Pro Asp Ala
1670	1675	1680
Val Val Gly His Ser Gln Gly	Glu Val Ala Ala Ala	Val Val Gly
1685	1690	1695
Gly Val Leu Ser Val Ala Asp	Gly Ala Arg Val Val	Ala Leu Arg
1700	1705	1710
Ser Arg Val Ile Gly Glu Val	Leu Ala Gly Gly Gly	Ala Met Val

1715	1720	1725
Ser Val Gly Leu Pro Ile 1730	Val Asp Ala Gln Glu 1735	Arg Leu Ala Gly 1740
Trp Gly Gly Arg Leu Gly 1745	Val Ala Ala Val Asn 1750	Gly Pro Ser Leu 1755
Thr Val Val Ser Gly Asp 1760	Val Asp Ala Ala Val 1765	Gly Phe Val Gly 1770
Glu Cys Glu Arg Asp Gly 1775	Val Trp Val Arg Arg 1780	Val Ala Val Asp 1785
Tyr Ala Ser His Ser Ala 1790	His Val Glu Ala Val 1795	Glu Gly Met Leu 1800
Ser Gly Leu Leu Gly Gly 1805	Leu Cys Pro Gly Arg 1810	Gly Val Val Pro 1815
Phe Tyr Ser Ser Val Val 1820	Gly Gly Val Val Asp 1825	Gly Val Gly Leu 1830
Asp Gly Gly Tyr Trp Tyr 1835	Arg Asn Leu Arg Glu 1840	Arg Val Leu Phe 1845
Ser Asp Val Val Gly Arg 1850	Leu Val Gly Asp Gly 1855	Phe Ser Gly Phe 1860
Val Glu Cys Ser Gly His 1865	Pro Val Leu Ala Gly 1870	Gly Val Leu Glu 1875
Ser Val Ala Val Val Asp 1880	Pro Asp Val Arg Pro 1885	Val Val Val Gly 1890
Ser Leu Arg Arg Asp Asp 1895	Gly Gly Trp Gly Arg 1900	Phe Leu Thr Ser 1905
Val Gly Glu Ala Phe Val 1910	Gly Gly Met Ser Val 1915	Asp Trp Lys Gly 1920
Val Phe Ala Gly Ala Gly 1925	Ala Arg Leu Val Asp 1930	Leu Pro Thr Tyr 1935
Pro Phe Gln Arg Arg His 1940	Tyr Trp Ala Pro Thr 1945	Pro Thr Asn Pro 1950
Ala Thr Asn Pro Ala Thr 1955	Gly Asp Thr Thr Thr 1960	Ala Asp Pro Val 1965
Gly Gly Val Arg Tyr Arg 1970	Ile Thr Trp Lys Pro 1975	Leu Pro Thr Asp 1980
Asp Pro Arg Pro Leu Thr 1985	Asn Arg Trp Leu Leu 1990	Ile Ala Asp Pro 1995
Gly Thr Ala Gly Ser Glu Leu	Ala Ala Asp Ile Thr	Ala Ala Leu

2000	2005	2010
Ile Arg Arg Gly Ala Glu Val 2015	Glu Leu Leu Ala Val 2020	Asp Pro Leu 2025
Ala Gly Arg Ala Arg Ile Ala 2030	Glu Leu Leu Ala Thr 2035	Thr Thr Thr Ala 2040
Gly Pro Val Pro Leu Ser 2045	Gly Ala Val Ser Leu 2050	Leu Gly Leu Val 2055
Gln Asp Ala His Pro Gln His 2060	Pro Ser Ile Gly Met 2065	Gly Val Val 2070
Ser Ser Leu Ala Leu Val Gln 2075	Ala Ile Gly Asp Ala 2080	Gly Ala Glu 2085
Thr Pro Leu Trp Ser Val Thr 2090	Gln Gly Ala Val Ala 2095	Val Val Pro 2100
Gln Glu Ala Pro Asp Val Phe 2105	Gly Ala Gln Val Trp 2110	Ala Phe Gly 2115
Arg Val Ala Ala Leu Glu Leu 2120	Pro Asp Arg Trp Gly 2125	Gly Leu Val 2130
Asp Leu Pro Ser Val Pro Asn 2135	Ala Arg Met Leu Asp 2140	Gln Leu Ala 2145
Asn Ala Leu Ala Gly Ala Asp 2150	Gly Glu Asp Gln Ile 2155	Ala Val Arg 2160
Gly Ser Gly Ile Tyr Gly Arg 2165	Arg Val Thr Arg Ala 2170	Ala Gly Thr 2175
Ala Arg Arg Glu Trp Arg Pro 2180	Arg Gly Asn Ile Leu 2185	Val Thr Gly 2190
Gly Thr Gly Ser Leu Gly Gly 2195	Arg Val Ala Arg Trp 2200	Leu Ala Arg 2205
Asn Gly Ala Glu His Leu Val 2210	Leu Thr Ser Arg Arg 2215	Gly Ala Asp 2220
Ala Pro Gly Ala Ala Glu Leu 2225	Glu Ala Asp Leu Arg 2230	Ala Leu Gly 2235
Val Glu Val Thr Met Ala Ala 2240	Cys Asp Val Ala Asp 2245	Arg Ala Ala 2250
Leu Ser Asp Val Leu Ala Ala 2255	His Pro Pro Thr Ala 2260	Val Phe His 2265
Thr Ala Gly Val Leu His Asp 2270	Gly Val Ile Asp Thr 2275	Leu Ala Ala 2280
Gly His Ile Asp Glu Val Phe 2285	Arg Pro Lys Thr Ala 2290	Ala Ala Leu 2295

2285		2290		2295
Leu Leu Asp Glu Leu Thr Gln His Gln Glu Leu Asp Ala Phe Val				
2300		2305		2310
Leu Phe Ser Ser Val Thr Gly Val Trp Gly Asn Gly Gly Gln Ala				
2315		2320		2325
Ala Tyr Ala Ala Ala Asn Ala Ser Leu Asp Ala Leu Ala Glu Arg				
2330		2335		2340
Arg Arg Ala Ala Gly Leu Pro Ala Thr Ser Ile Ala Trp Gly Leu				
2345		2350		2355
Trp Gly Gly Gly Gly Met Ala Glu Gly Ile Gly Glu Gln Asn Leu				
2360		2365		2370
Asn Arg Arg Gly Ile Thr Ala Leu Asp Pro Glu Leu Gly Ile Ala				
2375		2380		2385
Ala Leu Gln Gln Ala Leu Asp Arg Asp Asp Val Ser Val Thr Val				
2390		2395		2400
Ala Asp Val Asp Trp Thr Val Phe Ala Pro Arg Leu Ala Asp Leu				
2405		2410		2415
Arg Ser Gly Arg Leu Phe Asp Gly Val Pro Glu Ala Arg Ser Ala				
2420		2425		2430
Leu Asp Ala Arg Lys Val Asp Thr Glu Ser Pro Ser Ala Gly Leu				
2435		2440		2445
Ala Gln Arg Val Ala Gly Met Pro Asp Ala Glu Arg Gln Arg Val				
2450		2455		2460
Leu Leu Glu Thr Val Arg Ala Ala Ala Ala Ala Val Leu Arg His				
2465		2470		2475
Glu Thr Val Asp Ala Val Ala Pro Thr Arg Ala Phe Lys Asp Ala				
2480		2485		2490
Gly Phe Asp Ser Leu Thr Ala Leu Glu Leu Arg Asn His Leu Asn				
2495		2500		2505
Ser Thr Thr Gly Leu Ser Leu Pro Pro Thr Val Val Phe Asp His				
2510		2515		2520
Pro Thr Pro Ser Thr Leu Ala Lys Phe Leu Glu Gly Val Leu Val				
2525		2530		2535
Gly Ala Ser Ala Glu Glu Val Pro Val Thr Ala Ala Ala Val Pro				
2540		2545		2550
Val Asp Glu Pro Ile Ala Ile Val Gly Met Ala Cys Arg Tyr Pro				
2555		2560		2565
Gly Gly Ala Asp Thr Pro Glu Lys Leu Trp Asp Leu Leu Leu Ala				

2570	2575	2580
Gly Ala Asp Val Ile Gly Pro Ala Pro Asp Asp Arg Gly Trp Asp		
2585	2590	2595
Val Asp Ser Phe Phe Asp Pro Val Pro Gly Ala Ala Gly Lys Ser		
2600	2605	2610
Tyr Ala Arg Glu Gly Gly Phe Val Tyr Asp Ala Gly Met Phe Asp		
2615	2620	2625
Ala Glu Phe Phe Gly Val Ser Pro Arg Glu Ala Val Ala Met Asp		
2630	2635	2640
Pro Gln Gln Arg Leu Leu Leu Glu Thr Ser Trp Glu Ala Leu Glu		
2645	2650	2655
Arg Ala Gly Ile Asp Pro Ala Gly Leu Arg Gly Ser Arg Thr Gly		
2660	2665	2670
Val Tyr Ser Gly Leu Thr His Gln Glu Tyr Ala Ala Arg Leu His		
2675	2680	2685
Glu Ala Pro Gln Glu Leu Glu Gly Tyr Leu Leu Thr Gly Lys Ser		
2690	2695	2700
Val Ser Val Ala Ser Gly Arg Val Ser Tyr Val Leu Gly Leu Glu		
2705	2710	2715
Gly Pro Ser Ile Ser Val Asp Thr Ala Cys Ser Ser Ser Leu Val		
2720	2725	2730
Ala Leu His Leu Ala Cys Gln Gly Leu Arg Leu Gly Glu Cys Asp		
2735	2740	2745
Val Ala Leu Ala Gly Gly Val Thr Val Ile Ala Ala Pro Gly Leu		
2750	2755	2760
Phe Val Glu Phe Ser Arg Gln Gly Gly Leu Ser Gly Asp Gly Arg		
2765	2770	2775
Cys Arg Ala Phe Ala Gly Gly Ala Asp Gly Thr Gly Trp Gly Glu		
2780	2785	2790
Gly Ala Gly Val Val Val Leu Glu Arg Leu Ser Val Ala Arg Glu		
2795	2800	2805
Arg Gly His Arg Val Leu Ala Val Val Arg Gly Ser Ala Val Asn		
2810	2815	2820
Gln Asp Gly Gly Ser Asn Gly Leu Thr Ala Pro Ser Gly Val Ala		
2825	2830	2835
Gln Arg Arg Val Ile Gly Ala Ala Leu Val Ala Ala Gly Leu Gly		
2840	2845	2850
Val Ser Asp Val Asp Val Val Glu Ala His Gly Thr Gly Thr Arg		

2855	2860	2865
Leu Gly Asp Pro Ile Glu Ala	Glu Ala Leu Leu Gly	Ser Tyr Gly
2870	2875	2880
Arg Gly Arg Val Gly Gly Ala	Leu Leu Leu Gly	Ser Val Lys Ser
2885	2890	2895
Asn Ile Gly His Thr Gln Ala	Ala Ala Gly Val	Ala Gly Val Ile
2900	2905	2910
Lys Met Val Met Ala Leu Arg	Ala Gly Val Val	Pro Ala Thr Leu
2915	2920	2925
His Val Asp Val Pro Ser Pro	Leu Val Asp Trp	Ser Ser Gly Gly
2930	2935	2940
Val Glu Leu Val Thr Glu Ala	Arg Asp Trp Pro	Val Val Gly Arg
2945	2950	2955
Val Arg Arg Ala Gly Val Ser	Ala Phe Gly Val	Ser Gly Thr Asn
2960	2965	2970
Ala His Leu Ile Leu Glu Gln	Ala Pro Glu Phe	Asp Asp Pro Ala
2975	2980	2985
Asp Ser Asp Ser Asp Ser Asp	Ser Asp Ser Asp	Ala Gly Val Val
2990	2995	3000
Asp Gly Gly Glu Gly Gly Val	Gly Arg Ser Leu	Ser Val Val Pro
3005	3010	3015
Val Val Val Ser Gly Arg Ser	Val Gly Ala Leu	Arg Ala Tyr Ala
3020	3025	3030
Gly Arg Leu Arg Glu Val Cys	Ala Gly Leu Ser	Asp Gly Gly Gly
3035	3040	3045
Ser Gly Gly Gly Ser Gly Leu	Val Asp Val Gly	Trp Ser Leu Val
3050	3055	3060
Ser Ser Arg Ser Val Phe Glu	His Arg Ala Val	Val Phe Gly Gly
3065	3070	3075
Gly Val Glu Glu Val Val Ala	Gly Leu Gly Ala	Val Ala Ser Gly
3080	3085	3090
Ala Val Ala Ser Gly Ser Val	Val Val Gly Ser	Val Ala Ser Gly
3095	3100	3105
Val Ala Gly Gly Gly Gly Arg	Val Val Phe Val	Phe Pro Gly Gln
3110	3115	3120
Gly Trp Gln Trp Val Gly Met	Gly Ala Ala Leu	Leu Asp Glu Ser
3125	3130	3135
Glu Val Phe Ala Glu Ser Met	Val Glu Cys Gly	Arg Ala Leu Ser

3140		3145		3150
Gly Phe Val Asp Trp Asp Leu Leu Glu Val Val Arg Gly Gly Ala				
3155		3160		3165
Gly Glu Gly Val Trp Gly Arg Val Asp Val Val Gln Pro Val Ser				
3170		3175		3180
Trp Ala Val Met Val Ser Leu Ala Arg Leu Trp Met Ser Val Gly				
3185		3190		3195
Val Val Pro Asp Ala Val Val Gly His Ser Gln Gly Glu Val Ala				
3200		3205		3210
Ala Ala Val Val Gly Gly Val Leu Ser Val Ala Asp Gly Ala Arg				
3215		3220		3225
Val Val Ala Leu Arg Ser Arg Val Ile Gly Glu Val Leu Ala Gly				
3230		3235		3240
Gly Gly Ala Met Val Ser Val Gly Leu Pro Ile Val Asp Val Gln				
3245		3250		3255
Glu Arg Leu Ala Gly Trp Gly Gly Arg Leu Gly Val Ala Ala Val				
3260		3265		3270
Asn Gly Pro Ser Leu Thr Val Val Ser Gly Asp Val Asp Ala Ala				
3275		3280		3285
Val Gly Phe Val Gly Glu Cys Glu Arg Asp Gly Val Trp Val Arg				
3290		3295		3300
Arg Val Ala Val Asp Tyr Ala Ser His Ser Ala His Val Glu Ala				
3305		3310		3315
Val Glu Gly Met Leu Ser Gly Leu Leu Gly Gly Leu Cys Pro Gly				
3320		3325		3330
Arg Gly Val Val Pro Phe Tyr Ser Ser Val Val Gly Gly Val Val				
3335		3340		3345
Asp Gly Val Gly Leu Asp Gly Gly Tyr Trp Tyr Arg Asn Leu Arg				
3350		3355		3360
Glu Arg Val Leu Phe Ser Asp Val Val Gly Arg Leu Val Gly Asp				
3365		3370		3375
Gly Phe Ser Gly Phe Val Glu Cys Ser Gly His Pro Val Leu Ala				
3380		3385		3390
Gly Gly Val Leu Glu Ser Val Ala Val Val Asp Pro Asp Val Arg				
3395		3400		3405
Pro Val Val Val Gly Ser Leu Arg Arg Asp Asp Gly Gly Trp Gly				
3410		3415		3420
Arg Phe Leu Thr Ser Val Gly Glu Ala Phe Val Gly Gly Met Ser				

3425		3430		3435
Val Asp Trp Lys Gly Val	Phe Ala Gly Ala Gly	Ala Arg Leu Val		
3440	3445	3450		
Asp Leu Pro Thr Tyr Pro	Phe Gln Arg Arg His	Tyr Trp Ala Gln		
3455	3460	3465		
Thr Ser Pro Ala Gly Val	Gly Thr Ala Ala Ala	Ala Arg Phe Gly		
3470	3475	3480		
Met Glu Trp Glu Asp His	Pro Leu Leu Gly Gly	Ala Leu Ser Val		
3485	3490	3495		
Gly Gly Ser Arg Ser Leu	Leu Leu Ala Gly His	Leu Ser Leu Ala		
3500	3505	3510		
Ser His Ala Trp Leu Thr	Asp His Ala Val Ser	Gly Thr Val Leu		
3515	3520	3525		
Leu Pro Gly Thr Ala Phe	Val Glu Leu Ala Leu	His Ala Ala Ala		
3530	3535	3540		
Ala Ala Gly Cys Pro Glu	Val Glu Glu Leu Arg	Leu Glu Ala Pro		
3545	3550	3555		
Leu Val Val Pro Ala Arg	Gly Gly Val Arg Leu	Gln Val Leu Val		
3560	3565	3570		
Asp Asp Pro Asp Asp Gly	Ser Asp Arg Arg Ala	Val Ser Val Phe		
3575	3580	3585		
Ser Arg Asp Asp Ala Ala	Pro Ala Glu Ser Ala	Trp Thr Arg His		
3590	3595	3600		
Ala Val Gly Val Leu Ala	Ala Arg Ser Arg Pro	Ala Pro Ala Ala		
3605	3610	3615		
Pro Trp His Thr Asp Ala	Trp Pro Pro Ser Gly	Thr Glu Pro Val		
3620	3625	3630		
Asp Val Ala Asp Leu Tyr	Glu Arg Phe Ala Ala	Leu Gly Tyr Glu		
3635	3640	3645		
Tyr Gly Glu Ala Phe Ala	Gly Leu Gln Gly Val	Trp Arg Gly Asp		
3650	3655	3660		
Gly Glu Val Phe Ala Glu	Val Arg Leu Pro Asp	Arg Val Ser Ala		
3665	3670	3675		
Glu Ala Ile Arg Phe Gly	Leu His Pro Ala Leu	Leu Asp Ala Ala		
3680	3685	3690		
Leu Gln Gly Trp Leu Ala	Gly Asp Leu Val Gly	Val Pro Glu Gly		
3695	3700	3705		
Ser Val Leu Leu Pro Phe	Ala Trp Gln Gly Val	Val Leu His Ala		

3710	3715	3720
Thr Gly Ala Asp Thr Leu Arg Val Arg Ile Gly Arg Ser Gly Asp 3725 3730 3735		
Ser Ala Val Cys Leu His Ala Val Asp Pro Ala Gly Ala Pro Val 3740 3745 3750		
Leu Ser Leu Asp Ala Leu Ala Leu Arg Pro Leu Val Arg Glu Arg 3755 3760 3765		
Leu Gly Leu Pro Ala Asp Ala Gly Ala Gly Ala Leu Tyr Arg Val 3770 3775 3780		
Gly Trp Arg Arg Gln Ala Ala Val Ala Gly Ala Ala Asp Arg Arg 3785 3790 3795		
Trp Ala Val Val Ala Pro Asn Gly Ala Glu Ala Asp Gly Ala Ala 3800 3805 3810		
Glu Pro His Arg Trp Pro Val Ala Ala Val Asp Val His Thr Asp 3815 3820 3825		
Val Asp Ser Leu Arg Ala Ala Leu Asp Ala Gly Ala Glu Leu Pro 3830 3835 3840		
Ala Val Val Leu Ala Asp Phe Arg Arg Ala Ala Gly Trp Ser Val 3845 3850 3855		
Asp Ser Ser Leu Ala Ala Gly Pro Ser Pro Asn Asp Gly Ala Val 3860 3865 3870		
Gly Asp Gly Ala Val Gly Asp Ala Arg Ala Gly Ala Val Arg Ala 3875 3880 3885		
Ala Thr Arg Ala Gly Leu Asp Leu Leu Gln Arg Trp Leu Ala Asp 3890 3895 3900		
Glu Arg Phe Ile Ala Ala Arg Leu Val Val Val Thr Glu Arg Ala 3905 3910 3915		
Val Ala Ala Gly Pro Asp Glu Asp Val Pro Gly Leu Val His Ala 3920 3925 3930		
Gly Leu Trp Gly Leu Leu Arg Ser Ala Gln Ser Glu His Pro Asp 3935 3940 3945		
Arg Phe Val Leu Val Asp Val Asp Ala Asp Asp Ser Ser Leu Ala 3950 3955 3960		
Ala Leu Pro Ser Ala Leu Ala Met Asp Ala Pro Gln Leu Val Val 3965 3970 3975		
Arg Ala Gly Gln Ile Leu Leu Pro Glu Ile Glu Pro Val Arg Pro 3980 3985 3990		
Val Pro Glu Pro Glu Gln Ala Glu Pro Glu Pro Gly Ala Val Leu		

3995	4000	4005
Asp Pro Asp Gly Thr Val 4010	Leu Leu Thr Gly Ala Thr 4015	Gly Thr Leu 4020
Gly Gly Leu Leu Ala Arg 4025	His Leu Val Thr Thr 4030	Arg Gly Ala Arg 4035
Arg Leu Leu Leu Val Ser 4040	Arg Ser Gly Pro Asp 4045	Ala Pro Asp Ala 4050
Gly Arg Leu Thr Glu Glu 4055	Leu Thr Gly Leu Gly 4060	Ala His Val Thr 4065
Leu Ala Ala Cys Asp Thr 4070	Thr Asp Arg Ala Ala 4075	Leu Ala Gly Val 4080
Leu Gly Gly Ile Pro Ala 4085	Glu His Pro Leu Thr 4090	Ala Val Val His 4095
Val Ala Gly Val Leu Asp 4100	Asp Gly Ala Val Gln 4105	Ala Leu Thr Pro 4110
Glu Arg Val Asp Ala Val 4115	Leu Arg Pro Lys Val 4120	Asp Ala Ala Leu 4125
His Leu His Glu Leu Thr 4130	Ala Gly Leu Pro Leu 4135	Ala Ala Phe Val 4140
Leu Phe Ser Gly Ala Ala 4145	Gly Ile Leu Gly Arg 4150	Pro Gly Gln Ala 4155
Asn Tyr Ala Ala Ala Asn 4160	Thr Phe Leu Asp Ala 4165	Leu Ala Gln His 4170
Arg Arg Ala Arg Gly Leu 4175	Pro Gly Val Ser Leu 4180	Ala Trp Gly Leu 4185
Trp Gly Leu Ala Ser Asp 4190	Met Thr Gly His Leu 4195	Gly Glu Gln Asp 4200
Leu Arg Arg Met Arg Arg 4205	Ser Gly Ile Ala Pro 4210	Met Thr Gly Glu 4215
Glu Gly Leu Ala Leu Phe 4220	Asp Leu Ala Leu Asp 4225	Leu Ala Arg Asp 4230
Glu Pro Val Leu Val Pro 4235	Ala Arg Leu Asp Pro 4240	Ala Ala Leu Arg 4245
Arg Glu Trp Ala Ala Asn 4250	Gly Pro Gly Ala Val 4255	Pro Val Leu Leu 4260
Arg Gly Leu Val Pro Ala 4265	Ala Pro Leu Arg Arg 4270	Ala Ala Pro Ser 4275
Gly Ala Ala Gly Gly Ala 4280	Pro Val Pro Ala Val 4285	Ala Ala Pro Gln 4290

4280		4285		4290
Gln Ala Asp Glu Leu Arg Gly	Gln Leu Ala Gly Lys Asp Ala Gln			
4295	4300	4305		
Ala Gln Val Arg Gln Leu Leu	Asp Leu Val Arg Ala His Val Ala			
4310	4315	4320		
Gly Val Leu Ala Leu Arg Glu	Ala Ala Asp Val Asp Pro Gly Arg			
4325	4330	4335		
Pro Phe Arg Glu Val Gly Phe	Asp Ser Leu Thr Ala Val Glu Leu			
4340	4345	4350		
Arg Asn Arg Leu Gly Ser Ala	Thr Gly Leu Arg Leu Ala Pro Ser			
4355	4360	4365		
Leu Val Phe Asp His Pro Thr	Pro Ser Ala Val Ala Glu His Leu			
4370	4375	4380		
Val Asp Arg Leu Ala Ala Glu	Gly Ala Ala Asp Glu Gly Ala Ala			
4385	4390	4395		
Ala Leu Thr Gly Leu Asp Ala	Val Ala Ala Ala Leu Gly Gly Met			
4400	4405	4410		
Arg Thr Asp Asp Val Arg Arg	Asp Ile Val Arg Arg Arg Leu Glu			
4415	4420	4425		
Glu Met Leu Ala Leu Val Gly	Gly Pro Arg Ser Gly Pro Ala Gly			
4430	4435	4440		
Asp Gly Leu Val Asp Ala Thr	Val Ala Glu Arg Leu Asp Ser Ala			
4445	4450	4455		
Ser Asp Asp Glu Leu Phe Ala	Leu Ile Glu Glu Gln Leu			
4460	4465	4470		

<210> 11

<211> 13416

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 11

atgcgagttg tgggcgcaga cgcgtgcagc gcagccgtcc ccgccggacc gcggatgggc	60
ttcccagcat cgttcttcga cccaggagac ctcatgaccg tgcagagtga cgtgttgccg	120
caccgcgata tcgccgtcat cgggatgtcc tgccggcttc ccggcgcgcc gagcatcgag	180
gaattctggg acctgctgtg cagcgggcgg agcgcggtcg accgccagcc cgacggcggt	240
tggcggggcg tgatcgatgg gaagggagaa tccgacgccg cgttcttcgg catgtccccg	300
cgccaggccg ccgcggtcga cccgcaacag cgcctgatgc tcgaactcgg ctgggaggca	360
ctggagaacg cccgcatccg gcccgccgac ctgaagggt cgcgactgg cgtcttcgtg	420

gggctcaccg ccgacgacta cgccacottg ctgcccgcgt ccggcacgcc catcagcggg 480
 cacaccgcga caggcctgaa ccgtagcctc acggccaacc gtctctcgta cctgctgggt 540
 ctgcccggcc ccagcttcac cgtggactcc gcgcagtcgt catccctggg ccgcgttcac 600
 ctggcgtgcg aaagcctgct gcggggcgag agcgcggtcg ccgtcgtcgg cggggtgagc 660
 ctcatcctgg cagaggagag caccgccgcc atggcgcgta tgggggcact ctctcctgac 720
 gggcgttgct tcaccttcga cggccgggcc aacggctacg tccgtggcga ggggtggcgtg 780
 gccatgggtcc tcaagccgct gatccgcgcg atcgaggacg gcgaccaggt gcaactgcgtc 840
 atccggggct gtgccgtcaa caacgacggc ggtggcccca gcctcaccca tcccgaccgg 900
 gaggcccagg aggcattgct gcgccgggcg tacgagcggg cgggggtggc ccccgaaacac 960
 gtcgactacg tcgagctgca cggcaccggg acgaaggccg gcgaccccg ctaggcggcg 1020
 gccctcgggg cgggtgctggg tgtcgccgcg ggctgcgaca acccactcgc ggtcggatcg 1080
 gtcaagacca acgtcggcca cctggagggg gcggccggca tcacgggcct gctgaaggcg 1140
 gtgctgtgcg tacgtgaggg ggtgctgccc ccgagcctca acttccgtac gccgaacccg 1200
 gacatccgcc tcgacgagct gaacctccgg gttcagacgg aactgcagcc gtggccgggc 1260
 gacgggacgg gccgccgcg tgtcgccgga gtgagttcct tcggcatggg cggtagaat 1320
 gcgcactga ttctcgagca ggctccgggtg gcggctgagg aaacggctgt taccgatgcc 1380
 ggtgtcggtt cgggttcgggt ggttcgggtg gtggtgtcgg gtcgttcggg gggggctttg 1440
 cgggcgtatg cgggtcgggt gcgtgaggtg tgcgcggggg tgtctgacgg tgggtggtcc 1500
 ggtggtggtt ctggtctggt ggatgtgggt tggctggttg tgtcgtcgcg gtcggtgttc 1560
 gagcatcggg cggctcgtgt cgggtggggg gtgcgcgagg tgggtggcggg tttggatgcg 1620
 gtggcttctg gggcggtgag ttcgggttcg gtggtggtgg gttcgggtgg gtcgggtgtt 1680
 gctggtggtg gtggtcgggt ggtgtttgtg tttccgggtc agggttggca gtgggtgggt 1740
 atgggtgccc ctctgttgga cgagtcggag gtgtttgctg agtcgatggg ggagtgtggg 1800
 cgggcgttgt cggggtttgt ggattgggat ttgttggaag tggccgcgg tgggtggggg 1860
 gacggatcgt ttggtcgggt tgatgtggtg cagccgggtg cgtgggcggg gatggtgtcg 1920
 ttggcgcggt tgtggatgtc ggtgggtgtg gtgccggatg cgggtggtggg tcattcgag 1980
 ggtgaggttg ctgcgccggt ggtggggggg gtgttgagtg tggctgatgg ggcgcgggtg 2040
 gtggcgttgc ggtcgcgggt gatcgggtgag gtgttgccgg gtggtggtgc gatggtgtcg 2100

gtgggggttgc cgggtggcgggt tgtgttggat cggttggcgg ggtgggggtgg tcggttgggt	2160
gtggcggcgg tgaatgggcc gtctgtgacg gtggtgtcgg gggatgtgga tgctgctgtg	2220
gggtttgttg gtgagtgtga gcgggatggg gtgtgggtgc ggcgggtggc ggtggattat	2280
gcgtcgatt cggcgcatgt ggaggcgggt gaggggatgc tgcgggggtt gttgggtgg	2340
ttgtgtccgg ggcgggggtgt ggtgccgttt tattegtcgg tgggtgggtgg tgtggtgat	2400
ggggtgggtt tggatgggtg gtattggtat cggaatctgc gtgagcgggt gttgttttcg	2460
gatgtggtgg ggcggcttgt tgggatggg ttttcgggggt ttgtggagtg ttcggggcat	2520
ccggtgttgg cgggtgggggt gttggagtgc gtggcgggtg tggatccgga tgcgcggccg	2580
gtggtggtgg ggtcgtcgc cgtgatgat ggtgggtggg gccggttttt gacgtcgggtg	2640
ggtgaggcgt tcgtcggcgg gatgagtgtt gactggaagg gtgtgttcgc gggggcgggc	2700
gcgcggttgg ttgacctgcc gacgtatccg ttccaacgac gccactactg ggcaccgaac	2760
accgacggcg cgccagctcc gatcctcgat gatcacgcgg aggcggagaa cgaaccagcc	2820
gaatccgagc cagggattcg ggccgagctt ctgacgttgg ccgagcccga gcaactgaac	2880
cgactcttgg cgaccgttcg cgccagcacc gccgtcgttc tgggcctcga ctcggcgcag	2940
gcggtcgate cggagcgcac gttcaaggag catggattcg aatcggtcac cgccgtcgag	3000
ctctgtaacc acctgcaacg cggcactggg ctgcgggttc ccgcctcgtt tgtatacaac	3060
catcccaccc cgatggccgc tgcccgggaag ctgcaggaag aaattcaggg ccggcaaccg	3120
gagaacgtcc ggcaggtcac ctccgtcgtt gctgtggatg atccggtggg ggtggtgggg	3180
atgggttgtc gttttccggg tggggtggtg tgtgcggagg gtttgtggga tttggtgttg	3240
gggggtgggg atgcggtgtc ggggtttccg gtggatcggg gttgggatgt ggaggggttg	3300
tttgatccgg tgcgggggtgt ggtggggaag tcgtatgtgc gggagggggg gtttgtgtat	3360
gacgcgggga tgttcgatgc ggagtttttt ggtgtgtcgc cgcgtgaggc ggtggcgatg	3420
gatccgcagc agcgtttgtt tttggagggt tcgtgggagg cgttgagcgc tgcggggatt	3480
gatccgttgg gtttgcgggg ttccgggacg ggtgtgtatg tgggggtgat gggtcaggag	3540
tatgggccgc ggttgggtga gtcgggtggg gggtttgagg gttatttgtt gacggggacg	3600
tcgccgagtg tgggtgcggg tcgtgtttcg tatgtgttgg ggttggaggg tccgtcgatt	3660
tcggttgata cggcgtgttc gtctcgttg gtggcgttc atttggcgtg tcaggggttg	3720
cggttgggtg agtgtgatgt ggcgttggcg ggtgggtgta cggtgattgc ggcgccgggg	3780
ttgtttgtgg agttttctcg gcagggtggg ttgtcgggtg atgggcgggtg tcgggcgttt	3840

gcgggtggtg cggatgggac ggggtggggg gaggtgcg ggggtggtggt gttggagcgg 3900
 ttgtcgggtg cgcgggagcg tggtcacggt gtgttgccgg tggcgcgggg ttctgcgggtg 3960
 aatcaggatg gtgggtcgaa tggtttgacg gcgccgtcgg ggggtggcgca gcgtcgggtg 4020
 attggtgctg cgttggtggc ggcgggtttg ggtgtgtcgg atgtggatgt ggtggagcgg 4080
 catgggacgg ggactcgggt ggggtgatccg attgaggtcg aggcgttggt ggggtcgtat 4140
 gggcggggtc gtgtgggtgg ggcgttggtg ttgggttcgg tgaagtcgaa tattggtcat 4200
 acgcaggcgg ctgcgggtgt ggcgggtgtg atcaagatgg tgatggcggt gcgggcgggg 4260
 gtggtgccgg cgacgttgca tgtggatgtg ccgtcgccgt tgggtggattg gtcttcgggt 4320
 ggggtggagt tgggtgacgga ggcgcgggat tggccgggtg tgggtcgtgt gcgtcgtgctg 4380
 ggtgtgtcgg cgtttggggg gtcggggacg aatgcgcacg tgattttgga gcaggcccc 4440
 gaattcgacg atccggttgt taccgacacc gacaccgatg ctggtgtggg taggggtcta 4500
 tcggtggttc cgggtggtgt ttccgggtcgt tcgacggcgg ctttgccgct ttatgcgggc 4560
 cggttgctgt aggtgtgctg gggcttttcc gatggtgccg gtctggtgaa tgtgggttgg 4620
 tcgttggtgt cgtcgcggtc ggtgttcgag catcgggcgg tcgtgtttgg tgggggtgtc 4680
 gccgaggtgg tggcgggttt ggatgcgggt gtttccgggg cggtaggttc ggggttcgggtg 4740
 gtggtgggtt cggtagcgctc ggggtgtgct ggtggtggtg gtcgggtggt gtttgtgttt 4800
 ccgggtcagg gttggcagtg ggtgggtatg ggtgcggcgc tgctggacga gtcggaggtg 4860
 tttgctgagt cgatggtgga gtgtggtcgg gcgttgctcg ggtttgtgga ttgggatttg 4920
 ttggaggtgg tcgggggtgg ggcgggtgag ggggtgtggg gtcgggttga tgtggtgcag 4980
 ccggtgtcgt gggcgggtgat ggtgtcgttg gcgcgggtgt ggatgtcggg ggggtgtggtg 5040
 ccggatgcgg tgggtgggtca ttccgaggtg gaggtgtctg cggcgggtgg ggggggtgtg 5100
 ttgagtgtgg ctgatggggc gcgggtggtg gcgttgcggt cgcgggtaat tggtaggtg 5160
 ttggccggtg gtggtgcgat ggtgtcggct ggactgccga tcgtggatgc gcaggaacgg 5220
 ttggcggggg ggggtggtcg gttgggtgtg gcggcggtga atggtccgtc gttgacgggtg 5280
 gtgtcggggg atgtggatgc tgctgtgggg tttgttggtg agtgtgagcg ggatgggggtg 5340
 tgggtgcggc ggggtggcgg ggtattatgc tcgcattcgg cgcattgtga ggcgggtggag 5400
 gggatgctgt cggggttggt ggggtggtttg tgtccggggc ggggtgtggt gccgttttat 5460
 tcgtcgggtg tgggtggtgt ggttgatggg gtgggtttgg atggtgggta ttggtatcgg 5520

aatctgcgtg agcgggtgtt gttttcggat gtggtggggc ggcttggttg ggatgggttt 5580
tcgggggtttg tggagtgttc ggggcatccg gtgttgccgg gtgggggtgtt ggagtcggtg 5640
gcggtggttg atccggatgt gcggccggtg gtggtggggc cgtcgcccg tgatgatggt 5700
gggtggggcc ggtttttgac gtcggtgggt gaggcgttcg tcggcgggat gagtgttgac 5760
tggaagggtg tgttcgcggg ggcgggcgcg cggttggttg acctgccgac gtatccgttc 5820
caacgccgcc actactgggc accgactccc accaaccgcc ccaccaacc cgccacgggc 5880
gacaccacca ccgccgaccc ggtgggtggc gtgcggtatc ggatcacctg gaaaccgttg 5940
ccgacggacg acccccgacc cctcaccaac cgtcggttac tcatcgccga cccggggacc 6000
gccggctcgg agcttgccgc agacatcaca gcagcgtca ttgcagggg cgccgaggtc 6060
gagttgctgg ccgtggaccc gctcgccggc cgggcccgga tcgccgaact gctcgccacc 6120
acgacggctg ggccggtgcc gctgtcgggc gccgtgtctc ttctcgggct tgtgcaggac 6180
gcgcatactc aacaccctc catcggaatg ggcgtggtct cgtcgctggc gctggtgcag 6240
gccatcggtg acgcgggagc cgagactcct ttgtggagcg tcacgcaggg ggcggtcgct 6300
gtggtgcccc aggaggcgcc ggatgtgttc ggtgcgcagg tgtgggcgtt cgggcgggtg 6360
gccgccctgg aactgccgga ccgctggggc ggcctggtcg accttcctc cgtaccgaat 6420
gcccggatgc tggaccagct cgccaacgcc ctgcgccgag cggacggcga ggaccagatc 6480
gcggtacgcg gctcggggat ctacgggcgt cgggtgacgc gcgcggcggg cactgcgcgc 6540
cgggaatggc gccctcgcgg gaacatcctg gtgaccggag gtacgggaag tctgggtggc 6600
cgggtggccc ggtggctcgc tcgcaacggt gccgaacacc tcgttctcac cagtcgtcgg 6660
ggtgccgacg ccccgggggc ggcagaactg gaagctgac ttgcgcgct cgggtgtcag 6720
gtgaccatgg ccgcctgcga tgtagcggac cgggctgcgc tgtccgacgt cctggcgggc 6780
catccgcca ctgcggtctt ccacaccgcc ggagtctgc acgacggtgt gatcgacacg 6840
ctgcgcccg gacacatcga cgaggtcttc cgtccgaaga ccgctgccgc gctgctgctc 6900
gacgaactca cccagcacca ggagctggac gccttcgtcc tcttctcact ggttaccgga 6960
gtctggggca acggcgcca ggcggcgtag gcggcgcgga acgcacgct ggacgccctg 7020
gcggagcgac gtcgtgccgc aggtcttccc gccacctca tagcttgggg actgtggggc 7080
ggcgggtggca tggcggaggg gatcggcgag cagaacctga accgcggtgg catcacggcc 7140
ttggaccgg agctcgcat cgccgctctg cagcaggccc tcgaccgga tgacgtgtct 7200
gtcacctcg ccgacgtcga ctggacggtt ttcgctccgc gtcttgccga cctgcgctcg 7260

gggcggtctt tcgacggggt gcccgaggcc aggagcgcg tcgatgccc gaaagtggac 7320
 accgagtcgc cgagcgccgg ccttgccgag cgggtggcgg ggatgcccga cgcggaacgg 7380
 cagcgggtcc tcctcgaaac ggtgcggggc gggccgcgg cggtcctgag gcacgagacg 7440
 gtggatgcgg tcgcgcccac ccgggccttc aaggacgccg gcttcgactc gctcacggcg 7500
 ctggaactgc gcaaccacct caacagcacg accggtctga gtctgcctcc gacggtggtc 7560
 ttcgaccacc ccaccccgtc cacgttggcg aagttcctgg agggcgctct cgtcggcgct 7620
 tctgccgagg aagtcccggt gactgccgca gccgtgcccg tcgacgagcc tattgccatc 7680
 gtcggcatgg cctgccgcta ccccgccgga gccgacactc ccgagaagct ctgggacctc 7740
 ctgctggccg gtgctgacgt catcgcccca gcccgcgacg accggggctg ggacgtggac 7800
 tccttctttg atcccggtgc gggcgccgcg ggggaagtcgt atgcgcggga gggggggttt 7860
 gtgtatgacg cggggatggt cgatgcggag ttctttggtg tgcgcgcgg tgaggcggtg 7920
 gcgatggatc cgcagcagcg cttgttgttg gagacgtcgt gggaggcgtt ggagcgtgcg 7980
 ggaatcgatc cggcgggtct gcggggtagc cggaccggcg tgtactccgg cctgacccac 8040
 caggagtatg ccgcccgtct gcacgaggct ccgcaggaac tcgagggcta tctgctcacc 8100
 ggcaagtccg tgagcgtcgc gtcgggtcgt gtttcgtatg tgttgggggt ggagggtccg 8160
 tcgatttcgg ttgatacggc gtgttcgtcg tcgttgggtg cgttgcatctt ggcgtgtcag 8220
 gggttgcggg tgggtgagtg tgatgtggcg ttggcgggtg gggtgacggg gattgcggcg 8280
 ccgggggttg ttgtggagtt ttctcggcag ggtgggttg cgggtgatgg gcgggtgtcg 8340
 gcgtttgcgg gtggtgcgga tgggacgggg tggggggagg gtgcgggggt ggtggtgttg 8400
 gagcggttgt cgggtggcgc ggagcgtggt catcgggtgt tggcgggtgt gcgggggtct 8460
 gcgggtgaatc aggatgggtg gtcgaatggt ttgacggcgc cgtcgggggt ggcgcagcgt 8520
 cgggtgattg gtgcggcggt ggtggcggcg ggtttgggtg tgcgggatgt ggatgtggtg 8580
 gaggcgcagt ggacggggac tcggttgggt gatccgattg aggctgaggc gttgttgggg 8640
 tcgtatgggc ggggtcgtgt ggggtggggc ttgttgttgg gtgcggtgaa gtcgaaatatt 8700
 ggtcatacgc aggcggctgc ggggtgtggc ggtgtgatca agatgggtgat ggcgttgcgg 8760
 gcgggggtgg tgccggcgac gttgcatgtg gatgtgccgt cgcggttggg ggattggtct 8820
 tcgggtgggg tggagtgggt gacggaggcg cgggattggc cgggtgtggg tcgtgtgcgt 8880
 cgtgcgggtg tgcggcggtt tggggtgtcg gggacgaatg cgcactgat tttggagcag 8940

gcccccgagt tcgacgatcc tgccgattcc gattccgatt ccgattccga ttccgatgcc 9000
 ggtgtcgtgg atggcggcga ggtggtggtt ggcaggagct tgcgggtggt tccggtggtg 9060
 gtgtcgggtc gttcgggtggg ggctttgcgg gcgtatgcgg gtcgggttcg tgagggtgtgc 9120
 gcgggggttg ctgacgggtg tggtcccggt ggtggttctg gtttgggtga tgtgggttg 9180
 tcgttggtgt cgtcgcggtc ggtgtttgag catcgggcgg tcgtgttcgg tgggggtgtg 9240
 gaggaggttg ttgctggtct tgggtcgggtg gcttctgggg cgggtggcttc gggttcgggtg 9300
 gtggtgggtt cgggtggcgc ggtgttgct ggtggtggtg gtcgggtggt gtttgtgttt 9360
 ccgggtcagg gttggcagt ggtgggtatg ggtgcggcgc tgctggacga gtcggaggtg 9420
 ttcgccgagt cgatggtgga gtgtggtcgg gcgttgctcg gggttgtgga ttgggatttg 9480
 ttggaggttg tgcgcggcgg ggcgggtgag ggggtgtggg gtcgggttga tgtggtgcag 9540
 ccggtgtcgt gggcgggtgat ggtgtcgttg gcgcgggttg ggatgtcggg ggggtgtggtg 9600
 ccggatgcgg tgggtgggtca ttgcagggt gaggttgctg cggcgggtgg tgggggtgtg 9660
 ttgagtgtgg ctgatggggc gcgggtggtg gcgttgccgg cgcgggtgat cggtgaggtg 9720
 ttggccgggt gtggtgcgat ggtgtcggtc ggactgccga tcgtggatgt gcaggaacgg 9780
 ttggcggggg ggggtggtcg gttgggtgtg gcggcgggtga atggtccgtc gttgacgggtg 9840
 gtgtcggggg atgtggatgc tgcgtggtgg tttgttggtg agtgtgagcg ggatgggggtg 9900
 tgggtgcggc ggggtggcgg ggattatgcg tcgcattcgg cgcattgtga ggcgggtggag 9960
 gggatgctgt cggggttgtt ggggtggtttg tgtccggggc ggggtgtggt gccgttttat 10020
 tcgtcgggtg tgggtggtgt ggttgatggg gtgggttttg atggtgggtg ttggtatcgg 10080
 aatctgcgtg agcgggtggt gttttcggat gtggtggggc ggcttggttg ggatgggttt 10140
 tcgggggttg tggagtgttc ggggcattcc gtgttggcgg gtgggtgtt ggagtccgtg 10200
 gcgggtggtg atccggatgt gcggccgggt gtggtggggc cgtgcgccg tgatgatggt 10260
 ggggtggggc ggtttctgac gtcggtgggt gaggcgttcg tcggcgggat gagtgttgac 10320
 tggaaggggt tgttcgcggg ggcgggcggc cggttggttg acctgccgac gtatccgttc 10380
 caacgacgcc actactgggc ccagacctc cccgctggcg tcgggacggc cgcggcggcc 10440
 cggttcggca tggagtggga ggaccatccc ctgctcggcg gtgcgctgtc ggtcgggggc 10500
 tccaggagcc tgcttctggc cgggcattct tcgctgcct cgacgcctg gctgaccgac 10560
 catgccgtct ccggcacgt gctgctgcc ggtacggcct tcgtggaact cgcctgcac 10620
 gccgccgctg cggctggctg tccggaggtc gaggagctgc ggctggaggc tcccctggtg 10680

gtgccggcca ggggcggggt gcggctccag gtgctcgtgg acgaccccga cgacggatcc 10740
 gaccgccgcg cggtaagcgt gttctcccgg gacgatgcgg cggcgccga gtccgcctgg 10800
 acgcggcacg cgggtggcgt cctggccgcg cggtcgcggc ctgcaccggc tgcgccctgg 10860
 cacaccgacg cctggccacc ttccggcacg gagccggtcg acgtggccga cctgtatgag 10920
 cggttcgcgg cgctgggcta cgagtacggg gaggcgttcg ccgggctcca gggggtctgg 10980
 cggggggacg gcgaggtgtt gcgcgaggtg cggctgcccg accgggtcag cgcggaggcc 11040
 attcgcttcg ggctgcatcc cgcgtgctc gacgccgcc tgcaggggtg gttggcgggc 11100
 gacctcgtcg gcgtccccga gggcagtgtg ctgctgccct tcgcctggca gggcgtcgtg 11160
 ctccacgcca ccggcgccga cactctgcgg gttcgcacg gccgggtccg tgactcggcc 11220
 gtctgcctgc acgcggtgga ccgggccggg gctccgggtc tctcgttga cgcctcggcc 11280
 ctgcgtccgc tcgtccggga acgcctcggg ctgccgcgcg atgccggagc cggggcgttg 11340
 taccgggtcg gctggcggcg gcaggccgcc gttgccgggg cagccgaccg gcgggtggcg 11400
 gtcgtggccc cgaacggtgc cgaggcggac ggggccgcgc agccgcaccg gtggccggtc 11460
 gccgccgtcg acgtgcacac cgacgtggac tcgctgcggg cggccctgga cgcgggcgcg 11520
 gaactgcccg ccgtcgtcct cgcgcacttc cggagggccg ccggttgag cgtcgacagt 11580
 tcgctggccg ccggcccgtc gcccacgac ggcgcgggtg gcgacggcgc ggtgggcgac 11640
 gcccgggccg gggccgtccg ggcggcgacc cgggccgggc tggatctgct gcaacgctgg 11700
 ctggccgacg agcggttcat cgcggccagg ctgctgggtg tcaccgaacg ggccgtggcc 11760
 gccggggccg acgaggacgt gccgggcctc gtccacgcgg gactgtgggg cctgctccgg 11820
 tcggcccaat cggagcacc gccacgcttc gtgctgggtg acgtcgacgc ggacgacagc 11880
 tcgctcgcgg cgctgccgtc ggccctcgcc atggacgcgc cccaactggt ggtgcgggcc 11940
 ggtcagatcc tgctgccga gatcgagccg gtgcggcccg tacccgagcc ggagcaggcg 12000
 gaaccgaac cgggggcccgt cctggacccc gacggcacgg tcctgtcac cggcgcgacc 12060
 ggcacgctcg gcgggtgct cggccggcac ctggtgacca ccggtggtgc gcgccggctg 12120
 ctgctggtca gccgcagcgg tccggacgcc ccgatgccg gccggctgac cgaggagctg 12180
 accgggctcg gcgccacgt gacgtggcc gctcgcgaca ccacggatcg cgcgcgctg 12240
 gccggcgtcc tgggcggcat ccccgccgag catccgctga ccgctggt gcacgtggcc 12300
 ggcgtactcg acgacggggc ggtgcaggcg ctacccccg agcgggtoga cgcggtgctc 12360

cggcogaagg tggacgcggc actgcacctg cacgaactga ccgcggggct gccgctggcc 12420
 gcgttcgtgc tgttctccgg ggcggcgggg atcctggggc ggcccggcca ggccaactac 12480
 gcggcggcga acaccttcct ggacgccctg gcgcagcacc gacggggccc gggcctgccc 12540
 ggcgtctccc tcgcctgggg cctgtggggg ctggccagcg acatgacggg ccacctgggc 12600
 gagcaggacc tgcggcggat gcggcgctcc ggcatcgccc cgatgaccgg cgaggagggc 12660
 ctgcgcgtgt tcgacctggc cctcgacctg gcccgggacg aaccggtgct cgtaccggcc 12720
 cgactggacc cggcggcgct gcgccgggag tgggcgcgca acggaccggg cgccgtcccg 12780
 gtcctgctgc ggggtctggt gccggcggct ccgctccgtc gcgcggcccc gtcgggcgcc 12840
 gccggcgggtg cggccgtgcc cgccgtcgcc gcgccgcagc aggcggacga gctgcgcggg 12900
 caactggccg ggaaggacgc gcaggcccag gtccggcagc tgctggatct ggtacgcgcc 12960
 catgtcgccg ggggtgctgc cctccgggaa gcggcggacg tggaccggg cagaccgttc 13020
 gcgcaggctc gattcgactc gttgaccgca gtcgaactgc gcaaccggct gggctcggcg 13080
 accggcctgc gggtggcacc gagcctggtg ttcgaccatc cgaccccgtc ggccgtggcc 13140
 gagcacctcg tggaccgcct cgccgccgag ggggcggctg acgagggcgc ggccggcactg 13200
 accgggctcg acgcagtggc cgccggcgtc ggcgggatgc ggacggacga cgttcgccgg 13260
 gacatcgctc gcaggcggct ggaggagatg ctcgccctgg tcggcgggcc acggtccggg 13320
 ccggcaggtg acgggctggt ggatgccacg gtcgccgagc gactggactc ggcttcgcac 13380
 gacgaactct tcgccctgat cgaggagcag ctgtga 13416

<210> 12

<211> 1925

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 12

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

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

```

370              375              380
Ala Ala Ala Gly Val Ala Gly Val Ile Lys Thr Val Leu Ala Leu Arg
385              390              395              400
His Ala Arg Leu Pro Arg Thr Leu His Val Asp Arg Pro Ser Thr Arg
405              410              415
Val Asp Trp Ser Ser Gly Ala Val Arg Leu Leu Thr Glu Gly Arg Pro
420              425              430
Trp Pro Asp His Gly Asp Arg Pro Arg Arg Ala Gly Val Ser Ser Phe
435              440              445
Gly Ala Ser Gly Thr Asn Ala His Val Ile Leu Glu Ser Ala Pro Gly
450              455              460
Ala Ala Ala Gly Ala Thr Gly Ala Thr Asp Leu Ser Ala Pro Pro Ala
465              470              475              480
Ser Val Ala His His Pro Ala Thr Ala Thr Ala Thr Ala Pro Ala Ala
485              490              495
Thr Val Pro Thr Ala His Glu Pro Ala Gly Thr Ala Gly Asp Asp Pro
500              505              510
Val Trp Val Leu Ser Gly Arg Thr Glu Ala Ala Leu Arg Glu Gln Ala
515              520              525
Arg Arg Leu His Ala His Leu Thr Ser Arg Ala Arg Pro Glu Pro Ala
530              535              540
Asp Ala Val Ala Arg Ala Leu Ala Arg Ser Arg Thr Ala Phe Ala Tyr
545              550              555              560
Arg Ala Ala Val Leu Gly Arg Asp Asp Thr Ala Arg Leu Asp Gly Leu
565              570              575
His Ala Leu Ala Ala Gly Arg Ser Ala Ala Gly Leu Val Thr Gly Arg
580              585              590
Ala Val Pro Glu Arg Arg Val Ala Phe Leu Phe Thr Gly Gln Gly Ser
595              600              605
Gln Arg Pro Gly Ala Gly Arg Glu Leu Tyr Ala Arg His Pro Ala Phe
610              615              620
Ala Gln Ala Leu Asp Gly Val Leu Ala Glu Leu Asp Arg His Leu Asp
625              630              635              640
Arg Pro Leu Arg Ala Val Met Leu Ala Glu Pro Gly Thr Glu Ala Ala
645              650              655
Ala Leu Leu Asp Asp Thr Ala Tyr Thr Gln Pro Ala Leu Phe Ala Leu
660              665              670
Glu Val Ala Leu Phe Arg Leu Val Thr Ser Trp Gly Leu Arg Pro Asp

```

675	680	685
Ala Leu Leu Gly His Ser Val Gly Glu Ile Thr Ala Ala Tyr Val Ala 690 695 700		
Gly Val Leu Thr Leu Pro Asp Ala Ala Arg Leu Val Ala Val Arg Gly 705 710 715 720		
Arg Leu Met Ala Asp Leu Arg Ala Gly Gly Ala Met Ala Ala Leu Gln 725 730 735		
Ala Ala Glu Ser Glu Val Asp Pro Leu Leu Ala Gly Arg Glu Gly Glu 740 745 750		
Leu Ser Ile Ala Ala Val Asn Gly Pro Gln Ala Thr Val Ile Ala Gly 755 760 765		
Asp Glu Ala Ala Val Glu Glu Gln Val Ala Leu Trp Arg Asp Arg Gly 770 775 780		
Arg Arg Ala Arg Arg Leu Arg Val Gly His Ala Phe His Ser Val Arg 785 790 795 800		
Met Asp Gly Met Leu Ala Glu Phe Glu Lys Ala Met Gly Asp Leu Arg 805 810 815		
Ala Gly Glu Pro Thr Ile Pro Val Val Ala Asn Val Arg Gly Ala Ile 820 825 830		
Ala Ser Gly Thr Asp Leu Arg Thr Ala Gly Tyr Trp Ile Arg His Ala 835 840 845		
Arg Glu Pro Val Arg Phe Leu Asp Gly Met Arg Ala Leu Arg Ala Glu 850 855 860		
Gly Val Asp Thr Phe Val Glu Leu Gly Pro Asp Gly Val Leu Thr Ala 865 870 875 880		
Met Ala Arg Asp Cys Leu Ala Asp Pro Ala Asp Pro Val Asp Leu Ala 885 890 895		
Asp Ala Ala Glu Pro Ala Gly Ala Ala Glu Pro Asp Arg Ser Leu Leu 900 905 910		
Phe Leu Pro Thr Leu Arg Arg Asp Arg Asp Asp Ala Val Ala Val Arg 915 920 925		
Glu Ala Leu Ala Ser Val His Val His Gly Leu Pro Val Asp Pro Val 930 935 940		
Ala Pro Leu Gly Asp Gly Pro Leu Ala Thr Asp Leu Pro Thr Tyr Pro 945 950 955 960		
Phe Gln Arg Ser Arg Tyr Trp Leu Asp Pro Arg Pro Gly Ala Arg Asp 965 970 975		
Leu Thr Ala Val Gly Leu Asp Val Ala Gly His Pro Leu Leu Ala Val		

980										985										990									
Ala Val Asp Leu Pro Asp Gly Ala	Gly Thr Val Trp Ser Gly Gln Leu																												
995	1000																												
Cys Val Arg Thr His Pro Trp	Leu Ala Asp His Ser Val Trp Gly																												
1010	1015																												
Arg Thr Val Val Pro Gly Thr	Ala Leu Leu Glu Ile Met His Arg																												
1025	1030																												
Val Arg Ala Glu Val Gly Cys	Thr Arg Val Ala Glu Leu Thr Phe																												
1040	1045																												
Glu Ala Pro Met Val Leu Ala	Asp Asp Gly Gly Val Arg Val Arg																												
1055	1060																												
Val Val Val Asp Gly Pro Asp	Ala Asp Gly Ala Arg Gln Val Arg																												
1070	1075																												
Ile His Ser Ala Pro Val Gly	Pro Glu Pro Pro His Trp Thr Arg																												
1085	1090																												
His Ala Ser Gly Arg Val Asp	Ser Ala Ala Pro Gly Pro Ala Ala																												
1100	1105																												
Gly Pro Pro Ala Trp Asp Ala	Gly Pro Gly Ser Asn Trp Pro Pro																												
1115	1120																												
Glu Gly Ala Glu Pro Val Gly	Val Glu Ser Glu Tyr Glu Arg Phe																												
1130	1135																												
Ala Asp Asn Gly Ile Gly Tyr	Gly Pro Ala Phe Arg Gly Leu Arg																												
1145	1150																												
Ala Ala Trp Arg Arg Gly Asn	Glu Thr Phe Ala Glu Val Arg Leu																												
1160	1165																												
Pro Glu Gly Tyr Ala Ala Glu	Ala Gly Asp Tyr Ala Val His Pro																												
1175	1180																												
Ala Leu Leu Asp Ala Ala Leu	His Ala Ile Val Phe Gly Asp Gln																												
1190	1195																												
Phe Pro Gly Gly Ala His Gly	Met Leu Pro Phe Ala Phe Thr Asp																												
1205	1210																												
Val Arg Val Phe Ser Ser Gly	Ala Asp Arg Leu Arg Val Arg Ile																												
1220	1225																												
Ala Pro Ala Asp Ala Asp Ser	Val Cys Val Thr Val Ala Asp Gly																												
1235	1240																												
Asp Gly Thr Pro Val Leu Ala	Ala Ala Thr Leu Ala Leu Arg Arg																												
1250	1255																												
Val Ala Ala Asp Arg Ile Ala	Ala Thr Val Thr Gly Gln Ala Pro																												

1265	1270	1275
Leu Tyr Arg Leu Glu Trp Ser Ala Val Arg Pro Ala Pro Val Ala 1280 1285 1290		
Thr Gly Ala Arg Phe Ala Val Val Gly Ala Asp Ala Pro Leu Pro 1295 1300 1305		
Ser Gly Ala Leu Gly Ala Gly Val Pro Val Gln Ala Tyr Pro Asp 1310 1315 1320		
Leu Gly Ala Leu Ala Gly Ala Leu Ala Thr Asn Gly Ala Pro Gly 1325 1330 1335		
His Val Leu Val Asp Phe Arg Arg Arg Ala Asp Gly Pro Ala Gly 1340 1345 1350		
Arg Gln Pro Gly Asp Val Gly Ala Arg Thr Arg Arg Ala Leu Ala 1355 1360 1365		
Val Val Gln Glu Trp Leu Ala Asp Asp Arg Phe Thr Gly Ser Arg 1370 1375 1380		
Leu Val Val Leu Thr Ser Gly Ala Val Asp Ala Gly Thr Ala Val 1385 1390 1395		
Thr Asp Pro Ala Ala Ala Gly Val Trp Gly Leu Leu Arg Val Ala 1400 1405 1410		
Gln Thr Glu His Pro Asp Arg Phe Val Leu Val Asp Thr Asp Asp 1415 1420 1425		
His Pro Asp Ser Leu Arg Ala Leu Pro Gly Ala Ile Val Ala Gly 1430 1435 1440		
Glu Pro Gln Leu Ala Leu Arg Ala Gly Thr Ala Ser Val Pro Gly 1445 1450 1455		
Leu Val Arg Val Pro Ala Gly Thr Gly Ala Ala Pro Pro Trp Ala 1460 1465 1470		
Ala Ala Gly Thr Val Leu Val Thr Gly Gly Thr Gly Met Leu Gly 1475 1480 1485		
Gly Ala Val Ala Arg His Leu Val Arg Arg His Gly Val Arg Arg 1490 1495 1500		
Leu Leu Leu Val Gly Arg Arg Gly Pro Asp Ala Pro Gly Ala Ala 1505 1510 1515		
Ala Leu Thr Arg Glu Leu Glu Glu Leu Gly Ala Ser Val Arg Val 1520 1525 1530		
Ala Ala Cys Asp Val Gly Asp Arg Gly Ala Val Thr Arg Leu Leu 1535 1540 1545		
Ala Gly Val Pro Ala Ala His Pro Leu Thr Ala Val Val His Ser		

1550		1555		1560
Ala Gly 1565	Leu Pro Asp Asp	Gly 1570	Val Leu Thr Ala	Gln Thr Gly Glu 1575
Arg Val 1580	Ala Ala Val Leu	Arg 1585	Ala Lys Ala Asp	Ala Ala Val Asn 1590
Leu His 1595	Glu Leu Thr Arg	His 1600	Leu Asp Leu Thr	Ala Phe Val Leu 1605
Phe Ser 1610	Ser Val Ala Gly	Thr 1615	Ile Gly Ser Ala	Gly Gln Ala Gly 1620
Tyr Ala 1625	Ala Ala Asn Ala	Phe 1630	Leu Asp Ala Phe	Ala Ser Trp Arg 1635
Gln Gly 1640	Gln Gly Leu Pro	Ala 1645	Thr Ala Leu Ala	Trp Gly Pro Leu 1650
Asp Gly 1655	Gly Met Ala Ala	Gly 1660	Leu Gly Thr Ala	Asp Val Ala Arg 1665
Leu Arg 1670	Arg Ser Gly Leu	Val 1675	Pro Leu Gly Val	Asp Asp Ala Leu 1680
Val Leu 1685	Phe Asp Ala Ala	Cys 1690	Ser Arg Pro Ala	Ala Ala Tyr His 1695
Pro Val 1700	Arg Leu Asp Pro	Ala 1705	Val Leu Arg Ser	His Ala Ala Ala 1710
Asp Ser 1715	Ala Val Pro Ala	Val 1720	Leu Leu Gly Pro	Ser Arg Ala His 1725
Pro Arg 1730	Asp Gly Thr Pro	Gly 1735	Lys Pro Ala Glu	Ala Ala Leu Ala 1740
Ala Leu 1745	Leu Thr Gly Arg	Ser 1750	Ala Ala Glu Arg	Thr Ala Ile Leu 1755
Thr Asp 1760	Leu Val Arg Thr	Glu 1765	Ala Ala Ala Val	Leu Gly His Gly 1770
Glu Ala 1775	Ala Met Leu Ser	Thr 1780	Gln Arg Ala Phe	Arg Asp Ala Gly 1785
Phe Asp 1790	Ser Leu Thr Ala	Val 1795	Asp Leu Arg Asn	Arg Leu Gly Ala 1800
Ala Thr 1805	Gly Leu Ser Leu	Pro 1810	Ala Ala Val Val	Phe Asp His Pro 1815
Thr Pro 1820	Ala Ala Leu Ala	Ala 1825	Tyr Leu Arg Thr	Glu Leu Asp Arg 1830
Arg Ser	Pro Thr Gly Gln	Gln	Phe Pro Thr Asp	Ala Ala Gly Val

1835	1840	1845
Leu Ala Met Leu Asp Arg	Leu Arg Asp Gly Ile	Ala Thr Val Val
1850	1855	1860
Arg Asp Asp Ala Asp Arg	Thr Arg Ala Ala Asp	Leu Leu Arg Val
1865	1870	1875
Leu Leu Ala Glu Val Gly	Gly Pro Gly Thr Gly	Pro Pro Arg Asp
1880	1885	1890
Thr Asp Gly Gly Ser Gly	Gly Glu Val Ser Asp	Arg Leu Arg Thr
1895	1900	1905
Ala Ser Asp Glu Glu Leu	Phe Asp Leu Leu Asp	Ser Asp Phe Arg
1910	1915	1920
Leu Ala		
1925		

<210> 13
 <211> 5778
 <212> DNA
 <213> micromonospora carbonacea subspecies aurantiaca

<400> 13
 gtgaccgcga acgaggaccg gatgcgtgag tacctcaagc gggtcaccgc cgagctggcc 60
 gggacgcggc gacgcctgcg cgagctggag gacagcgcgc gtgagcccat cgcgatcgtg 120
 ggcattgagct gccgggtgcc gggcgggggtg agcacgcccc aggacctgtg gcggctggtc 180
 gaggccggta ccgacgcgat ctccggcttc cccgacgacc ggggctggga tgcggggagg 240
 ctctacgacc cggatccgga ctcgaccgga acgagctacg tgcgcgaggg cggcttcctc 300
 tacgactgcg ccgagttcga cccggagttc ttcaccgtct cgcgccgcga ggcgctggcc 360
 atggaccgcg agcagcggct gctgctggag gccgcctggg agaccttcga acgggcgggg 420
 atgcccccg actcggcccc cggcaccgcg accgggggtct acgtcggggg gatgtacgac 480
 gactacggca gccggctgtc ggaggtgccg aaggacctgg agggctacct ggtcaacggc 540
 agcgcgggca gtgtcgcgtc gggccggatc gcgtacacgc tggggttgca ggggcgggcg 600
 gtgacggtcg acacggcctg ctctcgtcgt ctggtcgcgt tgcacctggc cgtgcaggcg 660
 ctgcggtcgg gcgagtgtga gctggccctg gcgggcgggg cgacggtgct cgccacgccg 720
 acgatgttcg tcgacttcgc ccggcagcgc ggtctcgccg aggacggccg ttgcaaggcg 780
 ttgcgggacg ccgccgacgg gaccgggttc ggcgagggcg tggggatgct gctgggtggaa 840
 cggctctcgg acgcgggtcc caaccgtcgc cagggtgctgg ccgtcgtgcg gggcagcgcg 900
 gtcaaccagg acggggcgag caacggcctg accgccccga acggtacggc ccagcaactg 960

gtcatccggc aggcgttgac caacgcgggg ctggccgcgg acgaggtgga cgcggtggag 1020
 gcacacggca ccggcaccgc gctgggcgat ccgatcgagg cgcaggcgct gctggcgacg 1080
 tacggccagg gccggccggc ggaccggccg ctctgtctgg gatccctgaa gtccaacatc 1140
 ggccacaccc aggccgcgcg aggggtcgcc ggggtgatca agaccgtgct ggcgctgcgt 1200
 cacgcgcggc tgccccggac cctgcacgtc gatcgccctt cgaccgggtt ggactggtcg 1260
 tcgggcgcgg tgccggtgct gaccgagggg cggccctggc ccgatcacgg cgaccggccc 1320
 cgcggggccg ggggtctctc gttcggcgcg agcggcacca acgcgcacgt catcctggag 1380
 agcgcggggc gtgcggcggc gggggcgacc ggggcgacgg acctctcggc ccgcccggca 1440
 tccgtcgccc accatccggc cacggccacg gccacggccc cggcggcgac ggtgcccact 1500
 gccacgaac cggcggggac ggccggcgac gacccgtctt gggtcctgtc cggccggacc 1560
 gaggcggccc tcgcgagca ggcccgcgcg ctacacgccc acctgacatc ccgggcgcgg 1620
 cccgagcccg ccgacgccgt ggcccgcgcg ctggcgcgct cccgcaccgc gttcgctac 1680
 cgggcgcgcg tgctggggcg ggacgacacc gcgcggctcg acggcctcca cgcgctcgcg 1740
 gcgggtcgca gcgcgcggg gctcgtcacc gggcggggcg tgccggagcg gcgcgtggcc 1800
 ttctcttca ccgggcaggg cagccagcga ccgggcgcgg gccgggaact gtacgcccgg 1860
 catccgcct tcgcacaggc cctggacggc gtcctcgcgg aactcgaccg gcacctggac 1920
 cggccgctgc gcgcgctcat gctcgccgag ccgggcaccg aggcggcggc gctgctggac 1980
 gacaccgct acaccagcc cgcctgttc gcgctggagg tggcgctgtt ccggctggtc 2040
 acgagctggg ggctgcggcc tgacgccttg ctgggccact cggtcgggga gatcacgcg 2100
 gcgtacgtcg cgggcgtcct caccctgccg gacgccgccg ggctgggtgg ggtgcgcggt 2160
 cgactcatgg cggacctgcg ggccggcggt gcgatggccg cgctccaggc cgccgagagc 2220
 gaggtcgacc ccctgttggc ggggcgggag ggcgaactgt cgatcgagc ggtcaacggg 2280
 ccgcaggcaa ccgtgatcgc gggcgacgag gcggccgtcg aggagcaggt cgcgctgtgg 2340
 cgtgaccggg gtgcgcgggc caggcgactg cgggtcggcc acgccttcca ctccgtacgg 2400
 atggacggga tgctcgcca gttcgagaag gcgatgggtg atctccgtgc cggcgagccg 2460
 acgatccccg tggtcgccaa cgtcaggggg gcgatcgct ccggcaccca cctccgtacg 2520
 gccgggtact ggatccggca cggcgcgag ccggtgcgtt tcctcgacgg catgcgtgcg 2580
 ctgcggggcg agggcgtcga cacgttcgtg gaactcggcc ccgacggagt gtcacggcg 2640

atggcgcgcg	actgcctggc	ggatcccgcc	gacccggtgg	atctcgcgga	cgccgcccag	2700
cccgcggggg	ccgcggagcc	cgaccgctcc	ctgctgttcc	tgcccaccct	gcgcccggac	2760
cgcgacgacg	cagtggccgt	gcgggaggcc	ctggcatccg	tccacgtgca	cgggcttccc	2820
gtcgacccgg	tcgcgccgct	cggcgacggc	ccgctcgcca	ccgacctgcc	cacctaccog	2880
ttccagcggg	cccgtactg	gctcgacccg	cgtcccgggg	cacgcgacct	gaccgccgtg	2940
ggcctcgacg	tggccgggca	cccgtgctc	gccgtcgccg	tggacctgcc	cgacggcgcc	3000
ggcacgggtct	ggagcgggtca	gctctgcgtg	cggacgcata	cgtggctcgc	cgaccacagc	3060
gtgtgggggg	gcacgggtgg	gccggggacc	gcgctgctgg	agatcatgca	ccgagtgcgc	3120
gccgaggtgg	gctgcacccg	ggtcgcgga	ctgaccttcg	aggcgccgat	ggtgctggcc	3180
gacgacgggg	gcgtccgct	gcgggtcgtc	gtcgacggac	cagacgccga	cggggcccgc	3240
caggctccga	tccactccgc	accggtgggg	cccgagcctc	cccactggac	ccggcacgcc	3300
tcggggccgcg	tcgacagcgc	cgcgccgggg	ccggccgccc	gcccacccgc	gtgggacgcc	3360
ggccctggca	gcaactggcc	gcccgagggg	gcggagcccg	tgggcgtcga	gagcgagtac	3420
gagcgcttcg	ccgacaacgg	catcgatac	ggccccgcct	tccgagggt	gcgcgccgcg	3480
tggcgctcgcg	ggaacgagac	gttcgcccag	gtccggctcc	ccgaggggta	cgccgccgag	3540
gcgggcgact	acgccgtcca	tccggcactg	ctggacgcgg	ccctgcacgc	gacgtcttc	3600
ggtgaccagt	ttcccgggtg	ggcacacggg	atgctgccgt	tcgccttcac	cgacgtgcgg	3660
gtgttcagct	ccggcgccga	ccggctccgg	gtgcgcatcg	cgcccgccga	tgccgactcg	3720
gtctgcgtga	ccgtcgccga	cggcgacggg	acgccgggtc	tcgccgcagc	cacctggcg	3780
ttgcgccggg	tcgccgccga	ccggatcgcg	gcgaccgtca	ccggccaggc	accgctgtac	3840
cggttggagt	ggtccgccgt	gcggcccgcc	ccggtggcca	ccggggcgcg	gttcgccgtc	3900
gtcggcgcg	acgccccgct	gccgtccggt	gcgctggggg	ccggggtgcc	cgtccaggcg	3960
tacccgga	tgggcgcgct	ggccggcgcg	ttggccacca	acggggcacc	gggccacgtg	4020
ctcgtcgact	tccgccgccc	cgccgacggc	ccggcagggc	ggcagcccgg	tgacgtgggt	4080
gcacggaccc	gacgggcgct	ggccgtcgtc	caggagtggc	tcgccgacga	ccgtttcacc	4140
ggctcacggc	tggtcgtgct	caccagcgga	gccgtggacg	ccggaacagc	cgtcaccgat	4200
ccggccgccc	ccgggggtgtg	gggcctgctg	cgggtcgccc	agaccgagca	tccggaccgg	4260
ttcgtcctcg	tggacaccga	cgaccacccc	gattcgctgc	gtgccctccc	cggggcgatc	4320
gttgcggggc	agccgcagct	ggcactgcgg	gccggcacgg	ccagcgttcc	gggcctggtg	4380

cgggtgccgg ccggcaccgg tgccgccccg ccgtggggccg cagccggcac cgtcctcgtc 4440
 accggggggca ccggcatgct cggcggcgcg gtggcccggc acctgggtccg ccggcacggg 4500
 gtccgcccgc tgctgctggt cggccggcgc gggccggacg caccggcgcg ggcggccctg 4560
 acccggggaac tggaggagct gggagcgtcc gtccgcgtcg ccgcctgcga cgtcggcgat 4620
 cgtggcgccg tgacgcgcct gttggccggg gttcccgcg cgcacccgt caccgcggtg 4680
 gtgcactcgg ccggcctgcc cgacgacggc gtgtgaccg cacagaccgg cgagcgggtc 4740
 gcggcggtgc tccgcgcaa ggcggacgca gcggtcaacc tgcacgaact caccggcat 4800
 ctcgacctca ccgccttcgt gctgttctcg tcggtagcgg ggacgatcgg cagcgccggg 4860
 caggccgggt acgccgcgc gaacgccttc ctcgacgcgt tcgcgagctg gcggcagggc 4920
 caggggctgc ccgccaccgc cctggcgtgg gggccgttg acggcgggat ggccgccggc 4980
 ctcggcactg cggacgtggc acggctgcgc cgggtccggc tcgtgccgt cggcgtggac 5040
 gacgcgctcg ttctcttcga cgcgcctgc tcccagaccg cggcggcgta ccaccccgtc 5100
 cgctcgatc cggcgggtgt gcggtccac gccgcgcgc acagcgcggt gcccgccgtc 5160
 ctgctcggtc cgagccgtgc gcaccgagg gacggtagc cggggaagcc tgccgaagcc 5220
 gccctcgccg cgctgctgac cggcaggtcg gcggccgagc gtacggcgat cctgaccgac 5280
 ctggtgcgga cggaggccgc cgcgttctc gggcatggcg aggcggcgat gctgagcacg 5340
 cagcgggcct tccgcgacgc cggcttcgac tcgtcaccg ccgtggacct ccgcaaccgg 5400
 ctcggcgcgg ccacgggcct cagcctgccg gccgcgcgtc tcttcgacca cccgaccccg 5460
 gcggccctgg ccgcctatct gcggaccgaa ctggaccgcc ggtcgccac cgggcaacag 5520
 ttcccagcgg acgccgcgg tgttctggcc atgtcgacc gcctgcggga cggaatcgcg 5580
 acggtcgtca gggacgacgc cgaccggacc cgcgcagccg acctgttgcg tgtcctgctc 5640
 gccgaggtcg gcggggcccg gacgggccc ccccgcgaca ccgacggcg ctccggcggc 5700
 gaggtcagcg accgcctccg gaccgcctcc gacgaggaac tggtcgacct gctcgacagc 5760
 gatttccgac tggcgtag 5778

<210> 14

<211> 3745

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 14

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

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

Gly Met Gly Ala Ala Leu Leu Asp Glu Ser Glu Val Phe Ala Glu Ser
 610 615 620
 Met Val Glu Cys Gly Arg Ala Leu Ser Gly Phe Val Asp Trp Asp Leu
 625 630 635 640
 Leu Glu Val Val Arg Gly Gly Ala Gly Glu Gly Val Trp Gly Arg Val
 645 650 655
 Asp Val Val Gln Pro Val Ser Trp Ala Val Met Val Ser Leu Ala Arg
 660 665 670
 Leu Trp Met Ser Val Gly Val Val Pro Asp Ala Val Val Gly His Ser
 675 680 685
 Gln Gly Glu Val Ala Ala Ala Val Val Gly Gly Val Leu Ser Val Ala
 690 695 700
 Asp Gly Ala Arg Val Val Ala Leu Arg Ser Arg Val Ile Gly Glu Val
 705 710 715 720
 Leu Ala Gly Gly Gly Ala Met Val Ser Val Gly Leu Pro Ile Val Asp
 725 730 735
 Val Gln Glu Arg Leu Ala Gly Trp Gly Gly Arg Leu Gly Val Ala Ala
 740 745 750
 Val Asn Gly Pro Ser Leu Thr Val Val Ser Gly Asp Val Asp Ala Ala
 755 760 765
 Val Gly Phe Val Gly Glu Cys Glu Arg Asp Gly Val Trp Val Arg Arg
 770 775 780
 Val Ala Val Asp Tyr Ala Ser His Ser Ala His Val Glu Ala Val Glu
 785 790 795 800
 Gly Met Leu Ser Gly Leu Leu Gly Gly Leu Cys Pro Gly Arg Gly Val
 805 810 815
 Val Pro Phe Tyr Ser Ser Val Val Gly Gly Val Val Asp Gly Val Gly
 820 825 830
 Leu Asp Gly Gly Tyr Trp Tyr Arg Asn Leu Arg Glu Arg Val Leu Phe
 835 840 845
 Ser Asp Val Val Gly Arg Leu Val Gly Asp Gly Phe Ser Gly Phe Val
 850 855 860
 Glu Cys Ser Gly His Pro Val Leu Ala Gly Gly Val Leu Glu Ser Val
 865 870 875 880
 Ala Val Val Asp Pro Asp Val Arg Pro Val Val Val Gly Ser Leu Arg
 885 890 895
 Arg Asp Asp Gly Gly Trp Gly Arg Phe Leu Thr Ser Val Gly Glu Ala
 900 905 910

Phe Val Gly Gly Met Ser Val Asp Trp Lys Gly Val Phe Ala Gly Ala
 915 920 925
 Gly Ala Arg Leu Val Asp Leu Pro Thr Tyr Pro Phe Gln Arg Arg His
 930 935 940
 Tyr Trp Ala Pro Thr Pro Thr Asn Pro Ala Thr Asn Pro Ala Thr Asn
 945 950 955 960
 Pro Ala Thr Asn Pro Ala Thr Gly Asp Thr Thr Thr Ala Asp Pro Ala
 965 970 975
 Gly Asp Leu Arg Tyr Arg Ile Thr Trp Lys Pro Leu Pro Thr Asp Asp
 980 985 990
 Pro Arg Pro Leu Thr Asn Arg Trp Leu Leu Met Val Pro Glu Ala Leu
 995 1000 1005
 Ala Gly Asp Gly Val Val Ala Gly Val Arg Gln Ala Leu Ala Ala
 1010 1015 1020
 Arg Gly Ala Ser Val Glu Leu Leu Thr Val Gly Thr Ala Asp Arg
 1025 1030 1035
 Ala Gly Leu Ala Ala Leu Leu Thr Ser Ala Ala Pro Gly Asp Pro
 1040 1045 1050
 Glu Ala Ala Gly Pro Ala Gly Val Val Ser Leu Leu Ala Leu Ala
 1055 1060 1065
 Glu Gly Ala Asp Ala Arg His Pro Ala Val Pro Leu Gly Leu Thr
 1070 1075 1080
 Ala Ser Leu Ala Leu Ile Gln Ala Leu Ala Asp Ala Gly Thr Gln
 1085 1090 1095
 Ala Arg Leu Trp Ala Val Thr Arg Gly Ala Val Ala Val Ser Ser
 1100 1105 1110
 Gly Glu Val Pro Asp Ala Gly Gln Ala Gln Val Trp Gly Leu Gly
 1115 1120 1125
 Arg Val Ala Ala Leu Glu Leu Pro Asp Arg Trp Gly Gly Leu Val
 1130 1135 1140
 Asp Leu Pro Ala Leu Thr Gly Glu Arg Ala Phe Ala Gln Leu Ala
 1145 1150 1155
 Asp Val Val Gly Gly Ser Asn Gly Glu Asp Gln Val Ala Val Arg
 1160 1165 1170
 Ala Ser Gly Val Tyr Gly Arg Arg Leu Val Arg Ser Arg Ala Thr
 1175 1180 1185
 Val Thr Ser Gly Asp Trp Pro Ala Arg Gly Thr Ile Leu Val Val
 1190 1195 1200

Gly	Asp	Thr	Gly	Pro	Val	Ala	Ala	Leu	Leu	Ala	Gly	Arg	Leu	Leu
1205						1210					1215			
Gly	Asp	Gly	Ala	Ala	His	Val	Val	Leu	Ala	Gly	Pro	Ala	Ala	Ala
1220						1225					1230			
Ser	Thr	Val	Gly	Leu	Thr	Gly	Gly	Ala	Asp	Arg	Val	Ala	Leu	Ile
1235						1240					1245			
Asp	Cys	Asp	Pro	Ser	Asp	Arg	Asp	Ala	Leu	Ala	Gly	Leu	Leu	Gly
1250						1255					1260			
Ala	Tyr	Arg	Pro	Thr	Thr	Ile	Val	Val	Ala	Pro	Pro	Ala	Val	Ala
1265						1270					1275			
Leu	Thr	Ala	Leu	Ala	Glu	Thr	Thr	Pro	Glu	Asp	Phe	Val	Ala	Ala
1280						1285					1290			
Val	Ala	Ala	Lys	Thr	Thr	Thr	Ala	Val	His	Leu	Asp	Ala	Leu	Ala
1295						1300					1305			
Ala	Glu	Ala	Glu	Leu	Glu	Leu	Asp	Ala	Phe	Val	Val	Phe	Ser	Ser
1310						1315					1320			
Val	Ser	Gly	Thr	Trp	Gly	Gly	Ala	Gly	His	Gly	Gly	Tyr	Ala	Ala
1325						1330					1335			
Gly	Thr	Ala	Arg	Leu	Asp	Ala	Leu	Val	Glu	Glu	Arg	Arg	Ala	Arg
1340						1345					1350			
Gly	Leu	Pro	Ala	Thr	Ala	Ile	Ala	Trp	Thr	Pro	Trp	Ala	Asp	Ala
1355						1360					1365			
Thr	Thr	Ala	Ala	Gly	Gly	Gln	Ala	Pro	Asp	Ala	Ser	Ala	Gly	Gly
1370						1375					1380			
His	Glu	Pro	Asp	Thr	Arg	Ala	Gly	Gly	Pro	Asp	Arg	Glu	Leu	Leu
1385						1390					1395			
Arg	Arg	Gly	Gly	Leu	Thr	Pro	Leu	Asp	Pro	Gly	Ala	Ala	Leu	Asp
1400						1405					1410			
Val	Leu	Arg	Gly	Ala	Val	Ala	Arg	Gly	Glu	Gly	Leu	Val	Thr	Val
1415						1420					1425			
Ala	Asp	Val	Asp	Trp	Ala	Arg	Phe	Val	Ala	Ser	Tyr	Thr	Ala	Ala
1430						1435					1440			
Arg	Pro	Thr	Thr	Leu	Phe	Asp	Glu	Leu	Pro	Glu	Leu	Arg	Ala	Thr
1445						1450					1455			
Arg	Glu	Ala	Glu	His	Thr	Pro	Ala	Glu	Asp	Ser	Ser	Ala	Gly	Gly
1460						1465					1470			
Glu	Leu	Val	Arg	Ala	Leu	Ser	Gly	Arg	Pro	Ala	Ala	Asp	Gln	His
1475						1480					1485			

Arg Thr	Leu Leu Arg Leu Val	Arg Ala His Val Ala	Ala Val Leu
1490	1495	1500	
Gly His	Asp Glu Ala Glu Ala	Ala Asp Pro Asp Arg	Ala Phe Arg
1505	1510	1515	
Glu Leu	Gly Phe Thr Ser Val	Thr Ala Val Asp Leu	Arg Asn Arg
1520	1525	1530	
Leu Asn	Ala Ala Thr Gly Leu	Asn Leu Pro Ala Ser	Val Val Phe
1535	1540	1545	
Asp His	Pro Ser Ala Arg Val	Leu Ala Ala Tyr Leu	Arg Ala Glu
1550	1555	1560	
Leu Leu	Gly Pro Glu Ala Asp	Glu Asp Thr Ala Glu	Ala Val Ala
1565	1570	1575	
Pro Pro	Ser Ala Pro Ala Gly	Ala Gly Asp Asp Glu	Pro Ile Ala
1580	1585	1590	
Val Ile	Gly Met Ala Cys Arg	Phe Pro Gly Gly Val	Asp Ala Pro
1595	1600	1605	
Asp Asp	Leu Trp Asp Leu Leu	Ala Lys Gly Arg Asp	Ala Ile Ser
1610	1615	1620	
Arg Phe	Pro Thr Asn Arg Gly	Trp Asp Val Asp Gly	Leu Tyr Asp
1625	1630	1635	
Pro Asp	Pro Glu Ala Pro Gly	Arg Thr Tyr Val Arg	Glu Gly Gly
1640	1645	1650	
Phe Leu	His Asp Ala Pro Asp	Phe Asp Ala Ala Phe	Phe Gly Ile
1655	1660	1665	
Ser Pro	Arg Glu Ala Leu Ala	Met Asp Pro Gln Gln	Arg Leu Leu
1670	1675	1680	
Leu Glu	Thr Thr Trp Glu Ser	Leu Glu Arg Ala Gly	Leu Asp Pro
1685	1690	1695	
Thr Ala	Leu Arg Gly Thr Arg	Thr Gly Val Phe Val	Gly Thr Asn
1700	1705	1710	
Gly Gln	His Tyr Met Pro Leu	Leu Arg Asp Gly Ala	Asp Asp Phe
1715	1720	1725	
Asp Gly	Tyr Leu Gly Thr Gly	Asn Ser Ala Ser Val	Met Ser Gly
1730	1735	1740	
Arg Leu	Ser Tyr Val Phe Gly	Leu Glu Gly Pro Ala	Val Thr Val
1745	1750	1755	
Asp Thr	Ala Cys Ser Ala Ser	Leu Val Ala Leu His	Leu Ala Val
1760	1765	1770	

Gln Ala	Leu Arg Arg Gly	Glu	Cys Thr Leu Ala Leu	Val Gly Gly
1775		1780		1785
Ala Thr	Val Met Ser Thr	Pro	Asp Met Leu Val	Glu Phe Ser Arg
1790		1795		1800
Gln Arg	Ala Met Ser Pro	Asp	Gly Arg Ser Lys	Ala Phe Ala Ala
1805		1810		1815
Ala Ala	Asp Gly Val Ala	Leu	Ser Glu Gly Ala	Ala Met Met Val
1820		1825		1830
Val Gln	Arg Leu Ala Asp	Ala	Glu Ala Ala Gly	His Glu Ile Leu
1835		1840		1845
Ala Val	Val Lys Gly Ser	Ala	Val Asn Gln Asp	Gly Ala Ser Asn
1850		1855		1860
Gly Leu	Thr Ala Pro Asn	Gly	Pro Ser Gln Glu	Arg Val Ile Arg
1865		1870		1875
Gln Ala	Leu Ala Asp Ala	Gly	Leu Arg Pro Asp	Gln Val Asp Ala
1880		1885		1890
Val Glu	Ala His Gly Thr	Gly	Thr Ala Leu Gly	Asp Pro Ile Glu
1895		1900		1905
Ala Gln	Ala Leu Leu Ala	Thr	Tyr Gly Arg Asp	Arg Pro Ala Gly
1910		1915		1920
Arg Pro	Leu Trp Leu Gly	Ser	Leu Lys Ser Asn	Ile Gly His Thr
1925		1930		1935
Gln Ala	Ala Ala Gly Ile	Ala	Gly Val Met Lys	Val Ile Leu Ala
1940		1945		1950
Leu Arg	His Asp Thr Leu	Pro	Arg Thr Leu His	Val Asp Arg Pro
1955		1960		1965
Thr Pro	Arg Val Asp Trp	Ala	Ser Gly Ala Val	Ser Leu Leu Thr
1970		1975		1980
Glu Pro	Val Pro Trp Pro	Gln	Gly Asp Glu Pro	Arg Arg Ala Ala
1985		1990		1995
Val Ser	Ser Phe Gly Ile	Ser	Gly Thr Asn Ala	His Val Ile Val
2000		2005		2010
Glu Gln	Ala Pro Pro Val	Val	Arg Glu Pro Ile	Asp His Glu Ala
2015		2020		2025
Asp Glu	Val Thr Val Pro	Leu	Phe Leu Ser Ala	Arg Gly Ser Ala
2030		2035		2040
Ala Leu	Cys Ala Gln Ala	Ala	Arg Leu Arg Ala	Arg Leu Ile Glu
2045		2050		2055

- 87 -

Ser Thr	Val Thr Gly Gly	Leu	Leu Asp Thr Ala	Leu	Leu Asp Ala
2345		2350		2355	
Ala Tyr	Trp Tyr Arg Asn Met	Arg Asp Pro Val	Glu	Phe Glu Gln	
2360		2365		2370	
Ala Thr	Arg Ala Met Leu Ala	Asp Gly His Glu Gly	Phe Leu Glu		
2375		2380		2385	
Pro Ser	Pro His Pro Met Leu	Ser Val Ser Leu Gln	Gly Thr Ala		
2390		2395		2400	
Ala Asp	Ala Gly Val Ala Ala	Thr Val Leu Gly Thr	Leu Arg Arg		
2405		2410		2415	
Gly Lys	Gly Gly Ala Arg Trp	Phe Gly Met Ala Leu	Gly Leu Ala		
2420		2425		2430	
His Ala	His Gly Ile Glu Ile	Asp Ala Ser Val Leu	Phe Gly Thr		
2435		2440		2445	
Asp Ser	Arg Arg Val Asp Leu	Pro Thr Tyr Pro Phe	Gln Arg Glu		
2450		2455		2460	
Arg Phe	Trp Tyr His Pro Pro	Ala Ala Arg Gly Asp	Val Ala Ser		
2465		2470		2475	
Ala Gly	Leu Ser Gly Ala Asp	His Pro Leu Leu Gly	Gly Ala Val		
2480		2485		2490	
Glu Leu	Pro Asp Arg Gly Gly	His Val Tyr Pro Ala	Arg Leu Gly		
2495		2500		2505	
Val Arg	His His Pro Trp Leu	Gly Glu His Ala Leu	Leu Gly Ala		
2510		2515		2520	
Ala Ile	Leu Pro Gly Ala Ala	Tyr Ala Glu Leu Ala	Leu Trp Ala		
2525		2530		2535	
Gly Arg	Arg Asp Gly Ala Gly	Arg Ile Glu Glu Leu	Thr Leu Asp		
2540		2545		2550	
Ala Pro	Leu Val Val Ala Asp	Glu Ser Ala Ala Gln	Leu Arg Leu		
2555		2560		2565	
Val Val	Gly Pro Ala Asp Ala	Glu Gly Arg Arg Gln	Leu Thr Val		
2570		2575		2580	
His Ser	Arg Ala Asp Gly Ala	Asp Ala Asp Thr Ala	Trp Thr Arg		
2585		2590		2595	
His Ala	Gln Gly Thr Leu Val	Pro Ala Asp Ala Asp	Ala Ala Gly		
2600		2605		2610	
Ser Gly	Asp Pro Gly Ala Pro	Trp Pro Pro Ala Gly	Ala Glu Pro		
2615		2620		2625	

Val Glu	Val Ala Gly Leu Tyr	Asp Arg Phe Ala Asp	Arg Gly Tyr
2630	2635	2640	
Gln Tyr	Gly Pro Ser Phe Arg	Gly Val Arg Ala Ala	Trp Arg Ala
2645	2650	2655	
Gly Asp	Thr Val Tyr Ala Glu	Val Ala Leu Pro Val	Pro Gln Pro
2660	2665	2670	
Gly Ser	Pro Arg Phe Gly Val	His Pro Ala Leu Leu	Asp Ala Ala
2675	2680	2685	
Phe Gln	Ala Met Ser Leu Gly	Ala Phe Phe Pro Glu	Asp Gly Gln
2690	2695	2700	
Val Arg	Met Pro Phe Ala Leu	Arg Gly Val Ser Ser	Ser Gly Val
2705	2710	2715	
Gly Ala	Asp Arg Leu Arg Val	Thr Ile Ser Pro Ala	Gly Ala Glu
2720	2725	2730	
Ala Val	Arg Ile Ala Cys Val	Asp Glu Arg Gly Asn	Pro Val Val
2735	2740	2745	
Val Ile	Asp Ser Leu Val Ala	Arg Ala Val Pro Val	Glu Ala Leu
2750	2755	2760	
Thr Pro	Gly Thr Pro Gly Thr	Gly Asp Gly Ala Leu	His His Val
2765	2770	2775	
Ala Trp	Thr Ala Arg Pro Glu	Pro Gly Val Ala Ala	Val Gln Arg
2780	2785	2790	
Trp Ala	Val Val Gly Ala Ala	Asp Pro Gly Leu Ala	Gly Gly Leu
2795	2800	2805	
Asp Arg	Ala Gly Gly Leu Cys	Gly Ala Tyr Pro Asp	Leu Ala Gly
2810	2815	2820	
Leu Val	Ala Ala Val Ala Glu	Gly Ala Ala Leu Pro	Asp Val Val
2825	2830	2835	
Ala Val	Pro Val Pro Ser Gly	Ala Pro Val Gly Pro	Asp Ala Val
2840	2845	2850	
Arg Ala	Thr Val Leu Gly Ala	Leu Asp Leu Ile Arg	Ala Trp Leu
2855	2860	2865	
Ala Val	Glu Gly Arg Leu Gly	Leu Ala Arg Leu Ala	Phe Val Thr
2870	2875	2880	
Thr Ser	Ala Val Ala Val Gly	Asp Gly Thr Glu His	Val Asp Pro
2885	2890	2895	
Val Ser	Ala Ala Leu Trp Gly	Leu Val Arg Ser Ala	Gln Ser Glu
2900	2905	2910	

Glu	Pro	Gly	Arg	Phe	Val	Leu	Val	Asp	Leu	Asp	Ala	Asp	Pro	Ala
2915						2920					2925			
Ser	Ala	Ser	Ala	Leu	Pro	Ala	Ala	Leu	Ala	Ala	Gly	Glu	Pro	Gln
2930						2935					2940			
Leu	Ala	Val	Arg	Ala	Gly	Ala	Val	His	Val	Pro	Arg	Leu	Val	Arg
2945						2950					2955			
His	Arg	Pro	Arg	Pro	Asp	Gly	Pro	Leu	Thr	Pro	Pro	Ala	Gly	Ala
2960						2965					2970			
Ala	Trp	Arg	Leu	Ala	Ala	Gly	Gly	Gln	Gly	Thr	Leu	Glu	Gly	Leu
2975						2980					2985			
Ala	Leu	Val	Pro	Ala	Pro	Asp	Ala	Leu	Ala	Pro	Leu	Ala	Pro	Gly
2990						2995					3000			
Gln	Val	Arg	Val	Ala	Val	Arg	Ala	Ala	Gly	Val	Asn	Phe	Arg	Asp
3005						3010					3015			
Thr	Leu	Ile	Ala	Leu	Gly	Met	Tyr	Pro	Gly	Thr	Pro	Val	Leu	Gly
3020						3025					3030			
Ala	Glu	Gly	Ala	Gly	Val	Ile	Thr	Glu	Val	Ala	Pro	Asp	Val	Ala
3035						3040					3045			
Gly	Phe	Ala	Pro	Gly	Asp	Arg	Val	Leu	Gly	Met	Trp	Thr	Gly	Gly
3050						3055					3060			
Leu	Gly	Pro	Val	Ala	Val	Ala	Asp	Ala	Arg	Met	Leu	Ala	Arg	Val
3065						3070					3075			
Pro	Arg	Gly	Trp	Ser	Tyr	Ala	Glu	Ala	Ala	Ser	Val	Pro	Ala	Val
3080						3085					3090			
Phe	Leu	Thr	Ala	His	Tyr	Ala	Leu	Thr	Arg	Leu	Ala	Gly	Ile	Arg
3095						3100					3105			
Pro	Gly	Gln	Ser	Leu	Leu	Val	His	Ala	Gly	Ala	Gly	Gly	Val	Gly
3110						3115					3120			
Met	Ala	Thr	Leu	Gln	Leu	Ala	Arg	His	Leu	Gly	Val	Glu	Val	Tyr
3125						3130					3135			
Ala	Thr	Ala	Ser	Arg	Gly	Lys	Trp	Asp	Thr	Leu	Arg	Gly	Leu	Gly
3140						3145					3150			
Leu	Asp	Asp	Ala	His	Ile	Ala	Asp	Ser	Arg	Ser	Leu	Asp	Phe	Ala
3155						3160					3165			
Gly	Arg	Phe	Leu	Ala	Ala	Thr	Gly	Gly	Arg	Gly	Val	Asp	Val	Val
3170						3175					3180			
Leu	Asn	Ser	Leu	Ala	Gly	Asp	Phe	Val	Asp	Ala	Ser	Leu	Arg	Leu
3185						3190					3195			

Leu Pro Arg Gly Gly His Phe	Leu Glu Leu Gly Lys	Ala Asp Val
3200	3205	3210
Arg Asp Pro Asp Arg Ile Ala	Ala Asp His Pro Gly	Val Gly Tyr
3215	3220	3225
Arg Ala Phe Asp Leu Val Glu	Ala Gly Pro Glu Leu	Val Gly Gln
3230	3235	3240
Leu Leu Gly Glu Leu Met Glu	Leu Phe Ala Ala Gly	Val Leu Ser
3245	3250	3255
Pro Leu Pro Leu Thr Val Arg	Asp Val Arg Arg Ala	Arg Glu Ala
3260	3265	3270
Phe Arg Leu Ile Ser Gln Ala	Arg His Val Gly Lys	Val Val Leu
3275	3280	3285
Thr Met Pro Pro Ala Phe Gly	Ala Tyr Gly Thr Val	Leu Val Thr
3290	3295	3300
Gly Gly Thr Gly Thr Leu Gly	Gly Ala Val Ala Arg	His Leu Val
3305	3310	3315
Ala Arg His Gly Val Arg His	Leu Val Leu Thr Gly	Arg Ser Gly
3320	3325	3330
Pro Ala Ala Asp Gly Ala Ser	Ala Leu Val Asp Glu	Leu Thr Ala
3335	3340	3345
Ser Gly Ala Ser Val Thr Val	Val Ala Cys Asp Ala	Ala Asp Arg
3350	3355	3360
Val Ala Leu Arg Arg Leu Leu	Asp Gly Ile Pro Ala	Ala His Pro
3365	3370	3375
Leu Thr Ala Val Val His Ala	Ala Gly Val Leu Asp	Asp Ala Thr
3380	3385	3390
Ile Thr Ala Leu Thr Ala Gly	Gln Val Asp Ala Val	Leu Arg Pro
3395	3400	3405
Lys Ala Asp Ala Val Ile Asn	Leu His Glu Leu Thr	Arg Asp Arg
3410	3415	3420
Glu Leu Ser Ala Phe Val Leu	Phe Ser Ser Ala Ala	Ala Leu Phe
3425	3430	3435
Gly Ser Pro Gly Gln Gly Asn	Tyr Ser Ala Ala Asn	Gly Phe Val
3440	3445	3450
Asp Ala Phe Ala Gln Tyr Arg	Arg Ala Gln Gly Leu	His Ala Val
3455	3460	3465
Ser Leu Ala Trp Gly Leu Trp	Ala Asp Ser Ser Arg	Met Ala Gly
3470	3475	3480

His Leu Asp Gln Glu Gly Met	Arg Arg Arg Met Ala Arg Gly Gly
3485	3490 3495
Val Leu Pro Leu Thr Thr Asp	Gln Gly Leu Ala Leu Phe Asp Ala
3500	3505 3510
Ala Gln Leu Val Asp Glu Ala	Leu Gln Val Pro Ile Arg Leu Asn
3515	3520 3525
Val Gly Ala Leu Arg Ala Ala	Gly Arg Val Pro Ala Leu Leu Ala
3530	3535 3540
Asp Leu Val Pro Ala Ala Ala	Ser Gly Ala Pro Ala Ala Thr Pro
3545	3550 3555
Thr Arg Asp Asp Ala Asp Arg	Thr Leu Ala Asp Arg Leu Ala Gly
3560	3565 3570
Leu Thr Val Ala Glu Gln Arg	Glu Leu Val Leu Glu Ser Val Arg
3575	3580 3585
Gly His Ala Ala Ala Val Leu	Gly His Ala Asp Pro Gln Ala Val
3590	3595 3600
Asp Ala Asp Arg Ala Phe Arg	Glu Leu Gly Phe Asp Ser Leu Thr
3605	3610 3615
Ala Val Glu Leu Arg Asn Arg	Leu Ala Thr Ala Ser Gly Leu Arg
3620	3625 3630
Leu Pro Ala Thr Leu Val Phe	Asp His Pro Thr Pro Glu Ala Leu
3635	3640 3645
Ala Glu His Leu Leu Ala Gly	Leu Ala Pro Glu Gln Ala Arg Ala
3650	3655 3660
Glu Leu Pro Leu Leu Ala Glu	Leu Gly Arg Leu Glu Ala Ala Leu
3665	3670 3675
Ala Ala Thr Asp Gly Ala Ala	Leu Asp Gly Leu Asp Asp Leu Val
3680	3685 3690
Arg Arg Glu Val Gly Val Arg	Ile Ala Ala Leu Ala Ala Arg Trp
3695	3700 3705
Gly Ala Ala Gly Asp Asp Val	Ala Gly Ser Asp Gly Gly Gly Thr
3710	3715 3720
Ala Asp Ala Leu Glu Ser Ala	Asp Asp Asp Glu Ile Phe Ala Phe
3725	3730 3735
Ile Asp Glu Arg Phe Arg Ala	
3740	3745

<210> 15
 <211> 11238
 <212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 15

```

gtgtctgtca acaacgaaga caagcttcgc gagtatctgc gtcgtgccat ggcggatctc      60
catgagtccc gcgagcgggt gcggcagtac gagtccgctg ctgctgtgga tgatccgggtg      120
gtggtggtgg ggatgggttg tcgttttccg ggtggggtgg tgtgtgcgga gggtttgtgg      180
gatttggtgt tgggggggtg ggatgcggtg tcggggtttc cggtgatcg gggttgggat      240
gtggaggggt tgtttgatcc ggtgcgggggt gtggtgggga agtcgtatgt gcgggagggg      300
gggtttgtgt atgacgcggg gatgttcgat gcggagtttt ttggtgtgtc gccgcgtgag      360
gcggtggcga tggatccgca gcagcgtttg tttttggagg tgtcgtggga ggcgttggag      420
cgtgcgggga ttgatccggt gggtttgcgg ggttcgcgga cgggtgtgta tgtgggggtg      480
atgggtcagg agtatgggcc gcggttggtg gagtcgggtg gtgggtttga gggttatttg      540
ttgacgggga cgtcgccgag tgtggtgtcg ggtcgtgttt cgtatgtgtt ggggttggag      600
ggtccgtcga tttcggttga tacggcgtgt tcgtcgtcgt tggtggcgtt gcatttggcg      660
tgtcaggggt tgcggttggg tgagtgtgat gtggcgttgg cgggtggggg gacggtgatt      720
gcggcgccgg ggttgtttgt ggagttttct cggcaggggt ggttgtcggg tgatgggcgg      780
tgtcgggcgt ttgcgggtgg tgcggatggg acgggggtgg gggaggggtg ggggggtggtg      840
gtgttgagac ggttgtcggg ggcgcgggag cgtgggtcatc ggggtgttggc ggtggtgcgg      900
ggttctgcgg tgaatcagga tggtggtcgt aatggtttga cggcgccgtc gggggtggcg      960
cagcgtcggg tgattggtgc ggcgttggtg gcggcgggtt tgggtgtgtc ggatgtggat     1020
gtggtggagg cgcattgggac ggggactcgg ttgggtgatc cgattgaggc tgaggcgttg     1080
ttggggtcgt atgggcgggg tcgtgtgggt ggggcgttgt tgttgggttc ggtgaagtcg     1140
aatattggtc atacgcaggc ggtgcgggt gtggcggttg tgatcaagat ggtgatggcg     1200
ttgcgggcgg ggggtggtgc ggcgacgttg catgtggatg tgccgtcgcc gttggtggat     1260
tggtcttcgg gtgggggtga gttggtgacg gaggcgcggg attggccggt ggtgggtcgt     1320
gtgcgtcgtg cgggtgtgtc ggcgtttggg gtgtcgggga cgaatgcgca totgattttg     1380
gagcagggcc ccgagttcga cgatcctgcc gattccgatt ccgattccga ttccgatgcc     1440
ggtgtcgtgg atggcggcga ggggtggtgt ggcaggagct tgtcgggtgt tccggtggtg     1500
gtgtcgggtc gttcgggtgg ggctttgcgg gcgtatgcgg gtcggttgcg tgagggtgtgc     1560
gcggggttgt ctgacggtgg tggctccggt ggtgggttctg gtttgggtgga tgtgggttgg     1620

```

tcgttggtgt	cgctcgcggtc	ggtgtttgag	catcggggcgg	tcgtgttcgg	tgggggtgtg	1680
gaggaggttg	ttgctggtct	tggcgcggtg	gcttctgggg	cggtggcttc	gggttcggtg	1740
gtggtgggtt	cggtggcgtc	gggtgttgct	ggtggtggtg	gtcgggtggt	gtttgtgttt	1800
ccgggtcagg	gttggcagtg	ggtgggtatg	ggtgcggcgc	tgctggacga	gtcggaggtg	1860
ttcgccgagt	cgatggtgga	gtgtggtcgg	gcgttgtcgg	ggtttgtgga	ttgggatttg	1920
ttggaggtgg	tgcgcggcgg	ggcgggtgag	gggggtgtgg	gtcgggttga	tgtggtgcag	1980
ccggtgtcgt	gggcgggtgat	ggtgtcgttg	gcgcggttgt	ggatgtcgg	gggtgtggtg	2040
ccggatgcgg	tgggtgggtca	ttcgccaggt	gaggttgctg	cgccgggtggt	gggggggtgtg	2100
ttgagtgtgg	ctgatggggc	gcgggtggtg	gcgttgccgg	cgccgggtgat	cggtgaggtg	2160
ttggccgggtg	gtggtgcgat	ggtgtcggtc	ggactgccga	tcgtggatgt	gcaggaacgg	2220
ttggcgggg	gggggtggtcg	gttgggtgtg	gcggcggtga	atggtccgtc	gttgacgggtg	2280
gtgtcggggg	atgtggatgc	tgctgtgggg	tttgttggtg	agtgtgagcg	ggatgggggtg	2340
tgggtgcggc	gggtggcggt	ggattatgcg	tcgcattcgg	cgcatgtgga	ggcgggtggag	2400
gggatgctgt	cggggttgtt	gggtggtttg	tgtccggggc	ggggtgtggt	gcggttttat	2460
tcgtcgggtg	tgggtggtgt	ggttgatggg	gtgggttttg	atggtgggta	ttggtatcgg	2520
aatctgcgtg	agcgggtgtt	gttttcggat	gtggtggggc	ggcttgttgg	ggatgggttt	2580
tcggggtttg	tggagtgttc	ggggcatccg	gtgttggcgg	gtgggtgtt	ggagtccggtg	2640
gcgggtggtg	atccggatgt	gcggccgggtg	gtggtgggg	cgctgcgcgg	tgatgatggt	2700
gggtggggcc	ggtttctgac	gtcgggtgggt	gaggcgttcg	tcggcgggat	gagtgttgac	2760
tggaaggggtg	tgttcgcggg	ggcggggcgcg	cggttggttg	acctgcgcgac	gtatccgttc	2820
caacgccgcc	actactgggc	accgaactcc	accaaccccg	ccaccaaccc	cgccaccaac	2880
cccgccacca	accccgccac	gggcgacacc	accaccgccc	acccggcggg	tgacctgcgg	2940
tatcgatca	cctggaaacc	gttgccgacc	gacgaccccc	gacccctcac	caaccgctgg	3000
ctgctgatgg	tgcccagaggc	gctggccgggt	gacgggggtg	tggcgggct	acggcaggcg	3060
ctggccgcgc	gtggcgctc	cgtcgaactg	ctgaccgtcg	gcaccgcga	ccgggcccggc	3120
cttgccgcgc	tcctgacctc	cgccgcccc	ggcgaccgg	aggcggccgg	cccggcgggc	3180
gtggtctccc	tgctggcgct	cgccgagggc	gcggacgcgc	gccacccggc	cgtaccgctc	3240
ggcctgaccg	cctcgctcgc	cctgatccag	gcattggcgg	acgcggggac	gcaggcccgc	3300
ctctgggcgg	tcacccgggg	ggccgtcgcc	gtgtcctccg	gcgaggtgcc	ggacgcgggg	3360

cagggcccagg	tgtgggggct	cggccggggtc	gcggccctcg	aactgccgga	ccgatggggc	3420
gggctggtgg	acctgccggc	gtcacccggg	gagcgtgcct	tcgcgcagct	cgccgatgtc	3480
gtgggcccgt	cgaacggcga	ggaccagggtc	gccgtacggg	cctccggcgt	ctaccgtcga	3540
cgcctcgtgc	gttcccgcgc	caccgtcacg	tccggcgact	ggccggccccg	gggcaccatc	3600
ctcgtcgtcg	gggacaccgg	cccggtcgcc	gcgtcctcgg	ccggccgcct	cctcggcgac	3660
ggggcggcgc	acgtggtgct	cgccggccccg	gccgccgcgt	ccaccgtcgg	gtcacccggc	3720
ggggccgacc	gggtggccct	gatcgactgc	gacccgagcg	accgggacgc	gctcgccggg	3780
ctgctcggcg	cgtaccggcc	cacgacgatc	gtggtggctc	cgcccgcctg	cgcgctcacc	3840
gccctcgccg	agaccacgcc	ggaggacttc	gtcgccgcgg	tcgccgcgaa	gacgacgacg	3900
gcagtgcacc	tcgacgccct	tgcggcggag	gcggaactgg	agctcgacgc	gttcgtcgtc	3960
ttctcctcgg	tctccggcac	ctggggcggc	gcggggcacg	gcggctacgc	ggcgggcacc	4020
gcccggctgg	acgcgtggt	cgaggagagg	cgggcccggtg	gcctgcccgc	cacggcgatc	4080
gcgtggacgc	cgtggggcca	cgcgaccaca	gccgccggcg	ggcaggcacc	cgatgccagc	4140
gccggcgggc	acgaacccca	cacgagggcc	ggggggccccg	accgcgaact	gctgcgccgg	4200
ggtggcctca	ccccgttggg	cccggggggc	gcgctggacg	tgctgcgcgg	ggcgggtggc	4260
cggggcgagg	gcctggtgac	cgtggccgac	gtcgactggg	cgcggttcgt	cgcctcgtac	4320
accgcggccc	ggcccaccac	gctcttcgac	gaactgcccc	agctgcgggc	gacccgggag	4380
gcggagcaca	ccccggccga	ggactcgtcg	gccggcggcg	aactggtccg	tgccctcagc	4440
ggccggcccc	cggccgatca	gcaccggacg	ctgctgcggc	tggtccgtgc	gcacgtcgcg	4500
gccgtcctgg	ggcacgacga	ggccgaggcg	gccgatccgg	accgggcgtt	ccgggaactc	4560
ggcttcacct	cggtgacggc	ggtggacctg	cgggaaccggc	tgaacgcggc	caccgggctg	4620
aacctgccgg	cgtccgtcgt	cttcgaccat	cccagcggcc	gggtgctggc	cgcgtacctg	4680
cgtgccgagc	tgctcggggc	ggaggccgac	gaggacacgg	cggaggccgt	cgccccgcgc	4740
tccgcgccgg	ccggggcggg	cgacgacgag	ccgatcgccg	tgatcgggat	ggcctgtcgg	4800
ttcccggggc	gggtcgacgc	ccccgacgac	ctgtgggatc	tgctggcgaa	gggcccgcgac	4860
gccatctcca	ggttccccac	gaaccggggc	tgggacgtcg	acggcctgta	cgacccggac	4920
ccggaggcgc	ccggccgcac	ctacgtccgc	gaggggcggt	tcctgcacga	cgcgccccgac	4980
ttcgatgccg	cgttcttcgg	gatctcgccc	cgcgaggccc	tcgccatgga	tccgcagcag	5040

cgctgctgc tggagaccac gtgggagtc ctggaacggg ccgggttggg cccgaccgcg 5100
 ttgcgcggca cccggaccgg ggtgttcgtg gggaccaacg gccagcacta catgccgctg 5160
 ctgcgagacg gcgcggacga cttcgacggc tacctcggca ccggcaactc ggccagcgtc 5220
 atgtccggcc ggctctccta cgtcttcggc ctggagggcc cggcgggtgac cgtggacacg 5280
 gcctgctccg cctccctcgt ggcgctgcac ctgcgggtgc aggcgctgcg ccggggcgag 5340
 tgcacgctgg ccctggtcgg cggggccacg gtgatgtcga cgccggacat gctggtggag 5400
 ttctcccggc agcggggcat gtccgggac ggcgggtcga aggcgttcgc cgccgccgcc 5460
 gacgggggtg cgctcagcga gggcgccgc atgatggtg tgcagcggct cgccgacgcg 5520
 gagggccgcg ggcacgagat cctggccgtg gtcaagggtc cggccgtcaa ccaggacggg 5580
 gccagcaacg gcctcacgc cccgaacggg ccctcccagg aacgggtcat ccggcaggcg 5640
 ctggccgacg ccggcctgcg gccggaccag gtggacgcgg tcgaggcgca cggcaccggc 5700
 accgccctgg gcgaccccat cgaggcgag gcgctgctc ccacgtacg ccgggaccgg 5760
 ccggcggggc ggccactgtg gtcggctcg ctgaagtcca acatcgtca caccaggcc 5820
 gccgccggca tcgccgggt gatgaagggt atcctgggcg tcgggcacga cacgtgccg 5880
 cgcacgctgc acgtggaccg gccgacgcc cgggtggact gggcttcgg ggcggtgtcg 5940
 ttgctgaccg agccggtgcc gtggccgcag ggcgacgaac cccgccgggc ggcggtgtcc 6000
 tcgttcggga tcagcggcac caacgcccac gtgatcgtc agcaggcgcc gccggtggtg 6060
 cgggaaccga tcgaccacga ggcggacgag gtcaccgtcc cgctgttct gtccggccgg 6120
 gggagcgccg cgctctgcg ccaggcggca cggctgcggg ccgggttgat cgaggaaccc 6180
 gacctggaca tcgccgaggt cggctacacg ctggcggcca cccgggcccg cttcgagcac 6240
 cgggccgtg tgatcgggga gagccgcgc gaggtcggcg acgcgtcgc cgcgtggcc 6300
 cggggcgagg agcaccgctc gctgctgcg gggcgggccc gcgcgagcga ccgggtcgcg 6360
 ttctctttc ccggccaggg ctgcagtg ggcgagatg ccgacggcct gctcgaccgc 6420
 tcccggcct tccgggcgag cgcgtcggcg tcgacgagg cgctcgggc gcacctcgac 6480
 tggctcgtgc tggacgtgct gcgtcgcgt cggacgcgc ctgcgtgag ccgggtcgac 6540
 gtggtccagc cgggtgctgt caccgatgat gtgtcgtg cggcggcctg gcgggcgctg 6600
 ggcgtgcacc cgtccgcctt ggtcggccac tcgaggggtg agatcgcggc ggcccacgtg 6660
 gcgggcggcc tctcgtgga cgacgcggcg cgcacgtcg ccctgcgcag ccaggcgtgg 6720
 ctgcggctgg ccgggcaggg cgggatggtg gcggtgtcg tcccgtcga cgcgtccgc 6780

gcccgcctgg	cgcggttcgg	cgaccggtg	tccgtcgccg	cggtcaacag	ccccggtacg	6840
gcggcggtga	gcggtaccc	cgacgcgtc	gccgaactcg	tcgacgagct	gaccgcccag	6900
ggcgtgcacg	ccaaggcgat	cccgggggtg	gacacggccg	ggcactccgc	gcagggtggag	6960
gtgctgaagg	accacctgat	ggccgccctc	gccccggtgt	cgccccgcag	ctcgcagatc	7020
cccttctact	cgaccgtcac	gggcggcctg	ctggacaccg	cgctgctgga	cgccgcctac	7080
tggtaacgca	acatgcgcga	cccgggtggag	ttcgagcagg	cgacccgggc	gatgctcgcg	7140
gacgggcacg	aggggttcct	ggagcccagc	ccgcacccca	tgctgtcggg	gtcgttgacg	7200
ggcaccgcgg	ccgatgccgg	ggtcgccgcg	acgggtgctg	ggacactgcg	gcgcggcaag	7260
ggcggcgccc	gctgggttcg	catggcgctc	gggctcgccc	acgcccacgg	gatcgagatc	7320
gacgcgagtg	tgctcttcgg	aaccgactcg	cgccgggtcg	acctgccgac	gtacccgttc	7380
cagcgcgagc	gcttctggta	tcaccgcgcg	gccgcgcgcg	gggacgtggc	ctccgcgggg	7440
ctcagcggtg	ccgaccatcc	gctgctgggc	ggggcgggtc	agctgcctga	ccggggcggc	7500
cacgtgtatc	cggcccggct	cggcgccga	caccaccctg	ggctcggcga	gcatgccctg	7560
ctgggcgcgg	cgatcctgcc	cggggccgcg	tacgcggaac	tcgccctgtg	ggccgggcgg	7620
cgtgacgggg	ccggccggat	cgaggagctg	accctcgacg	cgccgctggg	ggtggccgac	7680
gagtcggcgg	cgcaactgcg	gctcgtgggt	ggcccggcgg	acgcggaggg	gcgccggcag	7740
ctcaccgtcc	actcgcgcgc	cgacggcgcg	gacgcggaca	ccgcgtggac	ccggcacgcg	7800
cagggcaccc	tcgtgccggc	cgacgccgac	gccgcgggga	gcggggaccc	gggcgcgccc	7860
tggccgcggg	ccggggccga	gcccgctcag	gtggcggggc	tgtacgaccg	gttcgccgac	7920
cggggctacc	agtacggggc	gtcgttcggg	gggggtccggg	ccgcctggcg	ggccggcgac	7980
acgggtgtacg	ccgaggtggc	cctgcccgtc	ccgcagcccc	ggagcccgcg	cttcggtgtc	8040
caccocggcg	tgctcgacgc	ggcgttcag	gcgatgagcc	tcggcgcggt	cttccccgag	8100
gacgggcagg	tccggatgcc	gttcgccctg	cggggcggtg	cgtcgtccgg	ggtcggggcc	8160
gaccggctgc	gggtcaccat	cagcccggcc	ggtgccgagg	cggtccggat	cgcctgcgtc	8220
gacgagcggg	gcaaccgggt	cgtggtgatc	gactccctgg	tggcgcgcgc	ggtgccgggtg	8280
gaggcgctca	ccccggcac	ccccggcacc	ggggacggcg	cgctgcacca	cgtcgcctgg	8340
accgccgggc	cgggaaccggg	ggtcgccgcc	gtgcagcgct	gggcgggtcgt	gggcgcggcc	8400
gatccccggg	tggccggggg	cctggaccgg	gcgggcggcc	tctgcggggc	gtacccccgat	8460

ctgcgcgggtc tggtcgcgggc ggtggccgaa ggggcggcgc tgcccgcacgt ggtcgcgggtg 8520
 ccggtcccgt cgggcgcgcc ggtcggggccc gacgcgggtgc gcgccaccgt gtcggcgccc 8580
 ctggacctga tccgggcctg gtcgcgggtc gagggccggc tggggctggc caggctggcg 8640
 ttcgtcacca cctcggcggt ggcggtcggc gacggcaccg agcacgtgga cccggtgtcg 8700
 gccgccctgt gggggctggg gcgttcgcc cagtccgagg agcccggccg gttcgtcctc 8760
 gtcgacctgg acgccgaccc ggccagcgcc tcggccctgc ccgccgcgct cgcgcgggt 8820
 gagccgcaac tggccgttcg cgcggggcg gtgcacgtgc cccggctggg tcggcaccga 8880
 ccccgcccg acggcccgct gacgccccg gccggtgcc cgtggcggt cgcgcgggt 8940
 gggcagggca ccctggagg cctggcgctg gtcccggccc cggacgcctt ggcgccgctg 9000
 gccccgggc aggtccgggt cgcggtgcgc gccgccggag tgaacttcg ggacaccctc 9060
 atcgcgctcg gcatgtaccc gggcacgcc gtgctgggtg ccgagggggc cggggtgatc 9120
 accgaggtcg cgcggacgt ggccggcttc gccccggcg accgggtgct gggcatgtgg 9180
 accggcggcc tggggccggg ggcggtcgcc gacgcccgga tgctcgccg ggttcgcgc 9240
 ggctggctgt acgccgaggc cgcgtcgggt ccggccgtct tcctcacggc ccactacgcg 9300
 ctcaccaggc tcgccgggat ccgccgggg cagtgcgtgc tgggtgcacgc gggggccggc 9360
 ggcgtcggca tggcgaccct ccaactggcc cggcacctgg gcgtggagg ctacgccacg 9420
 gcgagccggg gcaagtggga caccctgcgt ggcctcggcc tggacgacgc gcacatcgcc 9480
 gactcccgca gcctcgactt cgcggacgg ttcttgccg ccaccggggg gcgcggcgtc 9540
 gacgtggtgc tgaactccct tgccggggac ttcgtggacg cgtccctgcg gctgctgccg 9600
 cgcggcggcc acttcttgga actgggcaag gccgacgtcc gcgacccga ccggtatcgcg 9660
 gccgaccacc cgggggtcgg ctaccgggcg ttcgacctcg tcgaggctgg tccggagctg 9720
 gtcgggcagc tgctcggcga gctgatggag ctgttcgcc cgggggtgct cagcccgctg 9780
 ccgttgaccg tcggggacgt ccggcgggcc cgggaggcgt tccgcctgat cagccaggcc 9840
 cggcacgtcg gcaagggtgt gctgaccatg ccgcccgct tcggcgcgta cggcaccgtc 9900
 ctggtcaccg gcggcacagg gacgctcggc ggcgccgtcg ccggcacct ggtcgcccg 9960
 cacggcgtag ggcacctgt gctcaccggc cgcagcggcc cggcggcgga cggggcgctc 10020
 gcgctcgtcg acgagctgac cgcgtccggc gcgtcggtag ccgtcgtcgc ctgcgacgcc 10080
 gccgaccggg tcgcgctgcg ccggctgctc gacggcattc cggccgcgca cccgctcacc 10140
 gccgtcgtgc acgtgccgg cgtcctcgac gacgccacca tcaccgcgct gaccgcccgg 10200

caggtggacg cgggtgctgcg gcccaaggcc gacgcggtga tcaacctgca cgagttgacc 10260
 cgggaccggg agctgtccgc gttcgtgctg ttctcctcgg cggcggccct gttcggcagc 10320
 ccggggcagg gcaactactc ggcgggccaac gggttcgtcg acgcgttcgc ccagtaccgc 10380
 cgcgcgcagg ggctccacgc ggtgtcgtcg gcctggggcc tgtgggcca cagcagccgg 10440
 atggccgggc acctcgacca ggaggggatg cggcgccgga tggcgcgcg cggcgtcctg 10500
 ccgctcacca ccgaccagg cctcgccctg ttcgacgcgc cgcagctggg ggacgaggcg 10560
 ctccaggtgc cgatccggct caacgtcggc gcgttgccgg cgcggggag ggtccccgcg 10620
 ctctcgccg acctggtgcc ggcggcggcg tcggggggccc cggccgccac cccgaccgg 10680
 gacgacgcg accgcacgct cgccgaccgc ctgcggggc tgaccgtggc cgaacagcg 10740
 gagctggtgc tggagagcgt gcgcggacac gcggcgccg tcctcggaca cgccgaccg 10800
 caggccgtcg acgccgaccg ggccttcgg gaactcggct tcgactcgt gacggcggtg 10860
 gagctgcgca atcggtggc caccgcgtcc gggctgcgcc tgccggcgac gctggtcttc 10920
 gaccaccca ccccggaagc gttggcggag cacctgctcg ccgggctcgc gcccgagcag 10980
 gccggggccg agttgccgtt gctggccgag ctgggcccgc tggaggcggc cctggccgcc 11040
 accgacgggg ccgccctcga cgggctggac gacctggtgc gccgggaggt cggcgctccg 11100
 atcgcggcgc tggccgccag gtggggcgcg gccggcgac acgtggccgg cagcgacggc 11160
 ggcgggacgg ccgacgcgct cgagtcgct gacgacgac agatcttcgc gttcatcgac 11220
 gagcggttcc gcgcctga 11238

<210> 16

<211> 1574

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 16

Met Ser Asn Glu Gln Lys Leu Arg Glu Tyr Leu Arg Leu Thr Thr Thr
1 5 10 15

Glu Leu Ala Arg Ala Thr Asp Arg Leu Arg Ala Val Glu Ala Arg Ala
20 25 30

His Glu Pro Ile Ala Ile Val Gly Met Ala Cys Arg Tyr Pro Gly Gly
35 40 45

Val Gly Ser Pro Glu Glu Leu Trp Glu Leu Val Ala Ser Gly Thr Asp
50 55 60

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

Pro Leu Trp Leu Gly Ser Leu Lys Ser Asn Ile Gly His Thr Gln Ala
 370 375 380
 Ala Ala Gly Val Gly Gly Val Ile Lys Thr Val Leu Ala Met Arg His
 385 390 395 400
 Gly Val Leu Pro Pro Thr Leu His Val Asp Ala Pro Thr Thr Glu Val
 405 410 415
 Asp Trp Ser Ala Gly Gln Val Ala Leu Leu Arg Ala Glu Thr Pro Trp
 420 425 430
 Pro Asp Thr Gly Arg Pro Arg Arg Ala Gly Val Ser Ser Phe Gly Val
 435 440 445
 Ser Gly Thr Asn Ala His Val Val Leu Glu Gln Ala Pro Gly Pro Ala
 450 455 460
 Ala Ala Pro Ala Gly Asp Ala Pro Pro Ala Glu Thr Arg Pro Val Gly
 465 470 475 480
 Asp Pro Pro Pro Val Val Pro Leu Val Leu Ser Ala Arg Ser Gln Pro
 485 490 495
 Ala Leu Ala Gly Gln Ala Arg Arg Leu Arg Asp Leu Leu Ala Ala Ala
 500 505 510
 Pro Glu Thr Asp Leu Ala Ser Ala Gly Leu Ala Leu Ala Thr Ala Arg
 515 520 525
 Ser Val Phe Asp His Arg Ala Val Val Thr Ala Ala Gly Arg Pro Gln
 530 535 540
 Ala Leu Asp Ala Leu Asp Leu Leu Ala Gly Gly Glu Pro Gly Pro Ala
 545 550 555 560
 Val Thr Thr Gly Val Ala Ala Pro Thr Gly Arg Thr Val Phe Val Phe
 565 570 575
 Pro Gly Gln Gly Thr His Trp Ala Gly Met Gly Ala Asp Leu Leu Asp
 580 585 590
 Gln Ser Pro Val Phe Ala Glu Ser Met Arg Arg Cys Glu Gln Ala Leu
 595 600 605
 Ser Ala His Thr Asp Trp Lys Leu Gly Glu Val Ile Arg Gly Ala Ala
 610 615 620
 Gly Ser Pro Pro Leu Asp Arg Val Asp Val Leu Gln Pro Val Ser Trp
 625 630 635 640
 Ala Val Met Val Ser Leu Ala Gln Val Trp Arg Ser Leu Gly Val Glu
 645 650 655
 Pro Asp Ala Val Val Gly His Ser Gln Gly Glu Ile Ala Ala Ala Val
 660 665 670

Val Cys Gly Ala Leu Thr Leu Pro Asp Ala Ala Arg Val Val Ala Leu
 675 680 685
 Arg Ser Gln Val Ile Gly Arg Val Leu Ser Gly Arg Gly Gly Met Ala
 690 695 700
 Ser Val Gln Leu Pro Ala Arg Glu Val Ala Gly Arg Leu Ala Ala Trp
 705 710 715 720
 Ala Gly Arg Leu Asp Val Ala Ala Val Asn Gly Pro Gln Ser Thr Val
 725 730 735
 Val Ser Gly Ala Ala Asp Ala Val Thr Glu Leu Val Glu Ala Phe Ala
 740 745 750
 Ala Glu Asp Val Arg Val Arg Arg Ile Pro Val Asp Tyr Ala Ser His
 755 760 765
 Ser Thr Gln Val Asp Arg Leu Arg Ala Glu Leu Leu Thr Val Leu Gly
 770 775 780
 Pro Val Asp Ala Arg Pro Ala Gln Val Pro Phe Tyr Ser Thr Val Gln
 785 790 795 800
 Gly Gly Arg Val Asp Thr Ala Gly Leu Asp Ala Gly Tyr Trp Tyr Arg
 805 810 815
 Asn Leu Arg Gly Gln Val Arg Phe Glu Glu Thr Val Arg Val Leu Leu
 820 825 830
 Asp Asp Gly His Arg Ala Phe Val Glu Ala Ala Ala His Ala Val Leu
 835 840 845
 Val Pro Ala Ile Gln Glu Leu Gly Asp Ser Ala Gly Val Arg Val Val
 850 855 860
 Ala Val Gly Ser Leu Arg Arg Glu Ala Gly Gly Leu Asp Arg Leu Leu
 865 870 875 880
 Ala Ser Ala Ala Glu Ala Phe Thr Gln Gly Val Ala Val Asp Trp Ser
 885 890 895
 Arg Ala Leu Ala Gly Ala Ala Arg Val Ala Val Asp Leu Pro Thr Tyr
 900 905 910
 Ala Phe Gln Arg Gln Arg Tyr Trp Leu Glu Pro Ala Ala Gln Ala Asp
 915 920 925
 Ser Gly Pro Ala Gly Asp Gly Trp Arg Tyr Arg Val Gly Trp Arg Arg
 930 935 940
 Leu Gln Arg Thr Gly Ala Ala Pro Ala Asp Arg Trp Leu Leu Val Thr
 945 950 955 960
 Gly Pro Glu Gln Pro Ala Glu Leu Val Glu Ala Val Arg Asp Ala Leu
 965 970 975

Thr Ala Arg Gly Ala Glu Val Arg Leu Val Thr Val Glu Pro Thr Ser
 980 985 990
 Thr Asp Arg Ala Ala Cys Ala Ala Leu Leu Thr Ala Ala Gly Ala Gly
 995 1000 1005
 Gly Ala Thr Arg Val Leu Ser Leu Leu Gly Thr Asp Arg Arg Pro
 1010 1015 1020
 His Pro Asp His Pro Ala Val Ser Val Gly Ala Ala Ala Thr Leu
 1025 1030 1035
 Leu Leu Thr Gln Ala Val Ala Asp Ala Leu Pro Ala Ala Arg Leu
 1040 1045 1050
 Trp Val Val Thr Arg Gly Ala Val Ser Val Gly Pro Gly Glu Thr
 1055 1060 1065
 Ala Asp Glu Arg Gln Ala Gln Val Trp Gly Phe Gly Arg Val Ala
 1070 1075 1080
 Ala Leu Glu Leu Pro Arg Thr Trp Gly Gly Leu Val Asp Leu Pro
 1085 1090 1095
 Ala Asp Ala Asp Gly Pro Val Trp Glu Ala Phe Val Asp Val Leu
 1100 1105 1110
 Ala Gly Asp Glu Asp Gln Val Ala Leu Arg Gly Pro Val Gly Tyr
 1115 1120 1125
 Gly Arg Arg Leu Arg Arg Ala Pro Ala Leu Pro Ala Lys Arg Arg
 1130 1135 1140
 Tyr Arg Pro Arg Gly Thr Val Leu Val Thr Gly Gly Thr Gly Ala
 1145 1150 1155
 Leu Gly Ala His Val Ala Arg Arg Leu Ala Ala Gly Gly Ala Ala
 1160 1165 1170
 His Leu Val Leu Thr Ser Arg Arg Gly Ala Asp Ala Pro Gly Ala
 1175 1180 1185
 Ala Gly Leu Val Gly Glu Leu Arg Ala Leu Gly Ala Glu Val Thr
 1190 1195 1200
 Val Ala Val Cys Asp Val Ala Asp Arg Ala Ala Val Ala Ala Leu
 1205 1210 1215
 Leu Ala Gly Leu Pro Ala Asp Ala Pro Leu Ser Ala Val Phe His
 1220 1225 1230
 Thr Ala Gly Val Ala His Ser Met Pro Ile Gly Glu Thr Gly Leu
 1235 1240 1245
 Thr Asp Val Ala Glu Val Phe Ala Gly Lys Val Ala Gly Ala Arg
 1250 1255 1260

His	Leu	Asp	Glu	Leu	Thr	Arg	Gly	His	Asp	Leu	Asp	Ala	Phe	Val
1265						1270					1275			
Leu	Tyr	Ser	Ser	Asn	Ala	Gly	Val	Trp	Gly	Ser	Ser	Gly	Gln	Ser
1280						1285					1290			
Ala	Tyr	Gly	Ala	Ala	Asn	Ala	Ala	Leu	Asp	Ala	Leu	Ala	Glu	Arg
1295						1300					1305			
Arg	Arg	Ala	Ala	Gly	Leu	Thr	Ala	Thr	Ser	Val	Ala	Trp	Gly	Leu
1310						1315					1320			
Trp	Gly	Ser	Gly	Gly	Met	Gly	Glu	Gly	Asp	Ala	Glu	Glu	Tyr	Leu
1325						1330					1335			
Ser	Arg	Arg	Gly	Leu	Arg	Pro	Met	Pro	Pro	Glu	Arg	Gly	Val	Asp
1340						1345					1350			
Ala	Leu	Leu	Ala	Ala	Leu	Asp	Arg	Asp	Glu	Thr	Phe	Val	Ala	Val
1355						1360					1365			
Ala	Asp	Val	Asp	Trp	Thr	Leu	Phe	Thr	Ala	Gly	Phe	Thr	Ala	Phe
1370						1375					1380			
Arg	Pro	Ser	Pro	Leu	Leu	Gly	Asp	Leu	Pro	Glu	Ala	Arg	Ala	Thr
1385						1390					1395			
Leu	Ala	Asp	Ala	Gly	Pro	Ala	Gly	Ser	Asp	Leu	Pro	Ala	Trp	His
1400						1405					1410			
Ala	Ala	Ala	Ser	Pro	Asp	Glu	Arg	Arg	Arg	Gly	Leu	Leu	Asp	Leu
1415						1420					1425			
Val	Arg	Arg	Gln	Val	Ala	Ala	Val	Leu	Gly	His	Pro	Gly	Pro	Glu
1430						1435					1440			
His	Val	Gly	Pro	Asp	Ala	Ala	Phe	Arg	Glu	Ile	Gly	Phe	Asp	Ser
1445						1450					1455			
Leu	Thr	Ala	Val	Asp	Leu	Ala	Lys	Arg	Leu	Arg	Ala	Ala	Val	Gly
1460						1465					1470			
Val	Pro	Leu	Ser	Ala	Thr	Leu	Val	Phe	Asp	His	Pro	Thr	Ala	Thr
1475						1480					1485			
Ala	Val	Ala	Glu	His	Leu	Ala	Gly	Leu	Leu	Gly	Pro	Ala	Pro	Ala
1490						1495					1500			
Gly	Gly	Asp	Pro	Arg	Glu	Ala	Glu	Val	Arg	Arg	Ala	Leu	Ala	Asp
1505						1510					1515			
Leu	Pro	Leu	Ala	Arg	Leu	Arg	Asp	Ala	Gly	Leu	Leu	Asp	Gly	Leu
1520						1525					1530			
Leu	Ala	Leu	Ala	Gly	Leu	Asp	Ala	Asp	Ala	Val	Pro	Asp	Gly	Pro
1535						1540					1545			

Glu Pro Ala Pro Gly Asp Ala Ile Asp Glu Leu Asp Pro Glu Glu
 1550 1555 1560

Leu Val Arg Arg Val Leu Asp Asn Ala Ser Ser
 1565 1570

<210> 17

<211> 4725

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 17

```

atgtcgaacg agcagaagct ccgcgagtag ctgcggttga ccaccaccga gctggccagg      60
gccaccgacc ggctgcgcgc ggtcgaggcg cgggcgcacg agccgatcgc gatcgtcggc      120
atggcctgcc ggtaccccg gggggtcggc tcaccggagg aactgtggga gctggtcgcc      180
tcgggcacgg acgcgatctc ccggttcccc gacgaccacg gctgggacgg cgacgcgctg      240
tacgaccggg acccgaggcg ggggggcgc acctactgcc gcgagggcg gttcctcgcc      300
ggggtcggcg acttcgacgc cgcgttcttc ggcatctcgc ccgcgaggc gctggccatg      360
gaccgcgagc agcgccctgct gctggagacg tcctgggagg cgctggagcg ggccgggacg      420
ccccgggact cgctgcgcgg cagccgtacc ggggtgtgcg tcggggcggt gcacggcggc      480
tacaccgacg tcgtcgggca gccccggcg gaactggagg gccacctgct gaccggcggg      540
gtggtcagct tcacctggg gcgatctcg tacgcgctgg gcctggaggg gcccgcgctg      600
acggtggaca ccgcctgctc gtcctcgctg gtggccctgc acctggcggt gcgggcccctg      660
cggcagggcg agtgcgacct ggcgttggcc ggcggggcga cgggtgctggc cagcccggcg      720
gtgttcgtgc agttctcgcg gcagcggggg ctggccccgg acggccggtg caaggcgctc      780
gccgactcgg cggacggggt cgggccggcc gagggggtcg gcatgctggt cgtggagcgg      840
ctgtcggacg ccgtccgcca cgggcgcgg gtgctggccc tggtaaccgg cacggcggtc      900
aaccaggacg gggcgagcaa cggcctcacc gccccagcg gccggcgca ggagaagggtg      960
ctgcgccagg cgctcgtgga cggccgggtg acggccgccg acgtcgacgc ggtcgaggcg      1020
cacggcaccg gcaccgggct cggcgacccg atcgagggtg gggccctgat gaacgtgtac      1080
ggtgccggcc ggcccgccga ccgtccgctc tggctcggtt cgtgaagtc caacatcggc      1140
cacaccagg cggcgggcgg ggtcggcggg gtcacaaaga cgggtgctggc gatgcggcac      1200
ggcgtcctgc cggccacct gcacgtggac gccccacca ccgaggtcga ctggtcgcc      1260
ggccagggtg ccctgctgcg ggcagagaca ccgtggccgg acacgggtcg cccgcgcgc      1320
gccggggtct cctccttcgg ggtgagcggc accaacgcgc acgtggtgct ggagcaggcc      1380

```

cctgggcccc	ccgccgcccc	ggcgggtgac	gccccgcccc	ccgagacccc	gcccgtcggc	1440
gacccgccgc	cggtcgtacc	gctggtgttg	tccgccaggt	cgcagccggc	gctggccggg	1500
caggccccgc	ggctgcgcga	cctgctggcc	gcagcgccgg	agaccgacct	cgccagcgcc	1560
ggactcgcgc	tggccaccgc	gcggtcgggt	ttcgaccacc	gggcggtggt	gacggccgcc	1620
gggcgaccgc	aggcgctcga	cgcgctcgac	ctgctggccg	gcggcgaacc	cggaccggcg	1680
gtcacgaccg	gcgtcgccgc	ccccaccggg	cgcaccgtgt	tcgtctttcc	cgggcagggg	1740
acgcactggg	ccggcatggg	tgccgacctg	ctcgaccagt	caccggtgtt	cgcagagtgc	1800
atgcgacggt	gcgagcaggc	gctgtcggcg	cacaccgact	ggaagctcgg	cgaggtgata	1860
cggggcgccg	ccggcagccc	gccgctggac	cgcgtggacg	tgctccagcc	cgtctcctgg	1920
gcggtgatgg	tgtcgctggc	gcaggtgtgg	cggtcgctcg	gcgtcgagcc	ggacgcgggt	1980
gtcggccatt	cccagggcga	gatcgccgcc	gcggtggtct	gcggcgcgct	gacctgccg	2040
gacgcggccc	gggtggtcgc	gctgcgggcc	caggtcatcg	gtcgggtgct	ctccggtcgc	2100
ggcggcatgg	cgtccgtcca	gctgccggcc	cgggaggtcg	cgggcgggct	ggccgcctgg	2160
gcggggccgg	tcgacgtcgc	ggcgtcaaac	gggccacagt	cgaccgtcgt	gtccggtgcc	2220
gccgacgcgg	tcaccgaact	ggtcgaggcg	ttcgcgcccg	aggacgtccg	ggtgcggcgg	2280
atcccgggtg	actacgcgtc	ccactcgacg	caggtggacc	ggctgcgcgc	cagactgctc	2340
accgtcctgg	gcccggtcga	cgcccgctcc	gcgcaggtgc	ccttctactc	gacggtgcag	2400
ggcgggcgcg	tcgacactgc	cggcctggac	gccggctact	ggtaccgcaa	cctgcggggg	2460
caggtccgct	tcgaggagac	cgtgcgggtg	ctgctcgacg	acgggcaccg	cgccttcgtc	2520
gaggccgccg	cgcacgcctg	cctcgtaacc	gcgatccagg	agctggggga	cagcgccggc	2580
gtccgggtgg	tggccgtggg	gtcgctgcgc	cgggagggcg	gcggcctgga	ccggctcctg	2640
gcctcggcgg	ccgaggcggt	caccagggg	gtggccgtgg	actggtccc	ggctctggcc	2700
ggggccgcgc	gcgtcgccgt	ggacctgccc	acgtacgcgt	tcacagcgga	acgtactgg	2760
ctggagcccc	ccgcgcaggc	ggactccggc	ccggccgggg	acggctggcg	ctaccgggtc	2820
ggctggcggc	ggcttcagcg	caccggcgcc	gcgccggccg	accggtggct	gctggtgacc	2880
ggcccggagc	agccggcgga	gctggtcgag	gcggtgcgcg	acgcgctcac	cgcgcggggc	2940
gccgaggtgc	gcctggtgac	cgtcgagccg	accagcaccg	accgggccc	gtgcgcggcg	3000
ttgctcaccg	cggccgggtgc	gggcggggcg	accgggtgct	tgctcgctgct	cggcaccgat	3060

cgtcgccccg accccgacca cccggccgtg tccgtcggcg ccgcccgcgac gttgctgctg	3120
accagggccg tcgcccacgc cctgccggcc gcccggtctgt gggtcgtcac ccggggcgcg	3180
gtctccgtcg ggcccggcga gaccgcccac gagcgccagg cgcaggctctg ggggttcggc	3240
cgggtcgcgg ccctcgaaact gcccgcacg tggggcgggc tcgtcgacct gcccgccgac	3300
gcggacggcc cgggtgtggga ggcgttcgtg gacgtgctgg ccggggacga ggaccaggtc	3360
gcgctgcgcg gcccggtcgg gtacggtcgc cggctccggc gcgccccgc gctaccgcg	3420
aagcggcggt accggcccag gggcacgctc ctggtcaccg gcggcacccg cgcgctcggc	3480
gcgcacgtgg cccggcggtt ggccgcccgc ggggcgcgc acctcgtgct caccagccgg	3540
cgcggggccg acgcccccg tgcggccggg ctggtcgggg aactccgggc gctgggcgc	3600
gaggtgaccg tcgcggtctg cgacgtcgcc gaccgggccc ccgtggcggc gctgctcgcc	3660
gggctgcccg ccgacgcgc gctgagcgc gtcttcacac ccgcgggcgt ggcgcactcg	3720
atgccgatcg gcgagaccg gctcaccgac gtcgcccagg tgttcgccgg gaaggtcgcc	3780
ggagcccgc acctcgacga actcaccgg gggcacgacc tggacgcgtt cgtcctgtac	3840
tcgtcgaaac cgggcgtgtg gggcagcagc gggcagagcg cgtacggggc ggccaacgcg	3900
gccctcgacg cgctcgccga acggcggcgc gccgcgggc tgaccgccac ctccgtcgcc	3960
tggggcctgt ggggctccgg gggcatgggc gagggcgacg ccgaggagta cctgagccgc	4020
cggggcctgc ggccgatgcc tcccagcgt ggcgtggacg ccctcctggc cgccctggac	4080
cgggacgaga ccttcgtcgc cgtcgccgac gtggactgga cgctgttcac ggccgggttc	4140
accgcgttcc ggcccagccc gctgctcggc gacctcccgg agggccgcgc gacgctggcc	4200
gacgcgggac ccgcgggctc cgacctgccg gcctggcacg ccgcccgcag ccccgacgaa	4260
cgccgcgggg gcctgctcga cctggtacgc cggcaggtcg ccgccgtcct cgccacccg	4320
gggcccagac acgtcgccc cgacgccgcg ttccgggaga tcggattcga ctcgctgacc	4380
gccgtcgacc tggccaagcg gtcagggcg gcggtcggcg tgccgctgtc cgccaccctc	4440
gtcttcgacc accccaccgc gacggcggtc gccgagcacc tggccgggct gctcggctcc	4500
gcgcccggcg gcggcgaccc gcgcgaggcc gaggtgcgcc gggccctggc cgacctgcg	4560
ctggcccggc tcgggacgc cggcctactg gacggcctgc ttgcgcttgc ggggctggac	4620
gccgacgcgg tgccggacgg gcccgagccg gctcccggcg acgccatcga cgaactcgat	4680
ccagaggagc tgggtgcgcc ggtgctggac aacgccagct cctga	4725

<210> 18
 <211> 1784
 <212> PRT
 <213> micromonospora carbonacea subspecies aurantiaca

 <400> 18

 Met Val Met Pro Pro Asp Lys Val Ile Glu Ala Leu Arg Val Ser Val
 1 5 10 15

 Lys Glu Thr Glu Arg Leu Arg Arg Gln Asn His Glu Leu Leu Ala Ala
 20 25 30

 Leu His Gly Pro Ile Ala Val Val Gly Met Ala Cys Arg Tyr Pro Gly
 35 40 45

 Gly Val Ser Ser Pro Glu Asp Leu Trp Arg Leu Val Glu Thr Gly Thr
 50 55 60

 Asp Ala Ile Gly Gly Phe Pro Thr Asp Arg Gly Trp Asp Val Asp Ala
 65 70 75 80

 Val Tyr Asp Pro Asp Pro Glu Ser Arg Asn Thr Thr Tyr Cys Arg Glu
 85 90 95

 Gly Gly Phe Leu Ala Gly Ala Gly Asp Phe Asp Ala Ala Phe Phe Gly
 100 105 110

 Val Ser Pro His Glu Ala Val Val Met Asp Pro Gln Gln Arg Leu Leu
 115 120 125

 Leu Glu Val Ser Trp Glu Ala Leu Glu Arg Ser Gly Thr Asp Pro His
 130 135 140

 Ser Leu Arg Gly Ser Arg Thr Gly Val Tyr Val Gly Ala Ala His Gln
 145 150 155 160

 Gly Tyr Ala Val Asp Ala Gly Gln Val Pro Glu Gly Ala Glu Gly Phe
 165 170 175

 Arg Leu Thr Gly Ser Ala Asp Ala Val Leu Ser Gly Arg Ile Ser Tyr
 180 185 190

 Leu Leu Gly Leu Glu Gly Pro Ala Leu Thr Val Glu Thr Ala Cys Ser
 195 200 205

 Ser Ser Leu Val Ala Val His Leu Ala Val Gln Ala Leu Arg Arg Gly
 210 215 220

 Glu Cys Gly Leu Ala Leu Ala Gly Gly Val Ala Val Met Pro Asp Pro
 225 230 235 240

 Ala Ala Phe Val Glu Phe Ser Arg Gln Arg Gly Leu Ala Ala Asp Gly
 245 250 255

 Arg Cys Arg Ala Phe Gly Ala Gly Ala Asp Gly Thr Gly Trp Ala Glu
 260 265 270

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

Arg Val Gly Met Ala Arg Glu Leu Tyr Glu Ala Phe Pro Ala Phe Ala
 580 585 590
 Asp Ala Leu Asp Gln Leu Ala Ala Arg Leu Asp Arg His Leu Asp Arg
 595 600 605
 Pro Leu Leu Arg Val Leu Phe Ala Glu Pro Gly Ser Asp Asp Ala Arg
 610 615 620
 Leu Leu Asp Gly Thr Arg Tyr Ala Gln Ala Ala Leu Phe Ala Val Glu
 625 630 635 640
 Val Ala Leu Phe Arg Leu Val His Gly Trp Gly Val Arg Pro Asp Val
 645 650 655
 Leu Leu Gly His Ser Val Gly Glu Leu Ala Ala Ala His Val Ala Gly
 660 665 670
 Val Leu Asp Val Asp Asp Ala Cys Glu Leu Val Ala Ala Arg Gly Arg
 675 680 685
 Leu Met Gly Glu Leu Pro Ser Gly Gly Ala Met Val Ala Val Arg Ala
 690 695 700
 Thr Glu Glu Glu Val Gly Pro Leu Leu Asp Gly Gln Arg Val Ala Val
 705 710 715 720
 Ala Ala Val Asn Gly Pro Arg Ser Val Val Val Ser Gly Asp Glu Glu
 725 730 735
 Ala Val Leu Ala Val Ala Ala Arg Cys Ala Ala Leu Gly His Arg Thr
 740 745 750
 Arg Arg Leu Asn Val Ser His Ala Phe His Ser Pro His Val Glu Ala
 755 760 765
 Met Leu Glu Pro Phe Arg Arg Val Ala Arg Gly Leu Thr Tyr His Ala
 770 775 780
 Pro Thr Ile Pro Val Val Ser Asn Ala Thr Gly Arg Leu Ala Thr Ala
 785 790 795 800
 Asp Ala Leu Arg Asp Pro Gly Tyr Trp Val Arg His Val Arg Gln Pro
 805 810 815
 Val Arg Phe Arg Asp Gly Val Arg Ala Ala Arg Asp Gln Gly Ala Thr
 820 825 830
 Ala Phe Val Gly Leu Gly Pro Asp Gly Val Leu Cys Ala Leu Ala Glu
 835 840 845
 Glu Cys Leu Gly Pro Thr Gly Asp Val Leu Leu Leu Pro Val Leu Arg
 850 855 860
 Pro Gly Arg Pro Glu Pro Ala Thr Leu Leu Ala Ala Leu Ala Gly Ala
 865 870 875 880

Tyr Ala Gly Gly Ala Glu Met Asp Trp Ser Arg Val Phe Ala Gly Thr
 885 890 895
 Gly Ala Arg Arg Val Glu Leu Pro Thr Tyr Ala Phe Gln His Arg Arg
 900 905 910
 Tyr Trp Leu Ala Pro Gly Pro Pro Ser Ala Arg Arg Asp Asp Ala Trp
 915 920 925
 Arg Tyr Arg Ile Ala Trp Arg Pro Leu Pro Thr Val Pro Ala Ala Ala
 930 935 940
 Gly Thr Glu Thr Val Ala Gly Ala Trp Leu Leu Val Val Pro Ala His
 945 950 955 960
 Asp Gly Val Ala Ser Leu Ala Asp Ala Ala Glu Arg Ala Val His Arg
 965 970 975
 Gly Gly Ala Thr Val Thr Arg Leu Thr Val Asp Ala Ala Asp Val Asp
 980 985 990
 Arg Asp Thr Leu Ala Ala Val Leu Thr Glu Ala Ala Ala Asp Ala Asp
 995 1000 1005
 Gly Gly Pro Asp Gly Val Leu Cys Leu Leu Gly Leu Asp Asp Arg
 1010 1015 1020
 Ala His Pro Arg Ser Ala Ser Val Pro Arg Gly Val Leu Ala Thr
 1025 1030 1035
 Leu Ser Leu Ala Gln Ala Leu Thr Asp Leu Gly Ala Ser Ala Arg
 1040 1045 1050
 Leu Trp Cys Val Thr Arg Gly Ala Val Ala Val Thr Pro Gly Glu
 1055 1060 1065
 Ser Pro Ser Val Ala Gly Ala Gln Leu Trp Gly Phe Gly Arg Val
 1070 1075 1080
 Ala Ala Leu Glu Leu Pro Arg Ser Trp Gly Gly Leu Val Asp Leu
 1085 1090 1095
 Pro Val Asp Pro Asp Asp Arg Asp Trp Asp Leu Leu Arg Arg Ala
 1100 1105 1110
 Leu Arg Gly Pro Glu Asp Gln Val Ala Val Arg Gly Ala Val Gly
 1115 1120 1125
 Tyr Ala Arg Arg Leu Val Pro Ala Pro Ala Pro Arg Ala Glu Arg
 1130 1135 1140
 Ala Trp Arg Pro Arg Gly Thr Val Leu Val Thr Gly Gly Thr Gly
 1145 1150 1155
 Ala Leu Gly Ala His Thr Ala Arg Trp Leu Ala Arg Asn Gly Ala
 1160 1165 1170

Thr	His	Leu	Val	Leu	Thr	Ser	Arg	Arg	Gly	Gly	Asn	Ala	Pro	Gly
1175						1180					1185			
Val	Ala	Ala	Leu	Arg	Ala	Glu	Leu	Val	Thr	Leu	Gly	Ala	Glu	Val
1190						1195					1200			
Thr	Val	Val	Ala	Cys	Asp	Val	Ala	Asp	Arg	Glu	Ala	Val	Ala	Gly
1205						1210					1215			
Leu	Leu	Ala	Gly	Ile	Pro	Arg	Ala	Ala	Pro	Leu	Thr	Ala	Val	Phe
1220						1225					1230			
His	Ala	Ala	Gly	Val	Pro	Gln	Val	Thr	Pro	Leu	His	Glu	Thr	Thr
1235						1240					1245			
Pro	Glu	Leu	Phe	Ala	Gln	Val	Cys	Ala	Gly	Lys	Val	Ala	Gly	Ala
1250						1255					1260			
Val	His	Leu	His	Glu	Leu	Ala	Gly	Asp	Leu	Asp	Ala	Phe	Val	Thr
1265						1270					1275			
Phe	Ala	Ser	Ala	Ala	Gly	Val	Trp	Gly	Ser	Gly	Gly	Gln	Cys	Ala
1280						1285					1290			
Tyr	Ala	Ala	Ala	Asn	Ala	Ala	Leu	Asp	Ala	Leu	Ala	Glu	Arg	Arg
1295						1300					1305			
Arg	Ala	Ala	Gly	Leu	Pro	Ala	Thr	Ser	Val	Ala	Trp	Gly	Val	Trp
1310						1315					1320			
Gly	Gly	Pro	Gly	Met	Gly	Ala	Gly	Ala	Gly	Glu	Glu	Tyr	Leu	Arg
1325						1330					1335			
Arg	Arg	Gly	Val	Arg	Ala	Met	Pro	Pro	Ala	Ala	Ala	Leu	Ala	Ala
1340						1345					1350			
Leu	Gly	Arg	Ile	Leu	Asp	Ala	Asp	Glu	Thr	Gly	Val	Thr	Val	Ser
1355						1360					1365			
Asp	Thr	Glu	Trp	Gly	Arg	Phe	Ala	Ser	Gly	Phe	Ala	Ala	Ala	Arg
1370						1375					1380			
Pro	Ala	Pro	Leu	Leu	Ala	Glu	Leu	Pro	Gly	Gly	Asp	Val	Asp	Pro
1385						1390					1395			
Ala	Gly	Pro	Ala	His	Arg	Ala	Gln	Pro	Pro	Val	Pro	Arg	Pro	Ala
1400						1405					1410			
Pro	Ala	Ala	Thr	Asp	Arg	Pro	Gly	Leu	Leu	Ala	Leu	Val	Arg	Ala
1415						1420					1425			
Glu	Ala	Ala	Gly	Val	Leu	Gly	His	Asp	Gly	Ala	Asp	Asp	Val	Pro
1430						1435					1440			
Ala	Asp	Ala	Glu	Phe	Ser	Ala	Leu	Gly	Phe	Asp	Ser	Leu	Ala	Ala
1445						1450					1455			

Val	Gln	Leu	Arg	Arg	Arg	Leu	Ala	Glu	Ala	Thr	Gly	Leu	Ser	Leu
1460						1465					1470			
Ser	Ala	Pro	Val	Leu	Phe	Asp	His	Arg	Thr	Pro	Asp	Ala	Leu	Ala
1475						1480					1485			
Ala	His	Leu	His	Gly	Leu	Leu	Thr	Gly	Ala	Ala	Gly	Gly	Pro	Pro
1490						1495					1500			
Ala	Pro	Ala	Ala	Gly	Ser	Ala	Leu	Val	Glu	Met	Tyr	Arg	Arg	Ala
1505						1510					1515			
Val	Ala	Thr	Gly	Arg	Ala	Ala	Glu	Ala	Val	Glu	Val	Leu	Gly	Thr
1520						1525					1530			
Val	Ala	Thr	Phe	Arg	Pro	Val	Phe	Arg	Ser	Pro	Asp	Glu	Leu	Gly
1535						1540					1545			
Glu	Pro	Pro	Ala	Leu	Val	Pro	Leu	Gly	Thr	Gly	Ala	Gly	Gly	Pro
1550						1555					1560			
Ala	Leu	Val	Cys	Cys	Ala	Gly	Thr	Ala	Ala	Ala	Ser	Gly	Pro	Arg
1565						1570					1575			
Glu	Phe	Thr	Ala	Phe	Ala	Ala	Ala	Leu	Ala	Gly	Leu	Arg	Asp	Val
1580						1585					1590			
Thr	Val	Leu	Pro	Gln	Thr	Gly	Phe	Leu	Pro	Gly	Glu	Pro	Leu	Pro
1595						1600					1605			
Ala	Gly	Leu	Asp	Val	Leu	Leu	Asp	Ala	Gln	Ala	Asp	Ala	Val	Leu
1610						1615					1620			
Ala	His	Cys	Ala	Gly	Gly	Pro	Phe	Val	Leu	Val	Gly	His	Ser	Ala
1625						1630					1635			
Gly	Ala	Asn	Met	Ala	His	Ala	Leu	Thr	Val	Arg	Leu	Glu	Ala	Arg
1640						1645					1650			
Gly	Ala	Asp	Pro	Ala	Ala	Leu	Val	Leu	Met	Asp	Ile	Tyr	Thr	Pro
1655						1660					1665			
Ala	Ala	Pro	Gly	Ala	Met	Gly	Val	Trp	Arg	Glu	Glu	Met	Leu	Ala
1670						1675					1680			
Trp	Val	Ala	Glu	Arg	Ser	Val	Val	Pro	Val	Asp	Asp	Thr	Arg	Leu
1685						1690					1695			
Thr	Ala	Met	Gly	Ala	Tyr	His	Arg	Leu	Leu	Leu	Asp	Trp	Ala	Pro
1700						1705					1710			
Arg	Pro	Thr	Arg	Ala	Pro	Val	Leu	His	Leu	Tyr	Ala	Gly	Glu	Pro
1715						1720					1725			
Ala	Gly	Ala	Trp	Pro	Asp	Pro	Arg	Gln	Asp	Trp	Arg	Ser	Arg	Phe
1730						1735					1740			

Asp Gly Ala His Thr Ser Ala Glu Val Pro Gly Thr His Phe Ser
 1745 1750 1755

Met Met Thr Glu His Ala Pro Val Thr Ala Ala Thr Val His Lys
 1760 1765 1770

Trp Leu Asp Glu Val Cys Pro Pro Arg Val Pro
 1775 1780

<210> 19

<211> 5355

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 19

```

atggtcatgc cccccgacaa ggtgatcgag gcgctgcgtg tctccgtcaa ggagacggag      60
cggctgcgcc ggcagaacca cgagctgctc gccgccctgc acgggccgat cgccgtcgtg      120
ggcatggcct gccgctaccc gggcggggtg tcctctccgg aggacctgtg gcggtgggtc      180
gagacgggca cggacgcgat cggcggttc cccaccgacc gtggctggga cgtcgacgcc      240
gtgtacgacc cggatcctga gtgcggaac accacctact gccgggaggg cgggttcctg      300
gccggggcag gagacttcga cgccgcgttc ttcggggtgt cgccgcacga ggccgtggtc      360
atggaccccc agcagcggct gcttctggag gtgtcctggg aggcgctgga gcggtccggg      420
accgaccgcg acagcctgcg cggctcgcgc accggggtct acgtcgggtg gggccaccag      480
gggtacgcgg tcgacgccgg tcaggtgccg gaggcgcgcg aggggttcog gctgaccggc      540
agcggcgacg ccgtcctgtc cggacggatc tcgtacctgc tcgggctgga gggtcgggcc      600
ctgaccgtcg agacggcctg ctgctcctcg ctggtggcgg tgcacctcgc ggtgcaggcg      660
ctgcgccggg gcgagtgcgg gctggcactg gccggcgggg tcgcogtgat gcccgaaccg      720
gcggcattcg tggagttctc ccggcagcgg ggcctcgcgg cggacggggc ctgccgggcg      780
ttcggggcgg gcgcggacgg caccggctgg gcggagggcg tcggtgtgct ggtcctgcaa      840
cggctctccg acgcggtgcg cgacggccgc tgggtgctgg gcgtgatccg gggttcggcc      900
gtcaaccagg acggggccag caacgggctg accgccccga gcggccccgc ccagcagcgg      960
gtcatccggc aggcgctgac cgacggccgg ctcggcgcgg accagatcga cgcggtcgag     1020
gcgcacggca cgggcacccg gctcggcgac ccgatcgagg cgcaggcgct gatcgccgcc     1080
tacggcgccg accggacccc ggaccggccg ctctggctcg gctcgttgaa gtcgaacatc     1140
gggcacgccc agggggcggc cggcgtcggc ggcctgatac agatgctcct ggcgatgcgg     1200
gccgggaacg tcccaccac cctgcacgcc gacgtccga ccccgctggt cgactggtcc     1260

```

gccgggtgtcg tccgggtgtc gaccgggggtg gtgccctggc ccgcgttgcc cggggcgccc 1320
 cgcagggccg ggatctccgc gttcgggggtg agcggcacca acgcgcacgt gatcgctgag 1380
 cagccgcgcg cggccccggt cgacgacccg gcgccacca cgaggaccct gccgctgggtg 1440
 ccgtgggtgc tctccggccg gacggaggcg gcgctgcgcg cccaggcgga ccggttgctg 1500
 acgcacctgg cggcgacccc cgacgcggac ccgctggacg tgggattctc cctggccacc 1560
 agccggggccg cgctggagca ccggggccgtg ctggtggccg ccgaccgcga cggcctgctc 1620
 cgctcgctcg acgcgtggc cgccggcgag ccggcgggcg gcctgatccg gggcacggta 1680
 cgtcacgacg gccggaccgg gttcctcttc gccgggcagg gcggccagcg cgtcgggatg 1740
 gcgcgcgaac tgtacgaggc gttccccgcc ttcgccgacg ccctggacca gctgcgcgcc 1800
 cggctggacc ggcacctcga tcgtccgctg ctgcgggtgc tggtcgccga gccggggctc 1860
 gacgacgccc ggctgctcga cggcaccggg tacgcgcagg ccgccctctt cgcgctcgag 1920
 gtggcggtgt tccgactggt ccacggctgg ggggtccggc ccgacgtgct gctcggccac 1980
 tcggtgggcg agctggcggc cgcgcacgtg gccggcgtag tcgacgtgga cgacgcgtgc 2040
 gagctggctg cggcgcgggg ccggctgatg ggggagctgc cgtcgggcgg cgcgatgggtg 2100
 gccgtccggg ccaccgagga ggaggtcggg cccctgctcg acgggcagcg ggtcgcgggtg 2160
 gccggcggtca acggcccgcg ctcggtcgctg gtctccggcg acgaggaggc ggtgctggcc 2220
 gtggccgccc ggtgcgcgcg cctcggccac cggacgcgac gcctcaacgt cagccacgcg 2280
 ttccactccc cgcacgtgga ggcgatgctg gagccgttcc ggcgggtggc gcggggcctg 2340
 acgtaccatg ccccgacgat cccggtgggtg tcgaacgcga cgggccggct cgccaccgcc 2400
 gacgcgctgc gcgaccccg ttactgggtc cggcacgtcc gccagccgt ccggttccgg 2460
 gacgggggtg gggccgccc cgaccagggg gccaccgcct tcgtcgggct cggcccggac 2520
 ggggtgctgt gcgcgttggc cgaggagtgc ctcgggccca ccggcgacgt gctgctgctg 2580
 ccggtgctgc gcccggctg gccggagccc gccaccctgc tggccgccct ggccggggcg 2640
 tacgcggcg gcgcggaaat ggactggtcc cgggtgttcg cgggcaccgg cgcgcgcagg 2700
 gtcgagctgc ccacgtacgc cttccagcac cggcgctact ggctggcgcc gggcccgcg 2760
 tcggcccgcg gcgacgacg ctggcggtac cggatcgctt ggcgccccct gccgaccgtg 2820
 cccgccgcg ccgggaccga gacggtggcc ggggcgtggt tgctggtggt cccgcgccac 2880
 gacggcgctg cgtcgtcgc cgacgcgcgc gagcggggcg tgcaccgggg cggggccacg 2940

gtcaccg	ggc	tgacggtgga	cgccgccc	gac	gtggaccggg	acaccctcgc	cgccgtgctg	3000
accgaggccg	ccgccc	gacgc	ggacggcggg	ccggacgggg	tgctctgcct	gctgggcctc	3060	
gacgaccggg	cacatccccg	gtccgcctcg	gtgccccgcg	gggtgctggc	gaccctgtcc	3120		
ctcgcccagg	ccctgaccga	cctggggggc	tccgcgcggc	tgtggtgagt	gaccggggg	3180		
gcggtcgccg	tgacgcccgg	cgagtccccg	tcggtcgccg	gagcccagtt	gtggggcttc	3240		
ggccgcgtgg	ccgcgctcga	actcccccg	tcctggggcg	gcctggtgga	cctgccggtc	3300		
gaccgggacg	accgggactg	ggacctgctg	cggcgcgcg	tcgcggccc	ggaggaccag	3360		
gtcgcggtcc	ggggggcggt	cgggtacgcc	cggcggtg	ttccgcgc	cgcccccg	3420		
gccgagcggg	cctggcgctc	gcgcggcacg	gtcctggtga	ccggcggtac	gggcgcgctc	3480		
ggcgcgca	cggcccgtg	gctggcgcg	aacggcgcca	cgcacctcgt	cctcaccagc	3540		
cgccggggcg	ggaacgcccc	cggggtcgcc	gcgctcgggg	cggaactggt	cacgctcggt	3600		
gccgaggtga	ccgtggtcgc	ctgcgacgtc	gccgaccggg	aggccgtggc	cggcctgctc	3660		
gccgggattc	cccgcgcgc	tcgctcacc	gccgtgttcc	acgcggcggg	cgtgccccag	3720		
gtgacgccgc	ttcacgagac	gaccccgag	ttgttcgcgc	aggtctgcgc	aggcaaggtc	3780		
gccggggcg	tgacacctga	cgagttggcc	ggtgacctg	acgccttcgt	caccttcgcc	3840		
tccgcccgcg	gggtgtgggg	cagcggcggg	cagtgcgcgt	acgctgcggc	caacgcgcgc	3900		
ctcgacgcgc	tcgcccagcg	tcgtcgcgcc	gcagggctgc	ccgcgacctc	cgtcgctgg	3960		
ggggtctggg	gcgggcccgg	catggggggc	ggcgcggggg	aggagtacct	gcgcgcgcg	4020		
ggcgtccggg	cgatgcccc	ggcagccgcc	ctcgccgccc	tcgggcggat	cctggacgcc	4080		
gacgagaccg	gggtgacggt	ctccgacacc	gagtggggcc	ggttcgcgtc	cggcttcgcc	4140		
gccgcgcgtc	ccgccccgct	gctcgccgag	ctgcggggcg	gggacgtcga	tcgggcccgc	4200		
ccggcgacc	gggcgcagcc	gcccgtgccc	cgaccggccc	cggcagccac	cgaccgcccc	4260		
gggctgctgg	cgctggtccg	cgccgaggcc	gccgggggtgc	tggggcacga	cgggtgccgac	4320		
gacgttcggg	ccgacgcgga	gttctccgcc	ctcggttcg	actcgctcgc	cgcgctccag	4380		
ctgcgcgcgc	ggctcgccga	ggccaccggc	ctgagcctct	cggccccggg	tctgttcgac	4440		
caccgcaccc	ctgacgcgct	cgccgcgcac	ctgcacggcc	tgctcaccgg	cgcggcgggc	4500		
gggcccacccg	cgccggccgc	cgggagcgcc	ctggtcgaga	tgtaccggcg	ggccgtcgcc	4560		
accggccgcg	ccgcccaggc	ggtggagggtg	ctcggcaccg	tcgccacgtt	ccggccgggtg	4620		
ttccggtccc	cggacgaact	gggcgagcca	ccggccctcg	ttccgctcgg	caccggggcg	4680		

```

gggggacccg cgctggtctg ctgcgcgggc acggccgcgg cgtccggccc ccgcgagttc 4740
acggcggttcg ccgcgcgcgt ggccgggtctc cgggacgtca ccgtccttcc gcagaccggc 4800
ttcctgcccg gcgagccgct gcccgccggg ctggacgtgc tgctcgacgc ccaggccgac 4860
gccgtcctgg cccactgogc cgggggaccc ttcgtcctgg tcggccactc ggccggggcg 4920
aacatggcgc acgcgctgac ggtccgcctg gaggcgcggg gcgcggaccc cgccgcgctg 4980
gtgctgatgg acatctacac gcccgccgcc ccgggggcga tgggggtgtg gcgcgaggag 5040
atgctggcct gggtcgccga gcgggtccgtc gtccccgtcg acgacacgcg gctgaccgcg 5100
atgggcgcct atcaccggct gtcctggac tgggcgcccc ggccgacccg ggcacccgtg 5160
ctgcacctgt atgccggtga accggcgggc gcctggccgg atccccggca ggactggcgt 5220
tcgcgcttcg acggcgcgca caccagcgcc gaggtgcccg gcacccactt ctcatgatg 5280
accgagcacg ccccgctcac cgccgcgacc gtgcacaagt ggctcgacga ggtgtgcccg 5340
ccccgcgttc cgtga 5355

```

<210> 20

<211> 464

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 20

```

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

```

Glu Ala Arg Arg Arg Leu Asp Ala Leu Ala Gly Phe Gly Gly Pro Val
 130 135 140
 Leu Arg Ala Asp Ala Ala Gly Ala Arg Thr Arg Val Val Glu Thr Thr
 145 150 155 160
 Ala Val Leu Leu Asp Gly Ile Gly Glu Arg Phe Asp Leu Ala Val Leu
 165 170 175
 Ala Arg Arg Leu Val Ala Ala Val Leu Ala Asp Leu Leu Gly Val Pro
 180 185 190
 Ala Ala Arg Arg Gly Arg Phe Ala Glu Ala Leu Ala Ala Ala Gly Arg
 195 200 205
 Thr Leu Asp Ser Arg Leu Cys Pro Gln Thr Val Ala Thr Ala Leu Ala
 210 215 220
 Thr Val Ala Ala Thr Ala Glu Leu Thr Asp Leu Leu Gly Glu Val Pro
 225 230 235 240
 Pro Pro Pro Ser Leu Ser Pro Ser Ala Ala Gly Ser Gly Pro Pro Arg
 245 250 255
 Pro Ser Ala Ala Gly Ser Trp Pro Pro Leu Pro Ala Asp Asp Arg Thr
 260 265 270
 Ala Ala Ala Leu Ala Leu Ala Val Gly Thr Ala Glu Pro Ala Ile Thr
 275 280 285
 Leu Leu Cys Asn Ala Val Gly Ala Leu Leu Asp Arg Pro Gly Gln Trp
 290 295 300
 Ala Leu Leu Gly Gly Asp Leu Asp Arg Ser Ala Ala Val Val Glu Glu
 305 310 315 320
 Thr Leu Arg Cys Leu Pro Pro Val Arg Leu Glu Ser Arg Val Ala Gln
 325 330 335
 Gln Asp Val Thr Leu Gly Gly Gln Phe Leu Pro Ala Asp Ser His Leu
 340 345 350
 Val Val Leu Val Ala Met Ala Asn Arg Gly Pro Arg Ala Ala Thr Ala
 355 360 365
 Pro Ser Pro Asp Ala Phe Asp Pro Gly Gly Ser Arg Val Pro Ala Arg
 370 375 380
 Asp Val Val Gly Leu Pro Gln Leu Ala Gly Ala Gly Pro Leu Ile Arg
 385 390 395 400
 Leu Val Val Thr Thr Ala Leu Arg Thr Leu Ala Glu Ala Leu Pro Thr
 405 410 415
 Leu Arg Arg Ala Ser Gly Gly Val Arg Trp Arg Arg Ser Pro Val Leu
 420 425 430

Leu Gly His Ala Arg Phe Pro Val Ala Arg Ala Glu Ser Gly Glu Gln
 435 440 445

Arg Ser Asp Asp Arg Pro Ala Leu Glu Glu Ala Ile Arg Cys Ala Ser
 450 455 460

<210> 21

<211> 1395

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 21

```

gtgaccgta cgccgggtcc gtcccgccga gtccgacgac agcaggagag gaagcgcatg      60
atcacagtcc cgcccgacgg ggatcccgcg acctggggccc gccggctgca actgaccgcg      120
gccgcgcagt gggtcgccgg caaccacggc gaccgtacg cgctgatcct gcgcgcggag      180
accgacgacc cgaccccgta cgagcagcgg gtggccgccc agccgctggt ccgcagcgag      240
cagttggaca cctgggtgac cggggacgcc gcgctggccc gggaggtggt gaccgacgac      300
cggttcggct ggctgacccg ggctgggcag cggcccgccg agcggaccct gccgctggcc      360
ggcacggcac tggaccacgg gccggaggcc cggcgtcggc tggacgcgct cgccgggttc      420
ggcgggcccgg tcctgcgggc cgacgccgca ggggcgcgta cccgggtcgt ggagaccacc      480
gcggtcctgc tcgacgggat cggggagcgg ttcgacctgg ccgtgctcgc ccggcggctg      540
gtcgtgcggg tgctggccga cctgctgggg gtgcccgcgg cgcgggcgggg ccgcttcgcc      600
gaggcactcg ccgccgccgg ccgtacgctg gacagccggc tgtgcccgca gaccgtggcg      660
accgctctcg ccaccgtcgc cgccaccgcc gagctgaccg acctgctggg cgaggtgccg      720
ccccgcggt cgctgtcccc gtccgccgcc ggctccgggc cgccgcgtcc gtccgcagcc      780
ggttcctggc cgccgctgcc ggctgacgac cggacggccg ccgcgctcgc gctggcggtc      840
ggcacggccg aaccggcgat caccctgctc tgcaacgcgg tcggtgcgct gctcgaccgc      900
cccgggcagt gggccctgct cggtagggac ctgacccggt ccgccgccgt cgtcgaggag      960
accctgcgct gccttcgcc ggtagcctg gagagccgcg tcgcgcagca ggacgtcacc     1020
ctgggcgggc agttcctccc ggcgacagc cacctggtcg tgctggtcgc catggcgaaac     1080
cggggtccgc gcgcggcgac cgccccgagc ccggacgcgt tcgaccctgg cgggtcgcgc     1140
gtcccgcccc gcgacgtggt gggcctgccg cagcttgccg gcgccgggccc gctgatcaga     1200
ctcgtcgtca cgaccgccct gccgaccctc gccgaggcgc tgcccacgct gcggcgggcg     1260
tccggcgggc tccggtggcg acgctgccc gtctgctcg gccacgccc ctttcccgtc     1320
gcacgggcgg agagcggcga acagcggctc gacgaccgcc cggcgctgga ggaggcgatc     1380

```

cgatgcgcggt cctga

1395

<210> 22

<211> 429

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 22

Met Thr Ser Phe Ala His Asn Thr His Tyr Tyr Ser Leu Val Pro Leu
1 5 10 15

Ala Trp Ala Leu Arg Ala Ala Gly His Glu Val Arg Val Ala Ser Gln
20 25 30

Pro Ser Leu Thr Asp Thr Ile Val Arg Ser Gly Leu Thr Ala Val Pro
35 40 45

Val Gly Asp Asp Gln Ala Ile Ile Asp Leu Leu Ala Glu Val Gly Gly
50 55 60

Asp Leu Val Pro Tyr Gln Arg Gly Leu Asp Phe Thr Glu Ala Arg Pro
65 70 75 80

Glu Val Leu Thr Trp Glu Tyr Leu Leu Gly Gln Gln Thr Met Leu Thr
85 90 95

Ala Leu Cys Phe Ala Pro Leu Asn Gly Val Ser Thr Met Asp Asp Met
100 105 110

Val Ala Leu Ala Arg Ser Trp Gln Pro Glu Leu Val Ile Trp Glu Pro
115 120 125

Phe Thr Tyr Ala Gly Pro Val Ala Ala Arg Val Val Gly Ala Thr His
130 135 140

Ala Arg Leu Leu Trp Gly Pro Asp Val Val Gly Asn Ala Arg Arg Leu
145 150 155 160

Phe Thr Glu Ser Leu Ala Arg Gln Pro Asp Glu Gln Arg Glu Asp Pro
165 170 175

Met Ala Glu Trp Leu Arg Cys Thr Leu His Arg Tyr Gly Cys Glu Leu
180 185 190

Gly Asp Asp Glu Val Glu Thr Leu Val Thr Gly Gly Trp Thr Ile Asp
195 200 205

Pro Thr Ala Asp Ser Thr Arg Leu Pro Val Pro Gly Arg Arg Val Ala
210 215 220

Met Arg Tyr Thr Pro Tyr Asn Ser Pro Ser Val Val Pro Glu Trp Val
225 230 235 240

Ala Lys Ala Asp Arg Pro Arg Val Cys Leu Thr Leu Gly Val Ser Ser

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

<210> 23

<211> 1290

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 23

atgacgtcct tcgcgcacaa caccactac tacagcctgg tgccgttggc ctgggcgctg 60
 cgcgcgggccg gccacgaggt acgggtggcg agccagccct cgctcaccga caccatcgtg 120
 cggtcggggc tgaccgcggt gccggtcggc gacgaccagg cgatcatcga cctgctcgcc 180
 gaggtcggcg gcgacctggt gccgtaccag cggggactgg acttcaccga ggcccgctccc 240
 gaagtgtga cctgggagta tctgctcggg cagcagacca tgctcaccgc gctgtgcttc 300
 gcgcgctca acggcgcttc cacgatggac gacatggtcg ccctggcccg gtcctggcag 360
 cccgagctgg tgatctggga gccgttcacc tacgccgggc cggtcgcggc gcgggtcgtc 420
 ggtgcgacgc acgcccggct gctctggggg ccggacgtgg tcggcaacgc ccggcggtg 480

```

ttcacccgaga gcctggcgcg gcagccggat gagcagcgcg aggacccgat ggccgagtgg      540
ttgcgctgca ccctgcaccg gtacggctgc gagctcggcg acgacgaggt ggagaccctg      600
gtcacccggcg ggtggaccat cgatcccacc gccgacagca cccggcttcc cgtccccggg      660
cgtcgggtgg ccatgcggta caccctgtac aacagcccgt ccgtggtgcc ggagtgggtg      720
gccaaggccg accggccccg cgtctgcctc accctcggcg tgcgagccg ggagacgtac      780
ggcagggacg tgggtctcctt ccaggagctg ctggcgccc tcggcgacct ggacgtcgag      840
gtcgtcgca cgctcagcga cgccagcgc gaggacctgg gtgacctgcc ggacaacgtc      900
cgggtgtgcg acttcgtgcc gctggacgtg ctgctgccga cctgtgccgc gatcatccac      960
cacggcgggg cgggcacgtg gtcgacggcc atgctctacg gggtgccgca gatcatgac      1020
gcgtcgctgt gggacgcccc gctcaaggcg cagcaggcgg agcgactcgg cacggggatc      1080
tcgatcccgc cggagcggct cgacccccg acgctgcggg cggccgtcgt ccggatcctc      1140
gacgaccgt cgatcgccgc cgcccccgc cgtcagcgcg acgagctgcg tgccgcgcg      1200
tcgcgggccg aggtggtccg catcctggaa cgctcgtcg cggacgaccg gcccggccg      1260
ccggccggaa ccgccaccga ccactcctga      1290

```

<210> 24

<211> 240

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 24

```

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

```

Gly His Leu Ala Thr Thr Ala Glu Leu Asp Ala Thr Leu Ala Arg Phe
 115 120 125
 Ala Ala His Leu Ser Pro Gly Gly Val Ala Ile Val Glu Pro Trp Trp
 130 135 140
 Phe Pro Glu Thr Phe Thr Pro Gly Tyr Val Gly Ala Ser Leu Val Glu
 145 150 155 160
 Val Asp Gly Arg Thr Ile Ser Arg Val Ser His Ser Val Arg Glu Gly
 165 170 175
 Gly Ala Thr Arg Ile Thr Val His Tyr Leu Val Ala Ser Pro Gly Gly
 180 185 190
 Gly Val Arg His Phe Asp Glu Ser His Leu Ile Thr Leu Phe Glu Arg
 195 200 205
 Ser Asp Tyr Glu Arg Ala Phe Ala Arg Ala Gly Phe Thr Thr Glu Tyr
 210 215 220
 Leu Thr Pro Gly Pro Ser Gly Arg Gly Leu Phe Val Gly Val His Pro
 225 230 235 240

<210> 25

<211> 723

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 25

atgtccatga tgtacgcgga cgccatcgcc gaggtctacg acctgatcta ccagggcaag 60
 ggcaaggact acgcggcgga ggcggcggag ctggaggcgc tggcccgggc cgcgcggccg 120
 cagccccgga cgtgctgga cgtggcgtgc ggcacggggc tgcacctgcg gcacctggcg 180
 gggctcttcg acgacgtggg cggcatcgag ctggcaccgg acatgctgag catcgcccag 240
 cagcgaaacc ccggggcggc cctgcacctc ggcgacatgc ggaccttcga cctggggcac 300
 cgctacgacg tcatcacctg catgttcagt tcggtgggcc acctggccac cacggccgag 360
 ctggacgcga cgttggcccg gttcgcgcg cacctgtccc ccgggggaggt ggcgatcgtc 420
 gagccgtggt ggttcccgga gaccttcacc cccgggtacg tgggcgcgag cctggtggag 480
 gtcgacggcc gtaccatctc gggggtctcc cattcggtgc gcgagggcgg cgcgaccg 540
 atcacctgc actacctgt ggccagcccc ggcgggggag tccggcactt cgacgagagc 600
 cacctgatca ccctcttcga acggtccgac tacgaacgtg ccttcgcccc ggcggggttc 660
 acgacggagt acctgacgcc cggcccgtcc ggccgcggtc tggtcgtcgg cgtccacccc 720
 tga 723

<210> 26
 <211> 1811
 <212> PRT
 <213> micromonospora carbonacea subspecies aurantiaca
 <400> 26

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

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

Ala Gly Ala Cys Gly Gly Gly Gly Ala Thr Gly Thr Thr Cys Cys Thr
 580 585 590
 Gly Gly Gly Cys Cys Thr Thr Cys Cys Cys Gly Cys Cys Ala Gly
 595 600 605
 Gly Ala Gly Ala Ala Gly Gly Ala Cys Cys Cys Gly Cys Gly Cys Ala
 610 615 620
 Ala Gly Thr Cys Gly Cys Thr Gly Cys Ala Cys Ala Cys Gly Gly Ala
 625 630 635 640
 Gly Gly Ala Gly Gly Thr Gly Cys Cys Gly Gly Thr Gly Cys Cys Cys
 645 650 655
 Gly Ala Gly Cys Thr Gly Gly Gly Cys Cys Cys Cys Gly Gly Gly Gly
 660 665 670
 Ala Gly Gly Cys Cys Cys Thr Cys Gly Thr Cys Gly Cys Gly Gly Thr
 675 680 685
 Cys Cys Thr Gly Gly Cys Cys Ala Gly Cys Thr Cys Gly Gly Thr Cys
 690 695 700
 Ala Ala Cys Thr Ala Cys Ala Ala Cys Ala Cys Gly Gly Thr Cys Thr
 705 710 715 720
 Gly Gly Thr Cys Gly Thr Cys Gly Thr Thr Gly Thr Thr Cys Gly Ala
 725 730 735
 Gly Cys Cys Gly Cys Thr Gly Cys Cys Cys Ala Cys Cys Thr Thr Cys
 740 745 750
 Gly Gly Cys Thr Thr Cys Cys Thr Gly Gly Ala Gly Cys Gly Cys Thr
 755 760 765
 Ala Cys Gly Gly Cys Cys Gly Gly Cys Thr Cys Thr Cys Cys Gly Ala
 770 775 780
 Gly Cys Thr Gly Gly Cys Cys Cys Gly Gly Cys Gly Gly Cys Ala Cys
 785 790 795 800
 Gly Ala Cys Cys Thr Gly Cys Cys Gly Thr Ala Cys Cys Ala Cys Ala
 805 810 815
 Thr Cys Cys Thr Cys Gly Gly Cys Thr Cys Gly Gly Ala Cys Cys Thr
 820 825 830
 Gly Gly Cys Cys Gly Gly Cys Gly Thr Gly Gly Thr Gly Cys Thr Gly
 835 840 845
 Ala Gly Gly Gly Thr Cys Gly Gly Gly Cys Cys Cys Gly Gly Cys Gly
 850 855 860
 Thr Cys Ala Ala Cys Cys Gly Cys Thr Gly Gly Cys Gly Gly Cys Cys
 865 870 875 880

Gly Gly Gly Thr Gly Ala Cys Gly Ala Gly Gly Thr Cys Gly Thr Gly
 885 890 895
 Gly Cys Gly Cys Ala Cys Thr Gly Cys Cys Thr Cys Thr Cys Gly Gly
 900 905 910
 Thr Gly Gly Ala Gly Cys Thr Gly Gly Ala Gly Thr Cys Cys Gly Cys
 915 920 925
 Cys Gly Ala Cys Gly Gly Cys Cys Ala Cys Gly Gly Cys Gly Ala Cys
 930 935 940
 Ala Cys Cys Ala Thr Gly Cys Thr Cys Gly Ala Cys Cys Cys Gly Gly
 945 950 955 960
 Ala Ala Cys Ala Gly Cys Gly Gly Ala Thr Cys Thr Gly Gly Gly Gly
 965 970 975
 Cys Thr Thr Cys Gly Ala Gly Ala Cys Cys Ala Ala Cys Thr Thr Cys
 980 985 990
 Gly Gly Cys Gly Gly Cys Cys Thr Cys Gly Cys Cys Gly Ala Gly Ala
 995 1000 1005
 Thr Cys Gly Cys Gly Thr Thr Gly Gly Thr Cys Ala Ala Gly Gly
 1010 1015 1020
 Cys Gly Ala Ala Cys Cys Ala Gly Cys Thr Gly Ala Thr Gly Cys
 1025 1030 1035
 Cys Cys Ala Ala Ala Cys Cys Cys Gly Ala Cys Cys Ala Cys Cys
 1040 1045 1050
 Thr Gly Ala Cys Cys Thr Gly Gly Gly Ala Gly Gly Ala Gly Gly
 1055 1060 1065
 Cys Cys Gly Cys Cys Gly Cys Gly Cys Cys Gly Gly Gly Ala Cys
 1070 1075 1080
 Thr Gly Gly Thr Cys Ala Ala Cys Thr Cys Cys Ala Cys Cys Gly
 1085 1090 1095
 Cys Cys Thr Ala Cys Cys Gly Cys Cys Ala Gly Cys Thr Gly Gly
 1100 1105 1110
 Thr Cys Thr Cys Cys Gly Gly Cys Ala Ala Cys Gly Gly Gly Gly
 1115 1120 1125
 Cys Cys Cys Gly Gly Ala Thr Gly Ala Ala Gly Cys Ala Gly Gly
 1130 1135 1140
 Gly Cys Gly Ala Cys Ala Ala Cys Gly Thr Cys Cys Thr Cys Gly
 1145 1150 1155
 Thr Cys Thr Gly Gly Gly Gly Gly Gly Cys Cys Ala Gly Cys Gly
 1160 1165 1170

Gly	Cys	Gly	Gly	Thr	Cys	Thr	Cys	Gly	Gly	Cys	Gly	Cys	Gly	Thr
	1175					1180					1185			
Thr	Cys	Gly	Cys	Cys	Ala	Cys	Cys	Cys	Ala	Gly	Cys	Thr	Cys	Gly
	1190					1195					1200			
Thr	Gly	Cys	Thr	Gly	Gly	Cys	Cys	Gly	Gly	Cys	Gly	Gly	Gly	Gly
	1205					1210					1215			
Cys	Cys	Ala	Ala	Thr	Cys	Cys	Cys	Gly	Thr	Cys	Thr	Gly	Cys	Gly
	1220					1225					1230			
Thr	Gly	Gly	Thr	Cys	Thr	Cys	Cys	Ala	Gly	Cys	Cys	Cys	Gly	Cys
	1235					1240					1245			
Gly	Cys	Ala	Ala	Gly	Gly	Cys	Cys	Gly	Ala	Cys	Ala	Thr	Cys	Thr
	1250					1255					1260			
Gly	Cys	Cys	Gly	Thr	Cys	Gly	Gly	Ala	Thr	Gly	Gly	Gly	Cys	Gly
	1265					1270					1275			
Cys	Cys	Gly	Ala	Gly	Gly	Cys	Cys	Gly	Thr	Cys	Ala	Thr	Cys	Gly
	1280					1285					1290			
Ala	Cys	Cys	Gly	Gly	Gly	Thr	Cys	Gly	Cys	Cys	Gly	Ala	Gly	Gly
	1295					1300					1305			
Ala	Cys	Thr	Ala	Cys	Cys	Gly	Cys	Thr	Thr	Cys	Thr	Gly	Gly	Thr
	1310					1315					1320			
Cys	Cys	Gly	Ala	Cys	Gly	Ala	Gly	Cys	Gly	Cys	Ala	Cys	Cys	Cys
	1325					1330					1335			
Ala	Gly	Ala	Ala	Thr	Cys	Cys	Cys	Cys	Gly	Gly	Gly	Ala	Gly	Thr
	1340					1345					1350			
Gly	Gly	Ala	Ala	Gly	Cys	Gly	Cys	Thr	Thr	Cys	Gly	Gly	Cys	Gly
	1355					1360					1365			
Cys	Ala	Cys	Gly	Cys	Ala	Thr	Thr	Cys	Gly	Gly	Gly	Ala	Gly	Cys
	1370					1375					1380			
Thr	Gly	Ala	Cys	Cys	Gly	Gly	Ala	Gly	Gly	Cys	Gly	Ala	Gly	Gly
	1385					1390					1395			
Ala	Cys	Gly	Thr	Cys	Gly	Ala	Cys	Ala	Thr	Cys	Gly	Thr	Cys	Thr
	1400					1405					1410			
Thr	Cys	Gly	Ala	Gly	Cys	Ala	Cys	Cys	Cys	Cys	Gly	Gly	Cys	Cys
	1415					1420					1425			
Gly	Gly	Gly	Ala	Gly	Ala	Cys	Gly	Thr	Thr	Cys	Gly	Gly	Cys	Gly
	1430					1435					1440			
Cys	Cys	Thr	Cys	Gly	Gly	Thr	Cys	Thr	Ala	Cys	Gly	Thr	Gly	Ala
	1445					1450					1455			

Cys Cys Cys Gly Cys Ala Ala Ala Gly Gly Ala Gly Gly Cys Ala	1460	1465	1470
Cys Cys Gly Thr Gly Gly Thr Cys Ala Cys Cys Thr Gly Cys Gly	1475	1480	1485
Cys Cys Thr Cys Gly Ala Cys Gly Ala Gly Cys Gly Gly Thr Thr	1490	1495	1500
Thr Cys Gly Ala Gly Cys Ala Cys Gly Thr Cys Thr Ala Cys Gly	1505	1510	1515
Ala Cys Ala Ala Cys Cys Gly Thr Thr Ala Cys Cys Thr Gly Thr	1520	1525	1530
Gly Gly Ala Thr Gly Thr Cys Cys Cys Thr Gly Ala Ala Gly Cys	1535	1540	1545
Gly Cys Ala Thr Cys Gly Thr Cys Gly Gly Cys Ala Cys Gly Cys	1550	1555	1560
Ala Cys Thr Thr Cys Gly Cys Cys Ala Ala Thr Thr Ala Cys Cys	1565	1570	1575
Gly Gly Gly Ala Gly Gly Cys Gly Thr Gly Gly Gly Ala Ala Gly	1580	1585	1590
Cys Cys Ala Ala Cys Cys Gly Gly Thr Thr Gly Gly Thr Gly Gly	1595	1600	1605
Thr Cys Ala Ala Gly Gly Gly Cys Ala Ala Gly Ala Thr Cys Cys	1610	1615	1620
Ala Cys Cys Cys Gly Ala Cys Gly Cys Thr Gly Thr Cys Gly Cys	1625	1630	1635
Gly Cys Thr Gly Cys Thr Ala Cys Cys Cys Gly Cys Thr Gly Gly	1640	1645	1650
Ala Gly Gly Ala Gly Gly Thr Cys Gly Gly Cys Cys Ala Gly Gly	1655	1660	1665
Cys Gly Gly Thr Cys Thr Ala Cys Gly Ala Cys Gly Thr Cys Cys	1670	1675	1680
Ala Thr Cys Ala Cys Ala Ala Cys Cys Thr Gly Cys Ala Cys Cys	1685	1690	1695
Ala Gly Gly Gly Cys Ala Ala Gly Gly Thr Cys Gly Gly Cys Gly	1700	1705	1710
Thr Gly Cys Thr Cys Gly Cys Gly Cys Thr Cys Gly Cys Gly Cys	1715	1720	1725
Cys Gly Cys Gly Cys Gly Ala Gly Gly Gly Gly Cys Thr Cys Gly	1730	1735	1740

Gly Gly Gly Thr Cys Cys Gly Gly Ala Ala Cys Cys Cys Gly Gly
1745 1750 1755

Ala Gly Cys Thr Gly Cys Gly Gly Gly Ala Ala Thr Gly Cys Cys
1760 1765 1770

Ala Thr Cys Thr Thr Gly Cys Cys Gly Cys Gly Ala Thr Cys Ala
1775 1780 1785

Ala Cys Cys Gly Cys Thr Thr Cys Cys Gly Gly Gly Thr Gly Cys
1790 1795 1800

Cys Gly Gly Cys Cys Thr Gly Ala
1805 1810

<210> 27

<211> 1359

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 27

```

atgccagaca cccccgagct gaaccggata ctgcacgcga tcctcgccca ggagaccgac      60
gcgcggggagc tggcggccct gccgctgccc tcctcctacc gggccgtgac ggtgcacaag      120
gacgagacgg ggatgttcct gggccttccc cgccaggaga aggaccgcg caagtcgctg      180
cacacggagg aggtgccggt gcccgagctg gggccggggg aggccctcgt cgcggtcctg      240
gccagctcgg tcaactacaa cacggtctgg tcgtcgttgt tcgagccgct gccaccttc      300
ggcttcctgg agcgctacgg ccggctctcc gagctggccc ggcggcacga cctgccgtac      360
cacatcctcg gctcggacct ggccggcgtg gtgctgaggg tcggggcccg cgtcaaccgc      420
tggcggcccg gtgacgaggt cgtggcgcac tgccctctcg tggagctgga gtccgccgac      480
ggccacggcg acaccatgct cgaccgcgaa cagcggatct ggggcttcga gaccaacttc      540
ggcggcctcg ccgagatcgc gttgggtcaag gcgaaccagc tgatgcccaa acccgaccac      600
ctgacctggg aggaggccgc cgcgccggga ctgggtcaact ccaccgccta ccgccagctg      660
gtctccggca acggggcccg gatgaagcag ggcgacaacg tcctcgtctg gggggccagc      720
ggcgggtctcg gcgcgttcgc caccagctc gtgctggccg gcggggccaa tcccgtctgc      780
gtggtctcca gcccgcgcaa ggccgacatc tgccgtcggg tgggcgccga ggccgtcatc      840
gaccgggtcg ccgaggacta ccgcttctgg tccgacgagc gcaccagaa tcccgggag      900
tggaagcgct tcggcgcacg cattcgggag ctgaccggag gcgaggacgt cgacatcgtc      960
ttcgagcacc ccggccggga gacgttcggc gcctcggtct acgtgaccgc caaaggaggc     1020
accgtggtca cctgcgcctc gacgagcggg ttcgagcacg tctacgaaa ccgttacctg     1080

```

tggatgtccc tgaagcgcac cgtcggcacg cacttcgcca attacccgga ggcgtgggaa 1140
 gccaacccggt tgggtgtcaa gggcaagatc caccgcacgc tgcgcgctg ctaccgctg 1200
 gaggaggtcg gccaggcggc ctacgacgtc catcacaacc tgcaccaggg caaggctggc 1260
 gtgctcgcgc tcgcgccgcg cgaggggctc ggggtccgga acccggagct gcgggaatgc 1320
 catcttgccg cgatcaaccg cttccgggtg ccggcctga 1359

<210> 28

<211> 636

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 28

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

Glu Leu Lys Gln Ala Gln Gln Glu Ala Ala Arg Ala Arg Ala Glu Leu
 210 215 220
 Asn Ala Val Arg Met Ala Asp Arg Leu Gln Pro Glu Leu Val Gln Leu
 225 230 235 240
 Ile Leu Asp Gly Gly Glu Leu Asp Asp Leu Val Gly Ser Ala Val Arg
 245 250 255
 Arg Leu Gly Gly Ala Leu His Val Arg Asp Arg Ala Asn Gly Val Leu
 260 265 270
 Ala Ala Ala Gly Glu Ile Pro Val Pro Asn Glu Arg Glu Leu Ala Arg
 275 280 285
 Val Arg Leu Asn Ala His Ala Thr Gly Arg Pro Gly Arg Leu Thr Thr
 290 295 300
 Gly Ser Trp Val Val Pro Leu Ala Ala Arg Ala Gly Asp Leu Gly Cys
 305 310 315 320
 Val Leu Phe His Ala Asp Glu Pro Ser Asp Asp Glu Arg Met Ala Ala
 325 330 335
 Leu Pro Ala Val Ala Gln Thr Val Ala Leu Leu Met Thr Arg Asn Gly
 340 345 350
 Gly Ser His Gly Gln Pro Gly Asp Gly Leu Leu Glu Asp Leu Leu Gly
 355 360 365
 Pro Trp Pro Asp Leu Glu Arg Gly Gly Lys Arg Arg Arg Tyr Thr Pro
 370 375 380
 Val Glu Phe Asp Arg Pro Tyr Val Val Val Val Ala Arg Pro Glu Gly
 385 390 395 400
 Ala Thr Ser Pro Arg Val Phe Glu Arg Ala Val Ser Val Ala His Gly
 405 410 415
 Leu Asn Gly Met Lys Ala Ile Arg Asp Gly Gln Ala Val Leu Leu Leu
 420 425 430
 Pro Gly Asp Asp Pro Gly Ala Arg Ala Arg Asp Val Thr Arg Glu Leu
 435 440 445
 Ser Gly Leu Leu Gly Leu Pro Val Thr Ala Gly Gly Ala Gly Pro Val
 450 455 460
 Arg Thr Ala Asp Ser Val Ser Arg Thr Tyr Gln Glu Ala Ala Arg Cys
 465 470 475 480
 Val Asp Ala Leu Ala Ala Leu Asp Ala Lys Gly Arg Ala Ala Cys Ser
 485 490 495
 Arg Asp Leu Gly Phe Leu Gly Leu Leu Val Ala Gly Gly His Asp Val
 500 505 510

Thr Gly Phe Val Asp Arg Val Ile Gly Pro Val Leu Ser Tyr Asp Ala
 515 520 525
 Arg Arg Leu Thr Asn Leu Arg Glu Thr Leu Gln Thr Tyr Phe Asp Ser
 530 535 540
 Ala Gly Ser Arg Thr Arg Ala Ala Glu Met Leu His Leu His Pro Asn
 545 550 555 560
 Thr Val Ser Arg Arg Leu Asp Arg Ile Ser Gln Leu Leu Gly Arg Asp
 565 570 575
 Trp Arg Gln Pro Asp Arg Ala Leu Asp Thr Gln Leu Ala Leu Arg Leu
 580 585 590
 His Arg Ile Arg Gly Leu Leu Cys Gln Glu Arg Gly Tyr Pro Gly Pro
 595 600 605
 Ser Gln Glu Pro Asp Gln Pro Ala Arg Pro Ile Arg Arg His Arg Pro
 610 615 620
 Pro Ala Ser Ala Gly Arg Ala Pro Arg Thr Pro Arg
 625 630 635

<210> 29

<211> 1911

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 29

gtgcaccagg cgcaccggga cggagtggac caggccacgc tcgaccgggt gatgatcgcc 60
 aagcgactcg cgttggagct tcgagaggct atcgggaggc ggtgtcagcg gcaggcggag 120
 ctggccgccc tcgtcgacac cgcccgtgac ctgcccgggg cgacgaacct ggaggccggg 180
 ctgcagctgg tgggtcgggc gacccaactg ctgctcgccg gggacgtggc gttcgtcagc 240
 ctgctcgacg acgcgaccgg cgaatcctac gtcgcctcgg ccgtcggggc ggccaccgcg 300
 ctgaccagcg gctaccggct gccctggcgc gacgggctgg tcgtggccgc cgcaccgcgc 360
 gagccactct cctggacggc ggaccacctc gccgacgagc gcctcgaacg acaccgggcc 420
 gccgacggcc tgggtccgcgc ggaagggctg cacgcggtgc tgtccgtggc tctgagcgct 480
 gagggccggc acctcggcaa cctgcacgtc ggccaccggc aggtccgcca cttcgccccg 540
 gacgaggtcg cgtcgctgcg cctgctcgcc gatctcgcg cgacggcagt ggagcggatc 600
 atgctgctcg acgacacgtg ggccgaactc aagcaggccc agcaggaggc ggccagggcc 660
 cgagccgagc tgaacgcggc ccgcatggcc gaccgcctgc aaccgaact cgtccagctc 720
 atcctcgacg ggggcgaact cgacgacctg gtgggcagcg ccgtgcggcg actgggcggc 780

```

gccctgcacg tgcgtgaccg ggccaacggc gtgctggcgg cggccgggtga aatccctgtc      840
ccgaacgagc gggaactggc ccgagtgcgg ctgaacgccc acgccaccgg ccgaccgggc      900
cgcttgacca cgggttcttg ggtgggtgcc ctggcgggccc gcgcgggtga cctcggctgt      960
gtgttgttcc acgccgacga gccgtccgac gacgagcggg tggcgggcct gccggcggtc     1020
gcgcagaccg tggcgctgct gatgaccagg aacggcgggg gccacggcca gccggggcgac     1080
gggctcctgg aggacctgct cggcccgtgg ccggacctgg agcgggggcg gaagcgccgt     1140
cggtaacacac ctgtcgagtt cgaccggccc tacgtcgtcg tggtcgcccg ccccgagggc     1200
gccacctcgc cccgggtggt cgaacggggc gtctccgtcg cccacggcct gaacggcatg     1260
aaggccatcc gggacggcca ggcgggtgct ctgctgcccg gtgacgacct gggggcccgg     1320
ggccgggacg tgacgcggga actgagcggg ctgctcgccc taccggtcac ggccggaggc     1380
gccggaccgg tgcgcacggc ggactcggtc agccgcacct accaggaggc ggcccgggtg     1440
gtcgacgccc tggcgcgct ggacgcgaag gggcgggcgg cctgctcacg ggacctgggc     1500
ttcctcgggc tgctggctgc cggcgggccac gacgtcaccc gtttcgtcga ccgggtcatc     1560
ggacctcgtg tgagctacga cgcgcgccgg ctacgaatc tcaggagagc cctccagacc     1620
tacttcgact cggcggggcag ccgtaccggg gcggcggaga tgctgcatct gcatccgaac     1680
accgtgtccc gccggctgga ccgcatctcc cagctgctcg gccgggactg gcggcagccg     1740
gaccggggccc tcgacacgca gctcgtcttg cgcctgcacc ggatccgtgg cctgctctgc     1800
caggaacggg gctacccggg cccatcgag gagccggacc aaccgcgcgc gcctatccgg     1860
cggcaccgcc ctocagcatc cgcagggcgt gcgccacgga cgccaagggtg a             1911

```

<210> 30

<211> 403

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 30

```

Met Val Pro Thr Leu Asp Val Arg Glu Glu Val Thr Ala Ala Arg Ser
1              5              10              15

```

```

Asp Pro Asp Thr Val Ser Arg Phe Cys Ala Ala Leu Leu Ala Ser Leu
20              25              30

```

```

Pro Arg Ala Asp Gln Arg Arg Lys Gly Glu Leu Tyr Val Arg Gly Leu
35              40              45

```

```

Leu Thr Ala Ser Gly Arg Lys Thr Met Arg Asn Leu Ala Ala Ile Ala
50              55              60

```

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

Gly Ile Gln Asp Phe Glu Gly Arg Ser Tyr Gln Gly Trp His Arg His
 370 375 380

Val Thr Leu Ala Ser Val Ala His Ala Leu Arg Met Leu Glu Gly Gly
 385 390 395 400

Ala Ala Gly

<210> 31

<211> 1212

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 31

```

atggtgccga cgctcgacgt ccgcgaggag gtgaccgcgg caaggtccga tccggacacc      60
gtgtcccggg tctgcgccgc cctgctggcc tcgctgcccc gcgccgacca gcgacgcaag      120
ggcgaactgt acgtccgggg gctgctgacc gcctccggcc gcaagaccat gcgcaacctg      180
gccgccatcg ccgacgatcc ggccggcgga cagagcatgc accacttcac cagttgctcc      240
acctgggact gggagaccgt ccgtgccgcg ctgcgccggc acctggaccg gacgctgtcg      300
ccccgggcct ggggtggtgc gtgatgctg gtgccgaaga ccggccggca ctcggtcggc      360
gtggaacgcc ggtacgtgcc cgcgctgggc gagacgggtc acagccagca gagctacggc      420
ctctggctgg cctcggagac cgtcgccgcg cccatcaact ggcagttgtc catcggttaag      480
ggttggtctc aggacaaccg cgcgcgcgcg agcgtaccgg cggacgagga cggcacgacc      540
agcgacggcg cggcgggtgca ggcgggtgct aaggccgcgg cctggggaat cggccctcgc      600
ccggtggtaa tggacgcccg gcaactcggc ctgccccgcg tgatcgaggc gttcaccacg      660
gcgggtctgc ccttcctgct acggatcaac agcggctgca ccctgctggc cgcggggccc      720
ggcccgcgcg agaaccgggt cgcggcgccc tccgccgagc acctgctcag cctgacgcgg      780
gccagcgccc gtccggtgga gtggatcgac ccggccagcc ccggcgacg gcgcacgagc      840
ctggtcgcac cgctacaggt ctattggccg ggctgtccg gtgccgccc gcccggtccg      900
tccgcccccg ccccgccggg ggcggcgcgc gccgccgcgc ccgggctgcc cctgacactg      960
ctcggcaagt ggcagacctc cgagcgcggc gtacggcaga tgtggctgac caacatgacc     1020
gacgccgggt acggcccaact gctgcggtcg agcaagctca cccggcgggt cgagaccgac     1080
ttctcccagg tcagcctcga cgtcggcatc caagacttcg agggtcggtc ataccaaggc     1140
tggcaccggc acgtcacctt ggcgtccgtg gcgcacgccc tgcggatgct ggagggcggt     1200
gccgcccgat ag                                     1212

```

<210> 32
 <211> 481
 <212> PRT
 <213> micromonospora carbonacea subspecies aurantiaca
 <400> 32

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

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

Thr

<210> 33

<211> 1446

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 33

atgacctccg cagcgcacca ttccccgcat ccggcgaagg ccgacgccct gatggacgac 60

gcccacgccg acatcggggc cgatgccgag gccgacggtc gacggctcga ccgggcccgc 120

ctgcggcggg tcgccgggct gtgcaccgag agggccgacg tcacggaggt cgagtaccgg 180

caggtgcggc tggagcgcgt cgtcctggtc ggcgtgtgga cctcgggcac cgccgacgag 240
 gccgaacggc ccctcgccga gctggcggca ctgcgccaga ccgcgggagc cgtgggtgctc 300
 gacgggggtga tccagcgcgc cgaccggccc gaccgggca cgtacatcgg ctccggcaag 360
 gcgcgggagt tgcgggacat cgtccaggag gtggggggccg acacgggtgat ctgcgacggc 420
 gagctgagcc cggcccaact ggtacgcctc gaagaggctc tgcacgcaa ggtgggtggac 480
 cgcaccgcgc tgatcctcga catcttcgcc cagcacgcca cgtcccgcga ggggaaggcg 540
 caggtggccc tggcacagat gcaatacatg ctgccgcggc tgcgcggctg gggccagtcg 600
 ctctcccggc agatgggcgg aggtgccggc ggcgggtggca tggccacccg ggggcccggc 660
 gagaccaaga tcgagaccga ccggcggcgc atccacgaga ggatggcccg gctccgacgg 720
 gagatcgcg agatgaagtc cggccgcgaa ctcaagcgcc gcgatcgcg ggcgaacagc 780
 gtcccgtcgg tcgcatcgc cggttacacc aacgccggca agtcctcgt gctcaaccgg 840
 ctactggcg cgagcgtgct ggtgcagaac gcgctgttcg ccaccctcga cccgacggcg 900
 cgccgggcca ccaccccag cgggcgcagc tacacgatca ccgacaccgt cggattcgtc 960
 cggcacctgc cgcaccacct ggtggaggcg ttccgctcca ccctggaaga ggtggccgag 1020
 gccgacctcc tgctgcacgt ggtggacggc gccaccccg ccccgctgga gcagctcgcc 1080
 tcggtgcgcg cggatcatccg ggacgtggac gcggcgggag tgcccgaact cgtcgtgac 1140
 aacaaggccg acgccgccac cccggccgcc ctggccgcgt tggcggaggc cgagccgcac 1200
 cacgtcgtcg tctcggcccg caccggtcag ggcacgaca cgcttcggca gttgctggag 1260
 gccgcgctgc cgcaccggga ggtccgggtc gacgtcctga tcccgtacgt cgcgggcagc 1320
 ctcgtggccc ggtgcacgc cgacggcgag gtgctggccg aggagcacac ggccgacggc 1380
 accctgctgc aggcgcgggt ggccccgcac ctggctgccg agctcagcgc gtacgccagg 1440
 acctga 1446

<210> 34

<211> 408

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 34

Met Lys Arg Asp Leu Gly Asp Leu Ala Leu Phe Gly Gly His Ala Ser
 1 5 10 15

Phe Leu Gln Gln Ile His Val Gly Arg Pro Asn Arg Ile Asp Arg Ala
 20 25 30

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

Leu Glu Pro Tyr Arg Ser Arg Gln Gln Phe Gln Leu Pro His Thr Glu
 340 345 350
 Arg Leu Ser Ala Arg Val Leu Ala Leu Pro Thr Gly Ser Ala Ile Ser
 355 360 365
 Arg Glu Asp Ile Arg Arg Val Cys Asn Ile Val Arg Leu Ala Val Ser
 370 375 380
 Arg Gly Phe Glu Leu Thr Ala Arg Trp Gln Gln Gln Pro Gly Pro Asp
 385 390 395 400
 Gly Gln Ser Val Val Ala Pro Gly
 405

<210> 35

<211> 1227

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 35

atgaagcgag atctcgggga tctggcactc ttcggaggac acgccagctt cctccagcag	60
atccacgtcg ggcgccccaa ccg gatcgat cgggccaggc tgttcgaccg gctgtcctgg	120
gcgctcgaca acgagtgggt gaccaacaac gggccgctgg cacgggagtt cgaggagcgg	180
gtcgccgaca tggtcgggggt cggcaactgc gtggcgacgt gcaacgccac ggtggccctc	240
cagctgctcg cgcacgccac cgagctgacc ggtgaggtga tcatgccatc gctcaccttc	300
gccgcgaccg cacacgcggt gcgctggctc gggctggagc cggctcttctg cgacatcgac	360
ccgcgccaccg gatgcctcga ccacgtggcg gtccgcgcgg ccacacgcc gcgcacgtcg	420
gcggtcttcg gcgtccacct ctggggccgc ccctgcgacg tcaacgcgct ggagaagggtg	480
accgccgacg cgggcctgcg cctgttcttc gacgccgcc acgccatcgg gtgcacctca	540
cagggccgcc cgggtggggcg gtccggccac gccgaggtgt tcagcttcca cgcgacgaag	600
gtcgtcaacg ccttcgaggg cggggcgatc gtcaccgacg acgacgacct cgcccaccgc	660
gtccgctccc tggcgaactt cggcttcggc ctgcacagcc ccagcgcggc cggcggcacc	720
aacgcgaaga tgagcgaggc gtccgcccgc atggggctca cctcgctcga cgcgttcccc	780
gaggtggccc gccacaacca ggccaactac gagcagtact gcggtgagct ggcccggatt	840
cccggcctca gcgtgatcga cttcgcccc gacgagcggc acaactacca gtacgtgatc	900
gtcgagatcg acccgacgt caccgggttg caccgcgacc tgctcgtcga cctgctccgg	960
gccgagaacg tcgtggcgca gcgtacttc tcgccggcct gtcaccaatt ggagccctac	1020
cgggtccggc agcagttcca gctgccgcac accgagcggc tctcggcgcg cgtcctggcg	1080

ctgccgaccg gctccgccat ctcccgggaa gacatccgca ggggtgtgcaa catcgtgcgg 1140
 ttggcggtct cccggggatt cgaattgacc gctcgggtggc agcagcagcc cgggccccgac 1200
 ggacagagcg tgggtggcacc cggttga 1227

<210> 36

<211> 488

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 36

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

Glu Gln Lys Leu Lys Gly Glu Arg Gly Leu Trp Ser Leu His Ala Ile
 225 230 235 240
 Arg Thr Ser Leu Tyr Gly Leu Asn Asp Glu Glu Tyr Arg Ala Thr Thr
 245 250 255
 Gly Lys Gln Gly Ala Phe Thr Arg Val Arg Ala Asn Leu Thr Arg Phe
 260 265 270
 Gln Gln Leu Arg Ala Glu Arg Gly Glu Pro Val Arg Leu Gly Leu Ser
 275 280 285
 Tyr Ile Val Leu Pro Gly Arg Ala Gly Arg Leu Ser Ala Leu Ile Asp
 290 295 300
 Phe Val Ala Glu Leu Asn Glu Ala Ala Pro Asp Arg Pro Leu Asp Tyr
 305 310 315 320
 Ile Asn Leu Arg Glu Asp Tyr Ser Gly Arg Pro Asp Gly Lys Leu Ser
 325 330 335
 Leu Asp Glu Arg Ala Glu Leu Gln Ala Glu Leu His Arg Phe Arg Glu
 340 345 350
 Arg Ala Met Gln Arg Thr Pro Thr Leu His Ile Asp Tyr Gly Tyr Ala
 355 360 365
 Leu His Ser Leu Met Thr Gly Ser Asp Val Glu Leu Val Arg Ile Arg
 370 375 380
 Pro Glu Thr Met Arg Pro Ala Ala His Pro Gln Val Ser Val Gln Val
 385 390 395 400
 Asp Ile Leu Gly Asp Val Tyr Leu Tyr Arg Glu Ala Ala Phe Pro Gly
 405 410 415
 Leu Ala Gly Ala Asp Arg Tyr Arg Ile Gly Thr Val Ser Pro Gly Thr
 420 425 430
 Thr Leu Ala Gln Val Val Glu Thr Phe Val Thr Ser Gly Gly Ser Val
 435 440 445
 Val Ala Lys Pro Gly Asp Glu Tyr Phe Leu Asp Gly Phe Asp Gln Ala
 450 455 460
 Val Thr Ala Arg Leu Asn Gln Met Glu Thr Asp Val Ala Asp Gly Trp
 465 470 475 480
 Gly Asp Arg Arg Gly Phe Leu Arg
 485

<210> 37

<211> 1467

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 37

```

gtgggagggc cctgaccat ggagatctcc gcctcgaatc cctggcgac ctgcgtgtc      60
cccgagcagc acccgaccgc ggcggcgcg gtgctgtacg acgaggtcgc cgggtcagga    120
atcgtgccgc cggcagagat cggggccgcc gccaggggt tggaggcatt ggcacgcac      180
tacgggacca caccttttct gccgcttgag caggcccgcc gcgaaatcgg cctggaccgg    240
gccgggttcg ggcggctgct ggacctgttc gcccggttc cggggttcg caccgcagt      300
gagaacggac cgtccggtcg ctactggacc aacacgggtgc tcggcctcga aagggccggc    360
gtcttcgacg cctgctcga ccggaggcgc gcgtttccgc atctcgtcgg gctctaccg      420
ggccccacgt gcatgttcg ctgtcacttc tgcgtaaggg tcaccggggc cgtctaccg      480
gcctcggcgc tggacgacgg gaacgccatg ttcgcctctg tcacgacga ggtccccgc      540
cacaaccgcg acgcggtgta cgtctccggt ggcctcgagc cactcaccaa ccccgggctc    600
ggtgcactgg tcagccgggc gcccgagcgg ggatttcgga tcacctcta caccaactcg      660
ttcgccctca cggagcagaa gctcaagggg gagcggggat tgtggagcct gcacgccatc    720
cgcacgtcgc tgtacgggtt gaacgacgag gaataccggg cgaccaccgg caagcagggg    780
gccttcaccc gggtaggggc gaacctcacg cggttccagc agctgcgtgc cgagcggggc    840
gagccggtgc ggctcggcct cagctacatc gtctgcccg gccgcgcgg gcggctgagc      900
gcgctgatcg acttcgtcgc cgagctcaac gaggcggcac cggaccgccc gctggactac     960
atcaacctgc gggaggacta cagcgggcgg ccggacggga agctctccct ggacgagcgc    1020
gccgagctcc aggccgagct gcaccggttc cgggagaggg caatgcagcg gacgccgacc    1080
ctgcacatcg actacggcta cgcctgcac agcctgatga cgggaagcga cgtggagctc    1140
gtgcgtatcc ggccggagac gatgcgccct gcggcccacc cgcaggtgtc ggtgcaggtg    1200
gatatactcg gtgatgtcta cctctatcgg gaggcggcgt ttccgggcct ggccggtgcc    1260
gaccgctatc gcatcggcac ggtatctccc ggcacgacgt tggcgaggt ggtggagacg    1320
ttcgtgacca gcggcggatc ggtggtcgcg aagcctggcg acgaatactt cctggacgga    1380
ttcgaccagg cggtagccgc gcggctgaac cagatggaga ccgacgtcgc cgatggctgg    1440
ggagaccgac ggggtttcct ccgctga                                         1467

```

<210> 38

<211> 277

<212> PRT

<213> micromonospora carbonacea subspecies aurantiaca

<400> 38

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

<210> 39

<211> 834

<212> DNA

<213> micromonospora carbonacea subspecies aurantiaca

<400> 39

atgccctaca tccagcacgc cgggcgacat gaattcggcc agaatttcct ggtcgaccgc	60
tcggtgatcg acgatttcgt cgaactcgtc gcccgaccgc acggccctat cgtggagatc	120
ggcgccggcg acggtgcgct gaccctaccc ctgagccggc agggaaggga gttgaccgca	180
gtggagatcg actccaagcg ttccaagcgg ctgagccggc agacaccgca caacgtcacc	240
gtggtctgcg cggatgtcct gagcttcggg tccccccagc atccgcacgt ggtcgtcggg	300
aacatcccct tccacgtgac ccccccatc gtgcgggctc tcctcgccgc ggaccactgg	360
cacacggcgg tgctgctggg gcagtgagg gtggcccgca ggcgggcccg cgtcggcggc	420
gcgacgctgc tgaccgcgag ctggtggccc tggtagact tcgaactgca ctcccgggtt	480
ccggcccgcg ccttcgggcc tgtcccttcc gtcgacggcg ggctgttctc catggtccgt	540
cgcgggaccc cgtggtcga cgaccggagg ggttaccagg aattcgtccg gctggtgttc	600
accggcaagg ggcacggatt gccggagatc cttcagcgga ccgggcggat cggccgcaag	660
gaccagcagg actggcaacg ggccaaccgg gtggggccgc agcacctgcc caaggacctg	720
accgcccacc agtgggcctc cctgtggcac ctggtggcac ccgcccggcc ggccggcccc	780
cgccgtccgg caccgcgcg gccaggaagc cccgcttcgg cgcgcggcg ctga	834

0-1	Form - PCT/RO/134 (EASY) Indications Relating to Deposited Microorganism(s) or Other Biological Material (PCT Rule 13bis)	
0-1-1	Prepared using	PCT-EASY Version 2.92 (updated 01.01.2002)
0-2	International Application No.	
0-3	Applicant's or agent's file reference	3016-3PCT
1	The indications made below relate to the deposited microorganism(s) or other biological material referred to in the description on:	
1-1	page	45
1-2	line	17
1-3	Identification of Deposit	
1-3-1	Name of depositary institution	National Microbiology Laboratory, Health Canada
1-3-2	Address of depositary institution	Federal Laboratories for Health Canada, Room H5190, 1015 Arlington Street, Winnipeg, Manitoba, Canada R3E 3R2
1-3-3	Date of deposit	10 July 2002 (10.07.2002)
1-3-4	Accession Number	NMLHC IDAC 100702-1
1-4	Additional Indications	A request to restrict access to a sample of any deposit of biological material pertaining to the above noted International application is made with respect to all designated states having enacted any such provisions in their respective national legislation. Australia Notice under Regulation 3.25 (3); CBE Rule 28(4); Canadian Patent Rules, Section 104(4) Notice.
1-5	Designated States for Which Indications are Made	all designated States
1-6	Separate Furnishing of Indications These indications will be submitted to the International Bureau later	NONE
2	The indications made below relate to the deposited microorganism(s) or other biological material referred to in the description on:	
2-1	page	45
2-2	line	17

2-3	Identification of Deposit	
2-3-1	Name of depositary institution	National Microbiology Laboratory, Health Canada
2-3-2	Address of depositary institution	Federal Laboratories for Health Canada, Room H5190, 1015 Arlington Street, Winnipeg, Manitoba, Canada R3E 3R2
2-3-3	Date of deposit	10 July 2002 (10.07.2002)
2-3-4	Accession Number	NMLHC IDAC 100702-2
2-4	Additional Indications	A request to restrict access to a sample of any deposit of biological material pertaining to the above noted International application is made with respect to all designated states having enacted any such provisions in their respective national legislation. Australia Notice under Regulation 3.25 (3); CBE Rule 28(4); Canadian Patent Rules, Section 104(4) Notice.
2-5	Designated States for Which Indications are Made	all designated States
2-6	Separate Furnishing of Indications These indications will be submitted to the International Bureau later	NONE
3	The indications made below relate to the deposited microorganism(s) or other biological material referred to in the description on:	
3-1	page	45
3-2	line	17
3-3	Identification of Deposit	
3-3-1	Name of depositary institution	National Microbiology Laboratory, Health Canada
3-3-2	Address of depositary institution	Federal Laboratories for Health Canada, Room H5190, 1015 Arlington Street, Winnipeg, Manitoba, Canada R3E 3R2
3-3-3	Date of deposit	10 July 2002 (10.07.2002)
3-3-4	Accession Number	NMLHC IDAC 100702-3

3-4	Additional Indications	A request to restrict access to a sample of any deposit of biological material pertaining to the above noted International application is made with respect to all designated states having enacted any such provisions in their respective national legislation. Australia Notice under Regulation 3.25 (3); CBE Rule 28(4); Canadian Patent Rules, Section 104(4) Notice.
3-5	Designated States for Which Indications are Made	all designated States
3-6	Separate Furnishing of Indications These indications will be submitted to the International Bureau later	NONE

FOR RECEIVING OFFICE USE ONLY

0-4	This form was received with the international application: (yes or no)	yes
0-4-1	Authorized officer	Sophie Madeau / RO/CA

FOR INTERNATIONAL BUREAU USE ONLY

0-5	This form was received by the international Bureau on:	
0-5-1	Authorized officer	