

Lantibiotic Biosynthetic Gene Clusters From *A. garbadinensis* and *A. liguriae*

Field Of The Invention

This invention relates to characterisation of the biosynthetic gene cluster for the lantibiotic actagardine, identification of a novel variant of actagardine and its biosynthetic cluster, and methods of production and use of actagardine, a novel actagardine variant produced in a strain of *A. liguriae*, and variants of both of these produced according to this invention, utilizing genes from the characterised biosynthetic gene clusters.

Background Of The Invention

Lantibiotics are peptides having antibiotic and other activities, produced by Gram-positive bacteria. They contain, among other modified residues, the thioether amino acids lanthionine and methyllanthionine, which cross-link the peptide chain into a polycyclic structure. They have been classified into two classes, type-A and type-B, though such classification is not unproblematic. Type-A lantibiotics are generally elongate amphiphiles that are capable of forming pores in bacterial and other plasma membranes. Examples are nisin and subtilin. Type-B lantibiotics, by contrast, are globular, conformationally defined peptides that inhibit enzyme functions. Examples are cinnamycin and duramycin.

Activities ascribed to type-B lantibiotics such as cinnamycin include antimicrobial activity (providing potential application as antibiotics), inhibition of angiotensin-converting enzyme (providing a potential application in blood pressure regulation), immunomodulation via inhibition of phospholipase A2 (providing a potential application as anti-inflammatories), and interference with prostaglandin and leucotriene biosynthesis.

Type-B lantibiotics appear to exert their activity by interfering with enzyme activities by blocking the respective substrates. For example, type B lantibiotics such as mersacidin and actagardine have been found to inhibit biosynthesis of peptidoglycan; transglycosylation was identified as the target reaction. The substrate for this reaction is the lipid-bound cell wall precursor lipid II. While this is a target for the lantibiotic vancomycin, the site of action is different and is a new target binding site not used by any current antibacterial drug.

For the cinnamycin class of type B lantibiotics antibacterial activity has been observed, in particular with *Bacillus* strains, with effects described on membrane functions, ATP-dependent proton translocation and  $\text{Ca}^{2+}$ -uptake, and on ATPases. Also, the formation of defined pores in phosphatidylethanolamine-containing planar membranes has been reported. These effects

can be attributed to the specific binding of these type-B lantibiotics to phosphatidylethanolamine.

Lantibiotics have been shown to have efficacy and utility as food additives and antibacterial agents against *Propionibacterium acnes* and problematic pathogens, e.g. methicillin-resistant *Staphylococcus aureus* (MRSA), which has or is developing resistance to many commonly used antibiotics, and *Streptococcus pneumoniae*. For reviews, see Sahl and Bierbaum (1998) Annual Rev. Microbiol. 52:41-79; Jack and Sahl (1995) TIBTECH 13:269-278; Gasson (1995) Chapter 10, Lantibiotics, in Vining and Stuttard (eds) Biotechnology Series: Genetics and Biochemistry of Antibiotic Production, Biotechnological I 30 Series 28, pages 283-306.

Within the field of antibiotics, there is a continuing need for the provision of new antibiotic compounds, to overcome issues such as resistance, bio-compatibility, toxicity and the like. Accordingly, methods of producing lantibiotics, and the production of variant forms of lantibiotics (which may have a different activity profile compared to native forms), are desirable.

Actagardine is a known type B tetracyclic lantibiotic, 19 amino acids in length (1890 Da). It has potent activity against important Gram positive pathogens such as *Staphylococcus aureus* and *Streptococcus pyogenes* both in vitro and in in vivo animal models. The structure of actagardine is shown in Figure 4. The compound is produced from a pre-pro-peptide, the C-terminal portion of which has the polypeptide sequence of SSGWVCTLTIECGTVICAC (SEQ ID NO:4). The polypeptide of SEQ ID NO:4 is modified by the following crosslinks, creating secondary and tertiary structure: CROSSLINK 1-6, Lanthionine (Ser-Cys); CROSSLINK 7-12, Beta-methyllanthionine (Thr-Cys); CROSSLINK 9-17, Beta-methyllanthionine (Thr-Cys); CROSSLINK 14-19, Beta-methyllanthionine sulfoxide (Thr-Cys).

Actagardine has been reported to be produced by two species of Actinoplanes; *A. garbadinensis* and *A. liguriae*. Also co-produced is an analogue in which the CROSSLINK 14-19 is not oxidized i.e. it is a beta-methyllanthionine not betamethyllanthionine sulfoxide which is named herein deoxy-actagardine.

US 6,022,851 describes the isolation of actagardine from isolated strains of *A. garbadinensis* and *A. liguriae*.

### Disclosure of the Invention

The present invention relates to the cloned, sequenced and elucidated structural and regulatory information relevant to the biosynthetic gene cluster for the type-B lantibiotic, actagardine, from *Actinoplanes garbadinensis* and *A. liguriae*.

5

We have also surprisingly found that in an isolate of *A. liguriae*, designated herein as *A. liguriae* NCIMB 41362, a novel form of actagardine is produced which we have termed actagardine B or, in the non-oxidised form, deoxy-actagardine B. These forms have similar anti-microbial activity to actagardine and are generated from the primary polypeptide  
10 sequence of SEQ ID NO:1, which undergoes similar cross-linking to actagardine. The variants provide new and useful alternatives to actagardine. In addition, the identification of residues in actagardine B which are different from actagardine leads to the provision of further lantibiotics based on these differences.

15 We have also isolated gene clusters from both actagardine-producing *A. garbadinensis* and *A. liguriae* NCIMB 41362 which comprise the genes for the production of actagardine and actagardine B.

In one aspect, the present invention provides the novel actagardine B and variants thereof,  
20 including variants based on the primary polypeptide sequences of SEQ ID NO:2 and SEQ ID NO:3, as well as variants thereof.

In a further aspect, the invention provides nucleic acids encoding actagardine B and its variants, sets of nucleic acids and variants thereof derived from the above-mentioned gene  
25 clusters, methods of making actagardine B and its variants, and methods of generating novel variants of actagardine B.

### Description Of The Drawings

30

Figure 1 provides a map of the actagardine encoding and regulatory gene cluster isolated from *A. garbadinensis*.

Figure 2 provides a map of the encoding and regulatory gene cluster isolated from *A. liguriae* which encodes a novel variant of actagardine, herein referred to as actagardine B.

Figure 3 provides a schematic showing a method disclosed herein for generation of actagardine variants utilizing nucleic acid sequences isolated from *A. garbadinensis* or from *A. liguriae*.

- 5 Figure 4 is a representation of the primary structure of mature actagardine where X1-X2 represent Val-Ile, Y is –S(O)– and Z is NH<sub>2</sub>. "Deoxy-actagardine B" is the Val<sup>15</sup>Leu Ile<sup>16</sup>Val variant with a non-oxidised methylanthionine bridge between AbuS<sup>14</sup> and AlaS<sup>19</sup>.

#### Description of the Sequences

- 10 For the convenience of the reader, the sequences of the present application have been numbered non-contiguously as follows:

**SEQ ID NO:1** is the primary polypeptide sequence of Actagardine B:

SSGWVCTLTIECGTLVCAC.

**SEQ ID NO:2** is the primary polypeptide sequence of Actagardine B variant VV:

- 15 SSGWVCTLTIECGTVVCAC.

**SEQ ID NO:3** is the primary polypeptide sequence of Actagardine B variant LI

SSGWVCTLTIECGTLICAC.

**SEQ ID NO:4** is the primary polypeptide sequence of Actagardine:

SSGWVCTLTIECGTVICAC;

- 20 **SEQ ID NO:11** is the primary polypeptide sequence of Ala-Actagardine B:

ASSGWVCTLTIECGTLVCAC.

**SEQ ID NO:12** is the primary polypeptide sequence of Ala-Actagardine B variant VV:

ASSGWVCTLTIECGTVVCAC.

**SEQ ID NO:13** is the primary polypeptide sequence of Ala-Actagardine B variant LI

- 25 ASSGWVCTLTIECGTLICAC.

**SEQ ID NO:14** is the primary polypeptide sequence of Ala-Actagardine:

ASSGWVCTLTIECGTVICAC.

**SEQ ID NO:212** is the primary polypeptide sequence of pre-pro-Actagardine B:

MSAITVETTWKNTDLREDLTAHPAGLGFGELSFEDLREDRTIYAASSGWVCTLTIECG

- 30 TLVCAC.

**SEQ ID NO:22** is the primary polypeptide sequence of pre-pro-Actagardine B variant VV:

MSAITVETTWKNTDLREDLTAHPAGLGFGELSFEDLREDRTIYAASSGWVCTLTIECG

TVVCAC.

**SEQ ID NO:23** is the primary polypeptide sequence of pre-pro-Actagardine B variant LI

- 35 MSALAIEKSWKDVDLRDGATSHAPAGLGFGELTFEDLREDRTIYAASSGWVCTLTIECG

TLICAC.

**SEQ ID NO:119** is the primary polypeptide sequence of pre-pro-Actagardine:

MSALAIEKSWKDVDLRDGATSHPAGLGFGELTFEDLREDRTIYAASSGWVCTLTIECG  
TVICAC.

**SEQ ID NO:100** is the non-vector, *A. garbadinensis*-derived, nucleotide sequence of the  
5 cosmid CosAG14.

**SEQ ID NOs:101 – 132** are the polypeptide sequences of the open reading frames *orf1* –  
*orf32* of SEQ ID NO:100 respectively.

**SEQ ID NO:200** is the non-vector, *A. liguriae*-derived, nucleotide sequence of the cosmid  
CosAL02.

10 **SEQ ID NOs:201 – 231** are the polypeptide sequences of the open reading frames *orf1* –  
*orf31* of SEQ ID NO:200 respectively.

**SEQ ID NOs:300-312** are primer sequences described herein below.

#### Detailed Description of the Invention

15 The present invention relates to the gene clusters of SEQ ID NO:100 and SEQ ID NO:200 and  
the polypeptides encoded by these clusters and variants thereof. The polypeptide of SEQ ID  
NO:119 is pre-pro-actagardine and the polypeptide of SEQ ID NO:212 is pre-pro-actagardine  
B. The remaining polypeptides and their variants (as defined herein) are referred to herein  
generically as "cluster polypeptides". Cluster polypeptides derived from SEQ ID NO:100 are  
20 referred to as "1xx polypeptides" and those derived from SEQ ID NO:200 are referred to as  
"2xx polypeptides". Polypeptides which are 100% identical in both sequence and length to a  
cluster polypeptide are referred to as "wild-type" polypeptides. A cluster polypeptide derived  
from SEQ ID NO:100 or SEQ ID NO:200 may be wild type or variant.

25 A polypeptide may be in substantially isolated form. Isolated polypeptides of the invention will  
be those as defined above in isolated form, free or substantially free of material with which it is  
naturally associated such as other polypeptides with which it is found in the cell. For example,  
the polypeptides may of course be formulated with diluents or adjuvants and still for practical  
purposes be isolated.

30 A polypeptide of the invention may also be in a substantially purified form, in which case it will  
generally comprise the polypeptide in a preparation in which more than 90%, e.g. 95%, 98%  
or 99% of the polypeptide in the preparation is a polypeptide of the invention.

#### ***Lantibiotic Polypeptide and LantibioticA gene***

35 In the present invention, reference to a lantibioticA or LanA polypeptide, the *LantibioticA* or  
*LanA* gene refers generically to a type B lantibiotic polypeptide or the gene encoding such a

peptide. Thus reference to these includes reference to cinnamycin, mersacidin, actagardine and actagardine B and the genes encoding these products. Reference to a lantibiotic producing host cell refers to any host cell which in its native form produces a LanA polypeptide, as further defined herein below.

5

A LanA polypeptide is a polypeptide with anti-microbial activity. Anti-microbial activity may be examined by determining the MIC value against a reference organism, e.g. *Micrococcus luteus*. A LanA polypeptide is considered to exhibit anti-microbial activity if it has a MIC value of less than or equal to 16-fold higher than that of actagardine against the same strain of the reference microorganism. In the present invention, the *A. garbadinensis* LanA gene is referred to as *actA* and the *A. liguriae* LanA gene is referred to as *LigA*.

10

### **Other Lan Polypeptides**

15

As used herein, reference to a "LanM" polypeptide is a polypeptide derived from a Lantibiotic gene cluster which is a modification factor required for the conversion of a precursor polypeptide to a lantibiotic compound. LanM polypeptides include those of SEQ ID NO:120 (ActM) or a variant thereof, SEQ ID NO:213 (LigM) or a variant thereof, a cinM polypeptide as defined in WO02/088367, a mrsM polypeptide as disclosed in Altena *et al*, 2000, or a homologous polypeptide from another gene cluster of a bacteria which produces a type B lantibiotic.

20

Reference to a "LanR" polypeptide is a polypeptide derived from a Lantibiotic gene cluster which is a regulatory factor required for the regulation of production of a precursor polypeptide. LanR polypeptides include those of SEQ ID NO:122 (ActR) or a variant thereof, SEQ ID NO:216 (LigR) or a variant thereof, a cinR1 polypeptide as defined in WO02/088367, a mrsR1 polypeptide as disclosed in Altena *et al*, 2000, or a homologous polypeptide from another gene cluster of a bacteria which produces a type B lantibiotic.

25

Reference to a "LanT" polypeptide is a polypeptide derived from a Lantibiotic gene cluster which is a transporter factor required for the production of a precursor polypeptide to a lantibiotic compound. LanT polypeptides include those of SEQ ID NO:123 (ActT) or a variant thereof, SEQ ID NO:214 (LigT) or a variant thereof, a cinT polypeptide as defined in WO02/088367, a mrsT polypeptide as disclosed in Altena *et al*, 2000, or a homologous polypeptide from another gene cluster of a bacteria which produces a type B lantibiotic.

30

Reference to a "LanO" polypeptide is a polypeptide derived from a Lantibiotic gene cluster which is a factor believed to be involved in the oxidation of the deoxy-form of actagardine and

35

compounds of the invention to actagardine or to compounds of the invention in which Y is - S(O)-.

LanO polypeptides include those of SEQ ID NO:122 (ActO) or a variant thereof, SEQ ID NO:215 (LigO) or a variant thereof, or a homologous polypeptide from another gene cluster of a bacteria which produces a type B lantibiotic.

### **Cluster Polypeptides**

In one aspect, the invention provides an isolated cluster polypeptide selected from any one of SEQ ID NOs: 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230 and 231. In another aspect, the invention provides a cluster polypeptide which is a variant of any of the above-mentioned sequences.

Cluster polypeptides of particular interest include 1xx and 2xx polypeptides which are LanM, LanR, LanT or LanO polypeptides.

A "variant", in relation to a cluster polypeptide, denotes: any polypeptide having an amino acid sequence which is different from, but which shows significant amino acid sequence identity with, the amino acid sequence of a reference polypeptide (in this case any wild type cluster polypeptide), or a fragment of that polypeptide.

Unless otherwise specified, significant amino acid sequence identity is preferably at least 80%, more preferably 85%, 90% or 95%, still more preferably 98% or 99%. A variant is preferably of a length which is the same as, or at least 70%, preferably at least 80%, more preferably at least 90% and most preferably at least 95% of the length of the wild type cluster polypeptide.

"Percent (%) amino acid sequence identity" is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the sequence with which it is being compared, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. The % identity values used herein are generated by BLAST- 2 which was obtained from Altschul et al. (1996); <http://blast.wustl.edu/blast/README.html>. WU-BLAST-2 uses several search parameters, most of which are set to the default values. The adjustable parameters are set with the following values: overlap span =1, overlap fraction = 0.125, word threshold (T) = 11. The HSPS and HSPS2

parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched; however, the values may be adjusted to increase sensitivity. A % amino acid sequence identity value is determined by the number of matching identical residues divided by the total number of residues of the "longer" sequence in the aligned region, multiplied by 100. The "longer" sequence is the one having the most actual residues in the aligned region (gaps introduced by WU BLAST-2 to maximize the alignment score are ignored).

Desirably, a variant will retain a biological function of the reference polypeptide. In the present invention, biological function is retained wherein the variant, when present in a host cell with the other members of its cluster, is capable of producing a lantibiotic. This may be determined, for example, by providing a host cell containing SEQ ID NO:100 in the case of a 1xx cluster polypeptide variant, or SEQ ID NO:200 in the case of a 2xx polypeptide variant, wherein the host cells produce actagardine or actagardine B respectively, modifying the sequence to encode the variant, and determining whether a lantibiotic polypeptide is still produced.

### ***Precursor Polypeptides***

In another aspect, the invention provides polypeptides, preferably in isolated form, which are precursors of the compounds of the present invention or of actagardine. The precursor polypeptides include the polypeptides of any one of SEQ ID NOs:1-4, SEQ ID NOs:11-14, SEQ ID NOs:212, 22, 23 and 119, as well as variants or derivatives thereof which can be converted to a lantibiotic polypeptide.

A variant of a precursor polypeptide of any one of SEQ ID NOs:1-4 is a polypeptide in which one or more, for example from 1 to 5, such as 1, 2, 3 or 4 amino acids are substituted by another amino acid. Preferably the amino acid is at a position selected from positions 2, 3, 4, 5, 8, 10, 11, 13 or 18 of any one of SEQ ID NOs:1-4.

A variant of a precursor polypeptide of any one of SEQ ID NOs:11-14 is a polypeptide in which one or more, for example from 1 to 5, such as 1, 2, 3 or 4 amino acids are substituted by another amino acid. Preferably the amino acid is at a position selected from positions 3, 4, 5, 6, 9, 11, 12, 14 or 19 of any one of SEQ ID NOs:11-14.

A variant of a precursor polypeptide of any one of SEQ ID NOs:212, 22, 23 and 119 is a polypeptide in which one or more, for example from 1 to 5, such as 1, 2, 3 or 4 amino acids of

the C-terminal region (residues 46-64) corresponding to SEQ ID NOs:1-4 respectively are substituted by another amino acid. Preferably the amino acid is at a position selected from positions corresponding to positions 2, 3, 4, 5, 8, 10, 11, 13 or 18 of any one of SEQ ID NOs:1-4. Such variants may further include changes to the N-terminal region which retain at least 70%, for example at least 80%, preferably at least 90%, for example at least 95% of the N-terminal regions (residues 1-45). For example, a variant of the N-terminal region of SEQ ID NO:212 or SEQ ID NO:119 may comprise one or more substitutions (e.g. from 1 to 12, such as from 1 to 5, e.g. 1, 2 or 3 substitutions at positions 4, 5, 6, 8, 9, 12, 13, 17, 18, 19, 21 and 32 which our data shows are varied between SEQ ID NO:212 and 119.

Substitutions may be of one amino acid by another naturally occurring amino acid and may be conservative or non-conservative substitutions. Conservative substitutions include those set out in the following table, where amino acids on the same block in the second column and preferably in the same line in the third column may be substituted for each other:

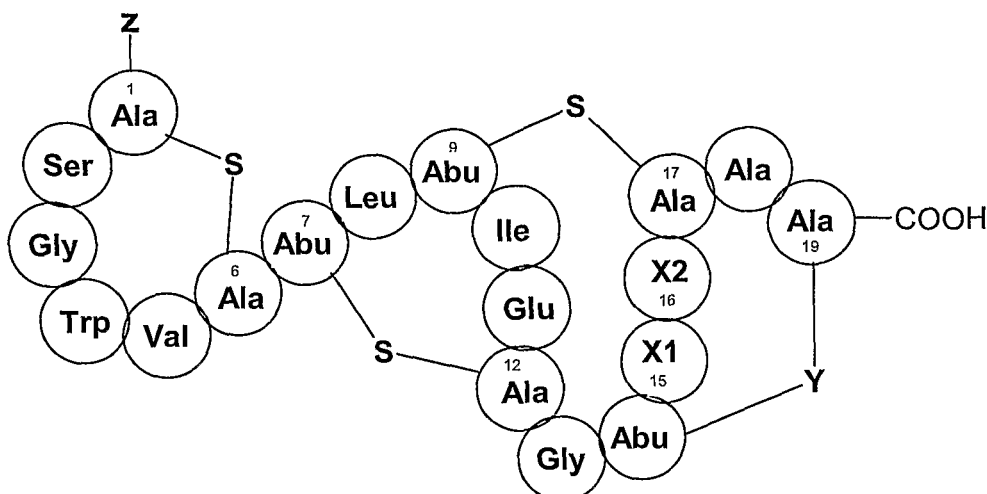
|           |                   |         |
|-----------|-------------------|---------|
| ALIPHATIC | Non-polar         | G A     |
|           |                   | I L V   |
|           | Polar - uncharged | C S T M |
|           |                   | N Q     |
|           | Polar - charged   | D E     |
|           |                   | K R     |
| AROMATIC  |                   | H F W Y |

For SEQ ID NO:212, the substitutions may be of an amino acid which differs from the amino acid residue located in the corresponding location of SEQ ID NO:119, or *vice versa*. In either case, the substitution may be to introduce the SEQ ID NO:119 amino acid into SEQ ID NO:212, or *vice versa* (e.g. Ile at position 4 of SEQ ID NO:212 may be substituted by Leu, and so on).

A precursor polypeptide may be obtained by expression of a nucleic acid encoding the polypeptide in a cell which is a non-producer of a lantibiotic.

### **Compounds**

In one aspect, the present invention provides a compound of the formula (I):



wherein:

-X1-X2- represent -Leu-Val-; -Val-Val- or -Leu-Ile-;

Y is -S- or -S(O)-; and

5 Z is either H<sub>2</sub>N- or Ala-,

or a pharmaceutically acceptable salt thereof. In a further aspect, the invention provides variants and biologically active derivatives of these compounds.

10 Where -X1-X2- represent -Leu-Val-, Y is -S(O)- and Z is NH<sub>2</sub> the compound of the invention is also referred to as actagardine B.

Where -X1-X2- represent -Leu-Val-, Y is -S(O)- and Z is Ala- the compound of the invention is also referred to as ala-actagardine B.

15 Where -X1-X2- represent -Leu-Val-, Y is -S- and Z is NH<sub>2</sub> the compound of the invention is also referred to as deoxy-actagardine B.

Where -X1-X2- represent -Leu-Val-, Y is -S- and Z is Ala- the compound of the invention is also referred to as ala-deoxy-actagardine B.

20 It will be understood by reference to Z being a group H<sub>2</sub>N-, that this moiety represents the N-terminus of the alanine residue at position 1 of the above compound. By reference to the group Z being Ala-, it will be understood that this moiety represents an alanine, conventionally referred to in the art as Ala(0), linked to the alanine at position 1 via an amide bond.

Variants

A variant of a compound of formula (I) is a compound which one or more, for example from 1 to 5, such as 1, 2, 3 or 4 amino acids are substituted by another amino acid. Preferably the amino acid is at a position selected from positions 2, 3, 4, 5, 8, 10, 11, 13 or 18 of the compound of formula (I).

A variant may also comprise a substitution at position 15 or 16, provided that when both positions 15 and 16 are substituted and none of the other positions are changes, 15 and 16 are not Val and Ile respectively.

Where Z is Ala-, variants of compounds of the invention include those in which Ala- is replaced by another amino acid (particularly a naturally occurring amino acid encoded by the genetic code or its D- isoform), more particularly an amino acid selected from the group Ile-, Lys-, Phe-, Val-, Glu-, Asp-, His-, Leu, Arg-, Ser- and Trp-. In one aspect, the amino acid may be selected from the group Ile-, Lys-, Phe-, Val-, Glu-, Asp-, His-, Leu-, Arg- and Ser-. Such variants may be produced by chemical addition of the residue to compounds where Z = H<sub>2</sub>N, as described in US 6,022,851. It will be appreciated that the chemical addition of an amino acid allows the amino acid to be in the L- or D- configuration. This includes D-Ala, in addition to the D- forms of other amino acids such as those mentioned above.

Derivatives

Derivatives of compounds of the invention (including variants) are those in which one or more amino acid side chain of the compound of the invention has been modified, for example by esterification, amidation or oxidation.

Derivatives of compounds of the invention may be monoamide derivatives at one of the carboxy functions of actagardine, particularly at the C-terminal. More particularly, a derivative may be a compound in which the C-terminal of the compound of the invention is of the formula -COR, in which R represents the group -NR<sup>1</sup>R<sup>2</sup>, wherein R<sup>1</sup> and R<sup>2</sup> independently represent:

(i) hydrogen;

(ii) a group of formula -(CH<sub>2</sub>)<sub>n</sub>-NR<sup>3</sup>R<sup>4</sup>, in which n represents an integer from 2 to 8 and R<sup>3</sup> and R<sup>4</sup> independently represent hydrogen or (C<sub>1</sub>-C<sub>4</sub>) alkyl or R<sup>3</sup> and R<sup>4</sup> taken together represent a group -(CH<sub>2</sub>)<sub>3</sub>-, -(CH<sub>2</sub>)<sub>4</sub>-, (CH<sub>2</sub>)<sub>2</sub>-O-(CH<sub>2</sub>)<sub>2</sub>-, -(CH<sub>2</sub>)<sub>2</sub>-S-(CH<sub>2</sub>)<sub>2</sub>- or -(CH<sub>2</sub>)<sub>5</sub>-; or R<sup>1</sup> and R<sup>2</sup> taken together with the adjacent nitrogen atom represent a piperazine moiety which may be substituted in position 4 with a substituent selected from:

(a) (C<sub>1</sub>-C<sub>4</sub>)alkyl;

(b) (C<sub>5</sub>-C<sub>7</sub>)-cycloalkyl,

(c) pyridyl,

(d)  $-(CH_2)_p-NR^5R^6$  in which p represents an integer from 1 to 8 and  $R^5$  and  $R^6$  independently represent hydrogen or (C<sub>1</sub>-C<sub>4</sub>) alkyl;

(e) piperidiny;

(f) substituted piperidiny, wherein the substituted piperidiny bears a N-substituent which is (C<sub>1-4</sub>)alkyl;

(g) benzyl; and

(h) substituted benzyl, wherein the phenyl moiety bears 1 or 2 substituents selected from chloro, bromo, nitro, (C<sub>1</sub>-C<sub>4</sub>)alkyl and (C<sub>1</sub>-C<sub>4</sub>)alkoxy.

In one embodiment, in the formula -COR, R represents the group -NR<sup>1</sup>R<sup>2</sup>, wherein R<sup>1</sup> and R<sup>2</sup> independently represent hydrogen, a group of formula  $-(CH_2)_n-NR^3R^4$ , in which n represents an integer from 2 to 8 and R<sup>3</sup> and R<sup>4</sup> independently represent hydrogen or (C<sub>1</sub>-C<sub>4</sub>) alkyl or R<sup>3</sup> and R<sup>4</sup> taken together represent a group  $-(CH_2)_3-$ ,  $-(CH_2)_4-$ ,  $(CH_2)_2-O-(CH_2)_2-$ ,  $-(CH_2)_2-S-$ ,  $(CH_2)_2-$  or  $-(CH_2)_5-$ , or R<sup>1</sup> and R<sup>2</sup> taken together with the adjacent nitrogen atom represent a piperazine moiety which may be substituted in position 4 with a substituent selected from (C<sub>1</sub>-C<sub>4</sub>)alkyl, (C<sub>5</sub>-C<sub>7</sub>)-cycloalkyl, pyridyl, benzyl, and substituted benzyl wherein the phenyl moiety bears 1 or 2 substituents selected from chloro, bromo, nitro, (C<sub>1</sub>-C<sub>4</sub>)alkyl and (C<sub>1</sub>-C<sub>4</sub>)alkoxy.

Further, a derivative may include a compound in which the carboxy function of a side chain of an internal residue, e.g. that of the residue Glu11, is modified from -COOH to a group -COOR<sup>5</sup> in which R<sup>5</sup> represents hydrogen, (C<sub>1</sub>-C<sub>4</sub>)alkyl or (C<sub>1</sub>-C<sub>4</sub>)alkoxy (C<sub>2</sub>-C<sub>4</sub>)alkyl.

N-terminal derivatives of compounds of the invention may be those in which the N-terminal amino group -NH<sub>2</sub> is instead a group -NHR<sup>6</sup> wherein R<sup>6</sup> represents C<sub>1-4</sub>alkyl.

The term "(C<sub>1</sub>-C<sub>4</sub>)alkyl" represents straight or branched alkyl chains of from 1 to 4 carbon atoms, such as: methyl, ethyl, propyl, 1-methylethyl, butyl, 1-methylpropyl or 1,1-dimethylethyl while the term "(C<sub>2</sub>-C<sub>4</sub>)alkyl" represents straight or branched alkyl chains of from 2 to 4 carbon atoms such as: ethyl, propyl, 1-methylethyl, butyl, 1-methylpropyl or 1,1-dimethylethyl. The term "(C<sub>5</sub>-C<sub>7</sub>)cycloalkyl" represents a cycloalkyl group selected from cyclopentyl, cyclohexyl and cycloheptyl.

The term "(C<sub>1</sub>-C<sub>4</sub>)alkoxy" represents a straight or branched alkoxy chain of 1 to 4 carbon atoms such as methoxy, ethoxy, propoxy, 1-methylethoxy, butoxy, 1-methylpropoxy and 1,1-dimethylethoxy.

Derivatives according to the present invention may be made according to the methods described for the manufacture of derivatives of actagardine in EP-0195359, the disclosure of which is incorporated herein by reference.

#### Further Embodiments

- 5 Where the derivative is a compound where the C-terminal is of the formula -COR, in which R represents the group -NR<sup>1</sup>R<sup>2</sup>, in some embodiments, R<sup>1</sup> is H and R<sup>2</sup> represents a group of formula -(CH<sub>2</sub>)<sub>n</sub>-NR<sup>3</sup>R<sup>4</sup>, in which n represents an integer from 2 to 8 and R<sup>3</sup> and R<sup>4</sup> independently represent hydrogen or (C<sub>1</sub>-C<sub>4</sub>) alkyl or R<sup>3</sup> and R<sup>4</sup> taken together represent a group -(CH<sub>2</sub>)<sub>3</sub>-, -(CH<sub>2</sub>)<sub>4</sub>-, (CH<sub>2</sub>)<sub>2</sub>-O-(CH<sub>2</sub>)<sub>2</sub>-, -(CH<sub>2</sub>)<sub>2</sub>-S-(CH<sub>2</sub>)<sub>2</sub>- or -(CH<sub>2</sub>)<sub>5</sub>-. In these  
 10 embodiments, R<sup>3</sup> and R<sup>4</sup> preferably represent hydrogen or (C<sub>1</sub>-C<sub>4</sub>) alkyl. More preferably R<sup>3</sup> and R<sup>4</sup> represent (C<sub>1</sub>-C<sub>2</sub>) alkyl, e.g. methyl. Integer n may be preferably from 2 to 5, and more preferably 2 to 4, e.g. 3.

- In other embodiments, R<sup>1</sup> and R<sup>2</sup> taken together with the adjacent nitrogen atom represent a  
 15 piperazine moiety. The N-substituent in the 4 position may preferably be selected from:

- (a) (C<sub>1</sub>-C<sub>4</sub>)alkyl;
- (b) (C<sub>5</sub>-C<sub>7</sub>)-cycloalkyl,
- (d) -(CH<sub>2</sub>)<sub>p</sub>-NR<sup>5</sup>R<sup>6</sup> in which p represents an integer from 1 to 8 and R<sup>5</sup> and R<sup>6</sup> independently represent hydrogen or (C<sub>1</sub>-C<sub>4</sub>) alkyl;
- 20 (e) piperidinyl; and
- (f) substituted piperidinyl, wherein the substituted piperidinyl bears a N-substituent which is (C<sub>1-4</sub>)alkyl.

- The piperdinyl and substituted piperidinyl groups preferably have their nitrogen atom at the 4-  
 25 position.

The N-substituent may more preferably be selected from:

- (d) -(CH<sub>2</sub>)<sub>p</sub>-NR<sup>5</sup>R<sup>6</sup> in which p represents an integer from 1 to 8 and R<sup>5</sup> and R<sup>6</sup> independently represent hydrogen or (C<sub>1</sub>-C<sub>4</sub>) alkyl; and
- 30 (f) substituted piperidinyl, wherein the substituted piperidinyl bears a N-substituent which is (C<sub>1-4</sub>)alkyl.

If the N substituent is -(CH<sub>2</sub>)<sub>p</sub>-NR<sup>5</sup>R<sup>6</sup>, then R<sup>5</sup> and R<sup>6</sup> may be preferably (C<sub>1</sub>-C<sub>4</sub>)alkyl, more preferably (C<sub>1</sub>-C<sub>2</sub>) alkyl, e.g. methyl. Integer p is preferably 1 to 4, e.g. 3.

35

If the N substituent is substituted piperidiny, then the N-substituent is preferably (C<sub>1</sub>-C<sub>2</sub>) alkyl, e.g. methyl. As mentioned above, the N is preferably in the 4-position.

### ***Nucleic Acid***

A nucleic acid of the invention may be a DNA or RNA, though preferably a DNA. A nucleic acid of the invention may be single- or double-stranded. In one aspect, the invention provides an isolated nucleic acid encoding a cluster polypeptide. In another aspect, the invention provides an isolated nucleic acid encoding a precursor polypeptide or variant or fragment thereof.

In a further aspect, the invention provides an isolated nucleic acid which may comprise all or a fragment of SEQ ID NO:100 or SEQ ID NO:200, including a fragment comprising an intergenic region disclosed herein. Such regions may include a promoter or other regulatory element for the expression of a cluster polypeptide or a precursor polypeptide of the present invention.

Twenty-five nucleotides is recognised by those skilled in the art as a sufficient number of nucleotides to be specific to the particular gene or gene cluster or sub-sequence thereof as disclosed herein. Thus fragments include fragments of SEQ ID NO:100 or SEQ ID NO:200, or variants thereof having significant sequence identity, which are at least 25, e.g. at least 30, e.g. at least 50, e.g. at least 100, e.g. at least 250 nucleotides in length.

Promoters that are variants of those intergenic sequences are also included and the specific intergenic sequences (or parts thereof) are preferred embodiments. In all cases, where a preferred embodiment of an orf, gene, nucleic acid, polypeptide or promoter is defined by reference to a specific sequence, the invention in its broader sense is intended to include embodiments having variants of that specific sequence.

The term "variant" as used herein in relation to a particular nucleic acid (the reference nucleic acid) denotes: any nucleic acid having a sequence which is different from that of the reference nucleic acid, but which is its complement or which shows significant nucleic acid sequence identity with, or hybridization under stringent conditions to, the reference nucleic acid or its complement or a fragment of the reference nucleic acid or its complement; or any nucleic acid which encodes an amino acid sequence having significant amino acid sequence identity with the amino acid sequence encoded by the reference nucleic acid, or a fragment of that nucleic acid. The term "variant" also refers to nucleic acids which differ from each other due only to the degeneracy of the genetic code, and which therefore encode identical deduced amino acid sequences. Variant nucleic acids of the invention are further defined as follows. If a variant

nucleic acid of the invention is introduced into the gene clusters identified herein, in place of the sequence of which it is a variant, and the recombinant fragment is introduced into a suitable host cell under suitable conditions for lantibiotic production (e.g. as shown in the Examples), then production of a molecule having one or more activities of a lantibiotic (especially antibiotic activity) will result. Preferably production will be regulated to occur at high cell density.

Significant nucleic acid sequence identity is preferably at least 50%, more preferably 60%, 70%, 80% or 90%, still more preferably 95%, 98% or 99%. Significant nucleic acid sequence identity is preferably shown between the variant nucleic acid (or a portion thereof) and a fragment of at least 30 residues of the reference nucleic acid, more preferably a fragment of at least 60, 90 or 120 residues, still more preferably a fragment of 180, 240 or 300 residues, more preferably the entire reference nucleic acid.

"Percent (%) nucleic acid sequence identity" is defined as the percentage of nucleotide residues in a candidate sequence that are identical with the nucleotide residues in the sequence under comparison. The identity values used herein were generated by the BLASTN module of WU BLAST-2 set to the default parameters, with overlap span and overlap fraction set to 1 and 0.125, respectively.

In relation to variants of the promoters used in the present invention, nucleic acid sequence identity is preferably assessed over a sequence of at least 30 residues, more preferably 40 or 50 residues, still more preferably 60 residues.

"Stringent conditions" or "high stringency conditions", as defined herein, may be identified by those that: (1) employ low ionic strength and high temperature for washing, for example 0.015 M sodium chloride/0.0015 M sodium citrate/ 0.1% sodium dodecyl sulfate at 50°C; (2) employ during hybridization a denaturing agent, such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 g/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC (sodium chloride/sodium citrate) and 50% formamide at 55°C, followed by a high-stringency wash consisting of 0.1 x SSC containing EDTA at 55°C.

When a nucleic acid of interest is said to be in "operative association" with a promoter or regulatory sequence, this means that the promoter/regulatory sequence is able to direct transcription of the nucleic acid of interest in an appropriate expression system, with the nucleic acid of interest in the correct reading frame for translation. Preferably when a nucleic acid of interest is in operative association with a promoter / regulatory sequence, the transcript of the nucleic acid of interest contains an appropriately located ribosome binding site for expression in an appropriate expression system of the polypeptide encoded by the nucleic acid of interest. See for example Sambrook et al. (1989) and Ausubel et al. 25 (1995).

When a nucleic acid is referred to as "isolated", this may mean substantially or completely isolated from some or all other nucleic acid normally present in *A. garbadinensis* and/or *A. liguriae*, especially nucleic acid from outside the gene cluster segments identified herein.

In light of the foregoing disclosure, it will be appreciated that this invention provides nucleotide sequences or a set of nucleotide sequences encoding the actagardine or actagardine B biosynthetic gene cluster. Accordingly, the entire gene cluster or portions thereof of at least twenty-five contiguous nucleotides may be used for a wide variety of applications, including but not limited to: expression of actagardine or actagardine B; use as probes to screen other organisms for related molecules and the like; use to induce gene silencing and the like.

#### **Expression Construct**

In a further aspect of the invention, there is provided an expression construct comprising a nucleic acid encoding a cluster polypeptide or a lantibiotic polypeptide of the invention operably linked to a promoter.

In a further aspect, there is provided a set of expression constructs. A set of expression constructs comprises two or more polypeptide coding sequences of the present invention and at least one promoter suitable for the expression of said sequences. The promoter(s) may be a promoter with which the polypeptide gene is naturally associated with (or in the case of a variant, the promoter of the gene from which the variant is derived), or may be a constitutive or inducible promoter functional in the host cell. Promoters thus include intergenic regions of SEQ ID NO:100 or SEQ ID NO:200 upstream of any of the open reading frames listed in Tables 1 and 2.

The promoter(s) will be operably linked to the nucleic acids of the set of expression constructs. By "operably linked" it will be understood that the promoter will be able to direct transcription of the nucleic acid of interest in an appropriate expression system, with the nucleic acid of

interest in the correct reading frame for translation. Preferably when a nucleic acid of interest is in operative association with a promoter / regulatory sequence, the transcript of the nucleic acid of interest contains an appropriately located ribosome binding site for expression in an appropriate expression system of the polypeptide encoded by the nucleic acid of interest. See  
5 for example Sambrook et al. (1989), Ausubel et al. (2002) and Kieser (2000).

Sets of expression constructs according to the invention include numerous permutations of genes encoding precursor and cluster polypeptides of the invention as defined above. In various aspects of the invention, the set will include at least a *LanA* gene. Examples of such  
10 sets are set out as "Set 1" to "Set 7" below, though these sets should be understood to be merely illustrative and not limiting.

Set 1: A *LanA* gene encoding a precursor polypeptide, preferably a precursor polypeptide capable of being converted to a compound of the invention, plus a *LanM* gene encoding a  
15 *LanM* polypeptide. The *LanM* polypeptide is preferably a *LanM* of SEQ ID NO:120 or a variant thereof, or SEQ ID NO:213 or a variant thereof.

Set 2: A *LanA* gene encoding a precursor polypeptide, preferably a precursor polypeptide capable of being converted to a compound of the invention, plus a *LanR* gene encoding a  
20 *LanR* polypeptide. The *LanR* polypeptide is preferably a *LanR* of SEQ ID NO:122 or a variant thereof, or SEQ ID NO:216 or a variant thereof.

Set 3: A *LanA* gene encoding a precursor polypeptide, preferably a precursor polypeptide capable of being converted to a compound of the invention, plus a *LanM* gene encoding a  
25 *LanM* polypeptide, plus a *LanR* gene encoding a *LanR* polypeptide. The *LanM* polypeptide is preferably a *LanM* of SEQ ID NO:120 or a variant thereof, or SEQ ID NO:213 or a variant thereof. The *LanR* polypeptide is preferably a *LanR* of SEQ ID NO:122 or a variant thereof, or SEQ ID NO:216 or a variant thereof.

Set 4: The genes of Set 3 together with a *LanO* gene encoding a *LanO* polypeptide. The *LanO* polypeptide is preferably SEQ ID NO:122 or a variant thereof, or SEQ ID NO:215 or a  
30 variant thereof.

Set 5: The genes of Set 3 or Set 4 together with a *LanT* gene encoding a *LanT* polypeptide.  
35 The *LanT* polypeptide is preferably SEQ ID NO:123 or a variant thereof, or SEQ ID NO:214 or a variant thereof.

Set 6: The genes of SEQ ID NOs:116 to 127 or variants thereof.

Set 7: The genes of SEQ ID NOs:206 to 220 or variants thereof.

- 5 In one aspect, a set will comprises sequences which all encode 1xx polypeptides or which all encode 2xx polypeptides. However sets which are made up of both 1xx and 2xx polypeptides are not excluded from the present invention.

### **Recombinant Expression Vector**

- 10 In another aspect, there is provided a recombinant vector comprising one or more expression constructs of the invention. In an alternative aspect, there is provided a set of recombinant vectors which comprise a set of expression constructs of the invention. Suitable vectors comprising nucleic acid for introduction into bacteria can be chosen or constructed, containing appropriate additional regulatory elements if necessary, including additional promoters, terminator fragments, enhancer elements, marker genes and other elements as appropriate.
- 15 Vectors may be plasmids, viral e.g. phage, or phagemid, as appropriate. For further details see, for example, Sambrook et al (1989) Molecular Cloning, A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor. Many known techniques and protocols for manipulation of nucleic acid, for example in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and
- 20 analysis of proteins, are described in detail in Ausubel et al. (1995) Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995). Many aspects of the employment of these techniques in the context of *Streptomyces* spp. are described in detail in Hopwood et al (1985) Genetic manipulation of *Streptomyces* a laboratory manual (Norwich: John Innes Foundation) and Practical Streptomyces Genetics (2000) Kieser T. et al., The John Innes
- 25 Foundation p.386. The disclosures of Sambrook et al, Ausubel et al, Hopwood et al and Kieser et al are all incorporated herein by reference for these and all other purposes.

### **Expression Cassettes**

- In another aspect, the inventors have developed a vector system useful for producing and screening lantibiotic derivatives of actagardine B. This is achieved by introducing one or more
- 30 restriction endonuclease recognition sites into the *LanA* gene which encodes SEQ ID NO:1, 11 or 212 in order to produce an expression cassette system. Thus in another aspect, the invention provides a recombinant DNA cassette which comprises a nucleotide sequence encoding an actagardine B precursor polypeptide, wherein said sequence comprises
- a first restriction site at or adjacent the N-terminal encoding region of the encoding
- 35 sequence;

optionally a second restriction site downstream of the first restriction site and within the encoding sequence; and

a third restriction site at or adjacent the C-terminal encoding region of the encoding sequence,

5 wherein at least one of said restriction sites does not occur within the *LanA* coding sequence shown as SEQ ID NO:200.

In a further aspect, there is provided a recombinant DNA cassette which comprises a nucleotide sequence encoding an actagardine precursor polypeptide, wherein said sequence  
10 comprises

a first restriction site at or adjacent the N-terminal encoding region of the encoding sequence;

optionally a second restriction site downstream of the first restriction site and within the encoding sequence; and

15 a third restriction site at or adjacent the C-terminal encoding region of the encoding sequence,

wherein at least one of said restriction sites does not occur within the *LanA* coding sequence shown as SEQ ID NO:100.

20 Generally, all two or three sites will be different from each other. It is also desirable that when the cassette is carried by a vector, the sites are unique for that vector.

In a preferred aspect, the non-naturally occurring restriction enzyme site is the second restriction site and is located between codons 5 and 16, such as between 6 and 15, of the  
25 encoding sequence of SEQ ID NO:1 or SEQ ID NO:4.

The cassette will desirably also include a *LanA* leader sequence and a *LanA* promoter, and may include in addition one or more cluster genes, particularly where such a cluster gene is required to complement the loss of the equivalent host cell gene.

30 The cassette of the invention described above may be engineered in a variety of ways. For example, the fragment obtained by cleaving the cassette between the first and second, first and third, or second and third, restriction sites may be replaced with a variant coding sequence encoding a .lantibiotic A variant. Thus the invention provides a variant of the  
35 cassette of the invention wherein said variant has from 1 to 15 nucleotide substitutions within the encoding region of the encoding sequence.

As an intermediate to the production of such a variant, the sequence of between the first and second, first and third, or second and third, restriction sites may be replaced by a larger stuffer fragment.

- 5 In another aspect, the cassette encoding a lantibiotic derivative may be used to transform a host cell to express the derivative, for example to assess its anti-bacterial properties.

In one aspect, a multiplicity of expression cassettes may be made to provide a library of different derivatives, which may then be screened for activity.

- 10 An expression cassette of the invention may be based on any cloning and expression vector used in the art for the expression of genes in host cells. Such vectors will include one or more origins of replication, which may be temperature sensitive. The vectors may include a selectable marker, such as the chloramphenicol acetyl transferase gene, the erythromycin resistance gene, the apramycin resistance gene or the tetracycline resistance gene. The vector may also contain a targeting region, this region being homologous to a genomic sequence present in the host cell outside the *LanA* gene cluster. Such a vector may be used to integrate the cassette into the genomic sequence homologous to the targeting region.

- 20 The expression cassette may also comprise one or more cluster genes in addition to the *LanA* gene or derivative thereof. Where the host cell is a  $\Delta$ *LanA* host cell in which the *LanA* gene has been inactivated in a manner which also inactivates such a cluster gene (e.g. in the strain disclosed in Altena et al, 2000), it is desirable that this gene or an equivalent gene is provided on the expression cassette.

- 25 As used herein, by "at or adjacent the N-terminal encoding region" it is meant that the first base of the restriction site is located at a position from six residues upstream of the ATG codon of the *LanA* leader sequence to no more than six codons downstream of the first codon of the propeptide. Preferably the first base of the restriction site is located at a position from twelve, preferably six, residues upstream to six residues downstream of the first codon of the propeptide encoding sequence.

In one aspect, the first restriction site is a *Bgl*II site.

- 35 Similarly, by "at or adjacent the C-terminal encoding region" it is meant that the first base of the restriction site either includes at least one of the nucleotides of the termination codon of

the propeptide or the 5' or 3' nucleotide of the restriction site is no more than twelve, preferably six, residues downstream or upstream respectively of the termination codon.

In one aspect, the third restriction site is a *AvrII* site.

5

The second restriction site, when present, will lie between the first and third restriction sites. Preferably the restriction site includes at least one nucleotide present from codon 5 to codon 16, preferably codon 8 to 16 of the propeptide-encoding sequence. In the accompanying examples, a *BsrGI* site has been introduced by altering codons 6 and 7 of the *ActA*-encoding sequence. However, other changes are also contemplated by the present invention.

10

It is also possible to introduce more than one change such that the expression cassette includes two or more sites between the first and third restriction sites.

15 The cassette may include two or three non-naturally occurring restriction sites. In the accompanying example, all three sites do not normally occur in the *ActA* sequence encoded by of SEQ ID NO:100.

20 The expression cassette simplifies the rapid production of lantibiotic derivatives, as discussed further herein below.

In one aspect, the region between the first and second sites, the first and third, or the second and third sites, may be replaced by a stuffer fragment. Where two or more sites between the first and third sites are present, the region between any pair of such sites may also be replaced by a stuffer fragment. A stuffer fragment is a piece of DNA which is larger than the sequence which it replaces. The stuffer fragment may be from 50 to 5000 nucleotides in size, for example from about 500 to 2000 nucleotides in size. The value of introducing these stuffer DNA fragments is that when the region is replaced by a lantibiotic-encoding oligonucleotide there is a significant decrease in plasmid size. The resulting plasmid can thus be readily purified away from any minor population of unrestricted plasmid thus eliminating any background.

25

30

A cassette of the invention may be used to introduce specific changes to the *ActA* sequence in a vector which can then be introduced into a host cell for expression of a lantibiotic. To achieve this, the sequence is desirably operably linked to the *LanA* (e.g. *ActA* or *LigA*) leader sequence, which in turn is operably linked to the *LanA* promoter (e.g. *ActA* or *LigA*)

35

In addition or as an alternative, the vector comprising the cassette may also include a *LanR* gene. The *LanR* gene will be located downstream of, and in tandem with, the lantibiotic A coding sequence.

### ***Expression Libraries***

5 Expression cassettes of the invention may be used to provide libraries of lantibiotic-encoding genes. Such libraries may be made by introducing into the cassette, between the first and second restriction sites, the first and third restriction sites, or the second and third restriction sites, a multiplicity of sequences each of which corresponds to the corresponding *ActA* or *LigA* sequence apart from having from 1 to 15, for example from 1 to 10, preferably from 1 to 6, for  
10 example from 1 to 3 nucleotide changes compared to the propeptide portion of SEQ ID NOs:100 or 200. Preferably such changes result in a change of the protein encoded by the sequence. However non-coding changes are not excluded.

Libraries form a further aspect of the invention. Such libraries may comprise from 10 to  
15 100,000, such as from 10 to 10,000 for example from 10 to 1,000 different coding sequences which are variants of the lantibiotic A coding sequence of an expression cassette.

An expression cassette encoding a lantibiotic A derivative may be introduced into a host cell for expression of the lantibiotic.

20 In one embodiment, the library may be transformed into host cells, and colonies isolated and/or screened for antibacterial activity. The sequences of the lantibiotic A expressed by individual colonies showing such activity can be determined. Where the lantibiotic A shows activity, the invention further provides a lantibiotic obtained by the method of the invention.

### ***Host Cell***

25 Two main types of host cells are envisaged by the present invention. The first type of host cell is a lantibiotic producing host cell. Alternatively the host cell may be a non-producer cell, i.e. does not contain a *LanA* gene or its associated cluster genes required for producing a *LanA* polypeptide.

30 In one embodiment, the invention provides a host cell transformed with a set of expression constructs of the invention. The set of constructs may be any one of Sets 1 to 7 as defined above, or a set based upon any other combination of precursor and cluster polypeptide-encoding nucleic acids. In another embodiment, the host cell may be transformed with a  
35 expression cassette of the invention.

In a further embodiment, there is provided a library of host cells, each one comprising a different expression cassette of the invention.

A Lantibiotic-producing host cell.

- 5 In one embodiment, the host cell may be a lantibiotic producing host cell. A lantibiotic producing host cell is one in which an expression construct comprising a *LanA* gene, if introduced into the cell in the absence of any cluster gene, would be expressed and a LanA polypeptide produced. Such cells include any type-B lantibiotic producing cell such as any strain of a bacillus, an actinomycete, or a streptomycete, (e.g. *S. lividans* or *S. coelicolor*)  
10 which produces a lantibiotic. Examples of such cells include a cinnamycin-producing host cell (*Streptomyces cinnamoneus cinnamoneus* DSM 40005), or an actagardine-producing *Actinoplanes garbadinensis* or *A. liguriae* NCIMB 41362.

- Where the invention relates to the productions of compounds of the formula (I) in which -X1-  
15 X2- represent -Leu-Val-, the host cell may be *A. liguriae* NCIMB 41362 without any further modification.

- In one aspect, a host cell of this class may comprise a mutation in its endogenous *LantibioticA* gene such that the gene is not expressed or the gene product is inactive. Such a host cell  
20 may be obtained by targeted homologous recombination to delete or mutate the *LanA* gene of the host cell. Methods to achieve this are known as such in the art and are illustrated in Altena *et al*, (2000) and WO2005/093069, the disclosures of which are incorporated herein by reference. The resulting host cell is referred to as a  $\Delta$ *LanA* host cell. In one particular embodiment, the host cell is a  $\Delta$ *LigA* *A. liguriae* NCIMB 41362 host cell in which the *ligA* gene  
25 has been inactivated, for example by mutation or deletion, e.g. deletion brought about by homologous recombination. In another embodiment, the host cell is a  $\Delta$ *ActA* *A. garbadinensis* host cell in which the *ActA* gene has been inactivated, for example by mutation or deletion, e.g. deletion brought about by homologous recombination.

- 30 The transformation of a host cell of this type with other cluster genes is also contemplated by the present invention, though where the host cell provides cluster genes necessary for the production of a lanA, the provision of such cluster genes is optional.

Non-producer cell

- A non-producer cell may be any host cell in which expression of a *LanA* gene encoding a  
35 precursor polypeptide capable of being converted to actagardine or a variant thereof, or to a compound of the invention, can produce such a product provided the *LanA* gene is introduced

into the cell as part of a set of expression constructs which are capable of converting a precursor polypeptide to actagardine or a variant thereof, or to a compound of the invention.

A non-producer host cell may be a bacterial host cell. Bacterial host cells include an actinomycete, or a streptomycete, e.g. *S. lividans*, *S. coelicolor* or *S. cinnamoneus*.

Host cells may be those in which the *lanO* gene is inactivated by mutation or deletion (or in the case of non-producer cells, not present), or those in which the expression of the *lanO* gene is increased, e.g. by provision of two or more copies of the gene or by linking the gene to a promoter which enhances expression in the host cell. Modulation of the *lanO* gene in this manner may be desirable to alter the relative levels of oxidized (Y = -S(O)) and reduced (Y = -S-) forms of compounds of the invention produced in the host cell.

### ***Production of Compounds of the Invention***

Compounds of the invention may be produced by expression of a nucleic acid, for example in the form of an expression construct encoding a precursor polypeptide carried in a recombinant expression vector, in a host cell which carries a *LanA* gene together with where necessary, associated cluster genes required for conversion of the precursor polypeptide to the product. As noted above, where the invention relates to the productions of compounds of the formula (I) in which -X1-X2- represent -Leu-Val-, the host cell may be *A. liguriae* NCIMB 41362 without any further modification.

The introduction of the expression cassette, or vector(s) into a host cell may (particularly for *in vitro* introduction) be generally referred to without limitation as "transformation". This may employ any available technique. For bacterial cells, suitable techniques may include calcium chloride transformation, polyethyleneglycol assisted transformation, electroporation, conjugation and transfection or transduction using bacteriophages.

In one aspect, the present invention provides a method of expressing nucleic acid of the invention, the method comprising providing a host cell (or other expression system) culturing the host cell, so as to express the nucleic acid of interest. The nucleic acid of interest will be in an expression cassette, such that culturing the host cell leads to the production of a product of the invention.

Preferably the nucleic acid of interest is expressed substantially only when the host cell culture reaches high cell density, more preferably at or close to the stationary phase of host cell culture. Cell cultures at or close to stationary phase may have OD650 values in the range of 1-

20. Known methods of culturing cells are well known in the art, for example from Sambrook et al (1989), Ausubel et al (2002), and (in particular for *Streptomyces* spp.) Kieser et al (2000).

The expression products of the expression systems may be collected and purified. Isolation methods may comprise capture from the fermentation medium using solvent extraction

5 techniques, adsorption resin such as hydrophobic resins or precipitation methods such as ammonium sulfate precipitation. Purification methods may include chromatography techniques such as ion exchange, hydrophobic interaction, reverse phase, normal phase, solid phase extraction and HPLC, e.g. as described in US 5,112,806 for the isolation of mersacidin

10 Following culture of the cell, the compounds of the invention may be recovered from the host cell culture. The recovered compounds may be formulated in the form of a pharmaceutical composition, optionally in the form of a pharmaceutically acceptable salt.

Where host cells produce a mixture of compounds of the invention, e.g. those in which Y is -S-  
15 or -S(O)- or those in which Z is NH<sub>2</sub> or Ala-, or mixtures of all four types, the products may be isolated using standard separation techniques such as hplc, e.g. as described in US 6,022,851 for the production of Actagardine and Ala-Actagardine.

The recovered compounds may be formulated in the form of a pharmaceutical composition,  
20 optionally in the form of a pharmaceutically acceptable salt.

#### ***Pharmaceutically acceptable salt***

A "pharmaceutically acceptable salt", may be an acid addition salt in which the base retains the biological effectiveness and properties of the compound and which is physiologically acceptable. Such salts include those formed with inorganic acids such as hydrochloric acid,  
25 hydrobromic acid, sulphuric acid, nitric acid, phosphoric acid and the like, and organic acids such as acetic acid, propionic acid, glycolic acid, pyruvic acid, oxalic acid, malic acid, malonic acid, succinic acid, maleic acid, fumaric acid, tartaric acid, citric acid, benzoic acid, cinnamic acid, mandelic acid, methanesulfonic acid, ethanesulfonic acid, p-toluenesulfonic acid, salicylic acid and the like.

30 Salts also include basic salts, such as an alkali or alkaline earth metal salt, e.g. a sodium, potassium, calcium or magnesium salt.

#### ***Pharmaceutical Compositions***

The lantibiotics of the present invention may be formulated together with one or more other  
35 pharmaceutically acceptable ingredients well known to those skilled in the art, including, but

not limited to, pharmaceutically acceptable carriers, adjuvants, excipients, diluents, fillers, buffers, preservatives, anti-oxidants, lubricants, stabilizers, solubilisers, surfactants (e.g., wetting agents), masking agents, colouring agents, flavouring agents, and sweetening agents. The formulation may further comprise other active agents, for example, other therapeutic or prophylactic agents. Thus, the present invention further provides pharmaceutical compositions, as defined above, and methods of making a pharmaceutical composition comprising admixing at least one active compound, as defined above, together with one or more other pharmaceutically acceptable ingredients well known to those skilled in the art, e.g., carriers, adjuvants, excipients, etc. If formulated as discrete units (e.g., tablets, etc.), each unit contains a predetermined amount (dosage) of the active compound.

The term "pharmaceutically acceptable" as used herein pertains to compounds, ingredients, materials, compositions, dosage forms, etc., which are, within the scope of sound medical judgement, suitable for use in contact with the tissues of the subject in question (e.g., human) without excessive toxicity, irritation, allergic response, or other problem or complication, commensurate with a reasonable benefit/risk ratio. Each carrier, adjuvant, excipient, etc. must also be "acceptable" in the sense of being compatible with the other ingredients of the formulation.

Compositions may be formulated for any suitable route and means of administration. Pharmaceutically acceptable carriers or diluents include those used in formulations suitable for oral, rectal, nasal, topical (including buccal and sublingual), vaginal or parenteral (including subcutaneous, intramuscular, intravenous, intradermal, intrathecal and epidural) administration. The formulations may conveniently be presented in unit dosage form and may be prepared by any of the methods well known in the art of pharmacy. Such methods include the step of bringing into association the active ingredient with the carrier which constitutes one or more accessory ingredients. In general the formulations are prepared by uniformly and intimately bringing into association the active ingredient with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

For solid compositions, conventional non-toxic solid carriers include, for example, pharmaceutical grades of mannitol, lactose, cellulose, cellulose derivatives, starch, magnesium stearate, sodium saccharin, talcum, glucose, sucrose, magnesium carbonate, and the like may be used. The active compound as defined above may be formulated as suppositories using, for example, polyalkylene glycols, acetylated triglycerides and the like, as the carrier. Liquid pharmaceutically administrable compositions can, for example, be prepared by dissolving, dispersing, etc, an active compound as defined above and optional

pharmaceutical adjuvants in a carrier, such as, for example, water, saline aqueous dextrose, glycerol, ethanol, and the like, to thereby form a solution or suspension. If desired, the pharmaceutical composition to be administered may also contain minor amounts of non-toxic auxiliary substances such as wetting or emulsifying agents, pH buffering agents and the like, for example, sodium acetate, sorbitan monolaurate, triethanolamine sodium acetate, sorbitan monolaurate, triethanolamine oleate, etc. Actual methods of preparing such dosage forms are known, or will be apparent, to those skilled in this art; for example, see "Remington: The Science and Practice of Pharmacy", 20th Edition, 2000, pub. Lippincott, Williams & Wilkins. The composition or formulation to be administered will, in any event, contain a quantity of the active compound(s) in an amount effective to alleviate the symptoms of the subject being treated.

Dosage forms or compositions containing active ingredient in the range of 0.25 to 95% with the balance made up from non-toxic carrier may be prepared.

For oral administration, a pharmaceutically acceptable non-toxic composition is formed by the incorporation of any of the normally employed excipients, such as, for example, pharmaceutical grades of mannitol, lactose, cellulose, cellulose derivatives, sodium crosscarmellose, starch, magnesium stearate, sodium saccharin, talcum, glucose, sucrose, magnesium, carbonate, and the like. Such compositions take the form of solutions, suspensions, tablets, pills, capsules, powders, sustained release formulations and the like. Such compositions may contain 1%-95% active ingredient, more preferably 2-50%, most preferably 5-8%.

Parenteral administration is generally characterized by injection, either subcutaneously, intramuscularly or intravenously. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol or the like. In addition, if desired, the pharmaceutical compositions to be administered may also contain minor amounts of non-toxic auxiliary substances such as wetting or emulsifying agents, pH buffering agents and the like, such as for example, sodium acetate, sorbitan monolaurate, triethanolamine oleate, triethanolamine sodium acetate, etc.

For topical applications, the pharmaceutically acceptable compositions may be formulated in a suitable ointment or gel containing the active component suspended or dissolved in one or more carriers. Carriers for topical administration of the compounds of this invention include, but are not limited to, mineral oil, liquid petrolatum, white petrolatum, propylene glycol,

polyoxyethylene, polyoxypropylene compound, emulsifying wax and water. Alternatively, the pharmaceutically acceptable compositions can be formulated in a suitable lotion or cream containing the active components suspended or dissolved in one or more pharmaceutically acceptable carriers. Suitable carriers include, but are not limited to, mineral oil, sorbitan monostearate, polysorbate 60, cetyl esters wax, cetearyl alcohol, 2-octyldodecanol, benzyl alcohol and water.

The percentage of active compound contained in such parental or topical compositions is highly dependent on the specific nature thereof, as well as the activity of the compound and the needs of the subject. However, percentages of active ingredient of 0.1% to 10% w/w employable, and will be higher if the composition is a solid which will be subsequently diluted to the above percentages. Preferably, the composition will comprise 0.2-2% w/w of the active agent in solution.

Further teaching regarding suitable carriers, adjuvants, excipients, etc. can be found in standard pharmaceutical texts, for example, Remington: The Science and Practice of Pharmacy", 20th Edition, 2000, pub. Lippincott, Williams & Wilkins; and Handbook of Pharmaceutical Excipients, 2nd edition, 1994.

### **Administration of Compounds**

Lantibiotic compounds and compositions of the invention may be administered to a subject in a method of medical treatment or prophylaxis. The subject may be a human or animal subject. The animal subject may be a mammal, or other vertebrate.

Thus there is provided a compound of the invention for use in a method of treatment or prophylaxis of a subject. There is also provided use of a compound of the invention for the manufacture of a medicament for use in a method of treatment or prophylaxis of a subject.

The method of treatment may be of a bacterial infection, including a skin, mucosal, enteric or systemic infection.

The variants and composition may be used for systemic treatment of bacteraemia (including catheter related bacteraemia), pneumonia, skin and skin structure infections (including surgical site infections), endocarditis and osteomyelitis. These and other such treatments may be directed against causative agents such as staphylococci, streptococci, enterococci. The compounds of the invention or compositions thereof may also be used for topical treatment of skin infections including acne ie. *Propionibacterium acnes*. The variants and compositions

thereof may also be used in the treatment of eye infections, such as conjunctivitis, and for oral treatment for gut super-infection, such as that caused by *Clostridium difficile* including multiply-resistant *C. difficile* (pseudomembranous colitis), or gut infections associated with *Helicobacter pylori*.

5 The variants may also be used in the treatment or prevention of infection of the skin in wounds or burns. In addition, the variants and compositions thereof may be used in prophylactic methods, such as for the clearance of the nares to prevent transmission of MRSA. This may be practiced on subjects at risk of infection (e.g. patients entering a hospital) or on health  
10 professionals or other carers at risk of being carriers of such infections. Prophylactic clearance of gut flora ahead of abdominal surgery is also contemplated.

The compounds according to the invention can be administered enterally (orally), parenterally (intramuscularly or intravenously), rectally or locally (topically). They can be administered in  
15 the form of solutions, powders (tablets, capsules including microcapsules), ointments (creams or gel), or suppositories. Possible auxiliaries for formulations of this type are the pharmaceutically customary liquid or solid fillers and extenders, solvents, emulsifiers, lubricants, flavor corrigents, colorants and/or buffer substances. As an expedient dose, 0.1-1000, preferably 0.2-100, mg/kg of body weight are administered. They are expediently  
20 administered in dose units which contain at least the efficacious daily amount of the compounds according to the invention, e.g. 30-3000, preferably 50-1000, mg.

The experimental basis of the present invention, including its best mode, will now be further described in detail, by way of example only, with reference to the accompanying drawings.

#### 25 EXAMPLE 1 - Cloning of Gene Clusters

Identifying and cloning the actagardine biosynthetic gene clusters from *A. garbadinensis* and *A. liguriae*.

30 O/SBDIG-1 is a digoxigenin (DIG)-labelled degenerate oligonucleotide composed of 48 bases. It was designed by translating the known amino acid sequence of actagardine and considering codon usage for Actinoplanes. Southern hybridisation analysis of genomic DNA isolated from *A. garbadinensis* and digested using the restriction enzyme *Nco*I, identified a ~3kb fragment which hybridised to O/SBDIG-1. The *Nco*I digest of the genomic DNA was repeated and DNA  
35 fragments of ~3kb were isolated and cloned into *Nco*I cut pLITMUS28 (NEB). The resulting plasmids were introduced into *E. coli* DH10B cells and then analysed by Southern

hybridisation using the probe O/SBDIG-1. A hybridising clone was identified and submitted for sequence analysis. Sequencing revealed that this plasmid designated pLITAG01 consists of DNA encoding the *lanA* structural gene for actagardine biosynthesis (*actA*) together with an upstream region believed to encode a portion of an ABC sugar transporter and a region downstream partially encoding lan M (*actM*).

The primers O/ACT08F and O/ACT09R were designed based upon sequence from pLITAG01. Using these primers in a PCR reaction together with DIG-labelled dNTPs (Roche) and pLITAG01 as a template, a 2296bp DIG-labelled DNA fragment was generated and designated SBDIG-2.

Two cosmid libraries were generated by cloning *Sau3AI* digested genomic DNA from *A. garbadinensis* ATCC 31049 and *A. liguriae* NCIMB 41362, into the cosmid SuperCos1 (Stratagene) previously digested using *BamHI*. Each cosmid library was analysed by Southern hybridisation using SBDIG-1. Twenty-five cosmids from each library believed to hybridise to SBDIG-1 were selected and re-analysed via Southern hybridisation using the probes O/SBDIG-1 and SBDIG-2. Nine cosmids derived from genomic DNA from *A. garbadinensis* and eleven cosmids derived from genomic DNA from *A. liguriae* hybridised to both probes. DNA was isolated from each cosmid and sequenced using the primers T3 and T7. The cosmids CosAL02 and CosAG14 were subsequently fully sequenced (Sequencing facility, Department of Biochemistry, University of Cambridge).

## **Materials and Methods**

### Strains

Bacterial strains used in the present invention are summarised in Table 5.

### Vectors

The cosmid SuperCos1 was obtained from Stratagene.

The plasmid pLITMUS was obtained from New England Biolabs.

Primers

| Primer name | SEQ. ID | Sequence 5'-3'                                       |
|-------------|---------|------------------------------------------------------|
| O/SBDIG-1   | 300     | TGGGTSTGCACSCTSACSATCGARTGCGGNACSG<br>TSATCTGCGCSTGC |
| O/ACT08F    | 301     | TCCAGCACGCGCGGGG                                     |
| O/ACT09R    | 302     | GTTTCGACCAGCCGCC                                     |

**Southern hybridisation**Labelling of DNA probe

DNA hybridisation probes were prepared using the Digoxigenin (DIG) PCR DNA labeling and  
5 detection kit supplied by Roche, according to their instructions.

Transfer of DNA to nylon membrane

The DNA of interest was initially separated by agarose gel electrophoresis and transferred to a  
nylon membrane (Hybond-N, Amersham Int., UK) using a vacuum blotter (Q BIO gene). The  
time taken for depurination of the DNA using 0.5 M HCl was judged by the time taken for the  
10 bromophenol blue marker band to turn completely yellow (typically 15-20 min for a 0.7%  
agarose gel). The DNA was then systematically denatured with 1.5 M NaCl, 1.5 M NaOH and  
then neutralised using 0.5 M Tris, 1.5 M NaCl, pH 8.0 for a further 15-20 min each. Complete  
blotting of the DNA was facilitated by flooding with 20 x SSC for a minimum of 60 min. After  
removing the blotted membrane from the vacuum it was left to air dry at room temperature.

15 The DNA was cross-linked by placing the membrane (DNA face down) on a UV  
transilluminator (UVP) and exposing it to UV at a wavelength of 365 nm for 5 min.

Colony lifts and hybridisation

Colonies to be screened by hybridisation were transferred onto a nylon membrane (Roche  
diagnostics). This was achieved by placing the positively charged nylon membrane over the  
20 colonies and pressing firmly for 1 min to ensure effective transfer. Reference points were  
marked on the membrane to indicate its orientation with respect to the colonies. Following this,  
the membrane was removed and prepared for hybridisation as directed in the Roche user's  
manual (DIG Application manual for filter hybridisation).

Hybridisation and development of membranes

25 DNA was hybridised with the prepared probe overnight (~16hr) at 68°C in a HB-1000  
hybridisation oven (UVP). Following hybridisation the membrane was washed for 2 periods of  
5 min at room temperature using 2 x salt sodium citrate (SSC) + 0.1% sodium dodecyl  
sulphate (SDS). These washes were followed with a second series of 2 x 15 min washes at  
68°C using 1 x SSC + 0.1% SDS for the membrane hybridised in the presence of SBDIG-1  
30 and 0.1 x SSC + 0.1% SDS for the membrane screened using SBDIG-2. Membranes were

then developed as recommended in the Roche user's manual (DIG Application manual for filter hybridisation).

### **Software**

5 The consensus sequences were analysed using FramePlot version 2.3.2, BioEdit sequence alignment editor, ClustalW (EMBL-EBI) and the Basic Local Alignment Search Tool (BLAST, NCBI).

### **Results and discussion**

#### **CosAG14**

10 The cosmid, CosAG14, contains a 38168bp fragment of genomic DNA isolated from *A. garbadinensis*. Sequence analysis has identified DNA encoding the leader and actagardine prepeptide, this gene has been assigned the name *actA*. Two alanine residues lie immediately upstream of the actagardine prepeptide. These residues are believed to represent the recognition site for protease cleavage of the leader peptide from actagardine. Partial cleavage at this position resulting in the retention of an alanine is thought to result in the production of  
15 ala-actagardine routinely observed in fermentation broths of *A. garbadinensis*. Downstream of the *actA* gene lies a 3162bp region of DNA with strong sequence similarity to several lanM proteins, for example, *mrsM* (30% identity) from the mersacidin gene cluster. This putative gene has been designated *actM* and is thought to be involved in the modification of the actagardine prepeptide, catalysing dehydration and thioether formation. An open reading  
20 frame designated *actO*, that is located 11bp downstream of *actM* encodes a 341-amino-acid protein with sequence similarity (~39% identity) to several luciferase-type monooxygenases. The role of the monooxygenase, ActO, is believed to be to catalyse the incorporation of oxygen generating actagardine from deoxy-actagardine. In reverse orientation to *actO* and located 62bp downstream is the open reading frame named *actR*. The protein product of this  
25 orf shows sequence similarity (~37% identity) to several two-component response regulators. Positioned 789bp downstream and in the same orientation to *actR* lies a putative 812 amino-acid protein that shows sequence similarity (~25% identity) to ABC transporter permeases. This putative protein designated *actT* is potentially responsible for the export of the modified lantibiotic from the cell. The amino acid sequences of the second and fourth orf downstream of  
30 *actT* show similarity (~30% identity) to response regulator kinases and penicillin binding proteins respectively. Recent work on the nisin biosynthetic gene cluster in *Listeria monocytogenes* (Gravesen et al., 2004) has demonstrated that a histidine kinase together with a penicillin-binding protein and protein of unknown function are involved in conferring nisin resistance. The presence of analogous genes within close proximity to the *actA* may indicate  
35 that these genes are involved in an actagardine resistance mechanism.

The cosmid CosAL02 contains a 40402bp fragment of genomic DNA isolated from *Actinoplanes liguriae*. Sequence analysis has identified a *lanA* gene encoding a 64-amino-acid protein with strong sequence homology (50 identical residues) to the *actA* gene identified in the cosmid CosAG14. We have termed this species of *lanA* gene as *ligA*. The amino acid sequence of the prepeptide of this *lanA* differs from that of actagardine by two residues indicated in the alignment of the two genes shown below (SEQ. ID 119 and SEQ. ID 212):

\* \* \* : : \* . : \*\* : . \*\*\* : . \* : \*\*\*\*\* : \*\*\*\*\* : : \*

Positioned downstream of *ligO* and in reverse orientation lies a putative 217 amino-acid protein that shows sequence similarity (~37%) to several two-component response regulators. This putative regulator has been designated *ligR*.

**Table 1 – Annotation of CosAG14 (38168bp fragment isolated from *A. garbadinensis*.  
The SuperCos1 vector backbone sequence is omitted)**

| Gene  | Description                               | Position (DNA) | Frame | Size AAs (bp) | Start – end (3AA) |
|-------|-------------------------------------------|----------------|-------|---------------|-------------------|
| orf1  | Hypothetical protein                      | 482 – 42       | -1    | 146 (441)     | MPR - RCG         |
| orf2  | Hypothetical protein                      | 2824 - 2375    | -2    | 149 (450)     | VSV - ERA         |
| orf3  | ATPase AAA involved in cell division      | 4432 – 2876    | -2    | 518 (1557)    | VER - TNR         |
| orf4  | Sugar hydrolase                           | 6002 - 5391    | -1    | 203 (612)     | VGE - NYS         |
| orf5  | Endoglucanase                             | 6484 – 5825    | -2    | 219 (660)     | MRR - TVR         |
| orf6  | Cytosine/adenine deaminase                | 6627 – 7112    | +3    | 161 (486)     | MTI - PAQ         |
| orf7  | Unknown                                   | 7756 – 8997    | +1    | 413 (1242)    | VTT - YDK         |
| orf8  | Unknown                                   | 9586 – 8933    | -2    | 217 (654)     | VGK - FRG         |
| orf9  | Pyruvate oxidase                          | 11886 – 10108  | -3    | 592 (1779)    | VSD - DPS         |
| orf10 | Hydrolase or acyltransferase              | 12066 – 12866  | +3    | 266 (801)     | VSR - SGT         |
| orf11 | Aldose epimerase                          | 13116 – 14306  | +3    | 396 (1191)    | MTE - TAD         |
| orf12 | ABC sugar transport periplasmic component | 14385 - 15521  | +3    | 378 (1137)    | MPR - AHG         |
| orf13 | ABC sugar transport permease              | 15514 – 16572  | +1    | 352 (1059)    | MDD - GRS         |
| orf14 | ABC transport protein. ATP-binding        | 16569 - 17330  | +3    | 253 (762)     | MTA - RGR         |
| orf15 | Hypothetical protein. Methyltransferase   | 18102 – 17335  | -3    | 255 (768)     | MES - RKR         |
| orf16 | ABC transport permease                    | 18962 - 18120  | -1    | 280 (843)     | MPP - RKG         |
| orf17 | ABC transport permease                    | 19896 – 18991  | -3    | 301 (906)     | MSA – ESE         |
| orf18 | ABC transport substrate binding           | 21236 – 19899  | -1    | 445 (1338)    | MFI – SGR         |
| orf19 | actA, structural gene                     | 21572 - 21766  | +2    | 64 (195)      | MSA - CAC         |
| orf20 | actM, modification gene                   | 21837 - 24998  | +3    | 1053 (3162)   | MSP - PLT         |
| orf21 | actO, monooxygenase                       | 25009 - 26034  | +1    | 341 (1026)    | MPE - PAA         |
| orf22 | actR, response regulator                  | 26791 - 26096  | -2    | 231 (696)     | MRS - CLS         |
| orf23 | actT, ABC transporter associated permease | 29323 - 26885  | -2    | 812 (2439bp)  | MLA - LTR         |

| Gene  | Description                      | Position<br>(DNA) | Frame | Size AAs<br>(bp) | Start – end<br>(3AA) |
|-------|----------------------------------|-------------------|-------|------------------|----------------------|
| orf24 | Hypothetical protein             | 29462 - 30196     | +2    | 244 (735)        | MIV - RNR            |
| orf25 | Reponse regulator kinase         | 30235 - 31338     | +1    | 367 (1104)       | VLR - ARA            |
| orf26 | Response regulator sensor        | 31335 – 31997     | +3    | 220 (663)        | MTR - AVG            |
| orf27 | Penicillin binding protein       | 32138 - 34486     | +2    | 782 (2349)       | MLI – PPR            |
| orf28 | Methyltransferase                | 35209 - 34448     | -2    | 253 (762)        | MAP - DRR            |
| orf29 | Hydrolase                        | 36030 - 35245     | -3    | 261 (786)        | VPR - PPP            |
| orf30 | Response regulator               | 36086 - 36820     | +2    | 244 (735)        | VSP - TGS            |
| orf31 | Fructose biphosphate<br>aldolase | 36844 - 37689     | +1    | 281 (846)        | MKD - RAW            |
| orf32 | Hydrolase                        | 37590 – 38168     | +3    | 192 (579)        | MGS - DPA            |

**Table 2 – Annotation of CosAL02 (40402bp fragment isolated from *A. liguriae*. The SuperCos1 vector backbone sequence is omitted).**

| Gene  | Description                          | Position (DNA) | Frame | Size AAs (bp) | Start-end (3AA) |
|-------|--------------------------------------|----------------|-------|---------------|-----------------|
| orf1  | Secretion system protein             | 1008 – 1       | -2    | 335 (1008)    | VRL – VDI       |
| orf2  | Response regulator                   | 2198 – 1122    | -3    | 358 (1077)    | MSE – LFP       |
| orf3  | Hypothetical protein                 | 3088 – 2288    | -1    | 266 (801)     | MRR – VVR       |
| orf4  | Hypothetical protein                 | 4410 – 3112    | -2    | 432 (1299)    | MRR – RTG       |
| orf5  | Response regulator ATP-binding       | 5205 – 4795    | -2    | 136 (411)     | MWK – SAR       |
| orf6  | ABC sugar transporter                | 5516 – 6607    | +2    | 363 (1092)    | MFN – SAY       |
| orf7  | ABC sugar transporter ATP-binding    | 6673 – 8178    | +1    | 501 (1506)    | MLL – DEH       |
| orf8  | ABC transport permease               | 8168 - 9127    | +2    | 319 (960)     | MST – RTR       |
| orf9  | ABC transporter permease protein     | 9130 – 10092   | +1    | 320 (963)     | MSI – RRS       |
| orf10 | Metallopeptidase                     | 12046 – 10586  | -1    | 486 (1461)    | MRT – PGS       |
| orf11 | Putative StrR-like regulator         | 12460 – 13320  | +1    | 286 (861)     | MDS – DAA       |
| orf12 | ligA                                 | 13641 - 13835  | +3    | 64 (195)      | MSA – CAC       |
| orf13 | ligM                                 | 13907 - 17047  | +2    | 1046 (3141)   | MSS – THV       |
| orf14 | ligT, ABC transporter                | 17040 - 18767  | +3    | 575 (1728)    | MSE – LLT       |
| orf15 | ligO, Luciferase type monooxygenase  | 18785 – 19828  | +2    | 347 (1044)    | MLS –RRW        |
| orf16 | ligR, Response regulator             | 20459 – 19806  | -3    | 217 (654)     | MAD- ELA        |
| orf17 | ABC-transporter associated permease  | 23069 – 20625  | -3    | 814 (2445)    | MIF – LVR       |
| orf18 | ABC-transporter. ATP-binding protein | 23788 – 23066  | -1    | 240 (723)     | MVS – VTS       |
| orf19 | Histidine kinase                     | 23980 - 25068  | +1    | 362 (1089)    | VIA – AVP       |
| orf20 | Response regulator                   | 25065 – 25721  | +3    | 218 (657)     | MTE – GPS       |

| Gene  | Description                               | Position (DNA) | Frame | Size AAs (bp) | Start-end (3AA) |
|-------|-------------------------------------------|----------------|-------|---------------|-----------------|
| orf21 | Putative membrane protein                 | 26673 – 25768  | -2    | 301 (906)     | MPI – RFP       |
| orf22 | alpha-beta hydrolase                      | 26697 – 27569  | +3    | 290 (873)     | MRN – ASR       |
| orf23 | Transcriptional regulator                 | 27574 – 28011  | +1    | 145 (438)     | VRL – RLG       |
| orf24 | Pyruvoyl-dependent arginine decarboxylase | 28102 – 28629  | +1    | 175 (528)     | MAD – GMN       |
| orf25 | Putative diaminopimelate decarboxylase    | 30946 – 29626  | -2    | 406 (1221)    | MTL – LYA       |
| orf26 | Kinase                                    | 31860 – 30931  | -2    | 309 (930)     | VRS – PDL       |
| orf27 | Transcriptional regulator                 | 33248 – 32145  | -3    | 367 (1104)    | VVF – ANS       |
| orf28 | Glycosyl transferase                      | 33600 – 34553  | +3    | 317 (954)     | MPS – NAG       |
| orf29 | Glycosyl transferase                      | 34543 – 35652  | +1    | 369 (1110)    | MPA – ARV       |
| orf30 | Dihydrolipoamide dehydrogenase            | 36432 – 37811  | +3    | 459 (1380)    | MGE – INF       |
| orf31 | Putative membrane protein                 | 37973 – 39019  | +2    | 348 (1047)    | MTT – TPG       |

## EXAMPLE 2 - Expression Cassette

### Generation of an Expression Cassette

5 This example illustrates the production of an expression cassette according to the present invention. This expression cassette, plasmid pAGvarX has been designed for the efficient generation of variant *lanA* genes of the present invention which can then be introduced into a host cell, such as a strain of *A. garbadinensis* in which the wild-type *actA* has been removed (*A. garbadinensis*  $\Delta$  *actA*). This plasmid, a derivative of the vector pSET152 (Bierman *et al.*, 1992) will integrate into the host's chromosome via the *attP* attachment site. Expression of the

10 mutated *actA* gene by the host organism together with the remaining wild-type genes of the actagardine biosynthetic gene cluster should generate actagardine variants.

Construction of the plasmid pAGvarX

Unless stated otherwise all quoted positions relate to SEQ ID NO:100. The scheme for the construction of plasmid pAGvarX is shown in Figure 3. The base adjacent to the orf lying upstream of the *actA* at position 21237 to the leucine residue within the *actA* encoding region at position 21672 was amplified by PCR using the primers O/AGvar01bF and O/AGvar02bR (primer table) and pLITAG01 as a template. The primers were designed to introduce a flanking *XbaI* site at the 5' end and a *BglII* site via a silent mutation at the 3' leucine region encoding the *actA*. This fragment was introduced into dephosphorylated pUC19 previously digested using *SmaI* to yield pAGvar1.

The region of DNA spanning from the C-terminus of the *actA* to the adjacent downstream orf (21758-21836 inclusive) was amplified by PCR using the primers O/AGvar05F and O/AGvar06R and pLITAG01 as a template. The primers were designed to introduce a flanking *AvrII* site at the 5' position and an *EcoRI* site at the 3' end. The resulting PCR product was cloned into dephosphorylated pUC19 previously digested using *SmaI* to yield pAGvar2. The plasmids pAGvar1 and pAGvar2 were then digested using *XbaI* and the PCR fragment from pAGvar1 recovered and cloned into dephosphorylated *XbaI* digested pAGvar2, the correct orientation of the incoming fragment was determined by restriction analysis. The resulting plasmid pAGvar3 was subsequently digested using *BglII* and *AvrII* and ligated to the annealed oligonucleotides O/AGvar03F and O/AGvar04R generating pAGvar4. The plasmid pAGvar4 was subsequently digested using *EcoRI* and *XbaI* and the resulting ~620bp fragment including the annealed oligonucleotides introduced into pSET152 previously digested using *EcoRI* and *XbaI* yielding the vector pAGvarX.

The region of pAGvarX constructed by annealing the respective oligonucleotides, introduce a *BsrGI* site via a silent mutation at the amino acids 6 and 7 (C and T respectively) with respect to the actagardine peptide. This site can be used in conjunction with either the upstream *BglII* site or downstream *AvrII* site to introduce DNA encoding targeted mutations to any of the amino acids encoded within the *actA* peptide.

EXAMPLE 3 - Host Cell

This example illustrates the production of a lantibiotic-producing host cell in which the *lanA* gene has been inactivated. In this example, the host cell is *A. garbadinensis* in which the *actA* gene has been deleted.

Construction of the strain *A. garbadinensis*  $\Delta$  *actA*

The strain *A. garbadinensis*  $\Delta$  *actA* is utilized as a host for expressing variants of the actagardine structural gene *actA*. This strain was generated from wild-type *A. garbadinensis* using the Redirect technology developed by Gust *et al.*, 2002. Firstly, the region of DNA from the cosmid CosAG14 encoding *actA* was replaced with the cassette SBdel-1. SBdel-1 consists of the apramycin resistance gene (*aac(3)/IV*) and *oriT* flanked by FLP recognition target (FRT) sites and was amplified by PCR using the plasmid pIJ773 as the template together with the primers O/SB50F and O/SB51R which bind at 21536 and 21802 of SEQ ID NO:100 respectively. Following the Redirect protocol (Gust *et al.*, 2004), *actA* of CosAG14 was replaced with SBdel-1 generating the cosmid CosAG14 $\Delta$ A. The central part of the SBdel-1 cassette was subsequently removed from CosAG14 $\Delta$ A by FLP-mediated excision following step 7 of the Redirect protocol generating CosAG14 $\Delta$ B. Removal of this region allows the generation of non-polar, unmarked in-frame deletions as well as repeated use of the same resistance marker (Gust *et al.*, 2003).

The second stage of construction was to engineer the cosmid so that it could be introduced into *A. garbadinensis* via conjugation. This began by first inserting CosAG14 $\Delta$ B into the *E. coli* strain BW25113/pIJ790 by transformation. The ampicillin gene of CosAG14 $\Delta$ B was then replaced with SBdel-2 following the Redirect protocol (Gust *et al.*, 2004) generating the cosmid CosAG14 $\Delta$ C. The cassette SBdel-2, like SBdel-1, houses the apramycin resistance gene (*aac(3)/IV*) and *oriT* flanked by FRT sites but was generated using the primers O/SB52F and O/SB53R together with the template pIJ773.

CosAG14 $\Delta$ C was used to transform electrocompetent cells of *E. coli* ET12567/pUZ8002 before being conjugated with *A. garbadinensis* following the Redirect protocol (Gust *et al.*, 2004; see also following paragraph). The resulting strain in which the *actA* gene has been removed from the chromosome of the wild-type producer is *A. garbadinensis*  $\Delta$  *actA*.

In more detail, to obtain the *A. garbadinensis*  $\Delta$  *actA* strain above, CosAG14 $\Delta$ C was used to transform electrocompetent cells of *E. coli* ET12567/pUZ8002 before being conjugated with *A. garbadinensis*. Apramycin resistant exconjugants were obtained and sub-cultured through six successive rounds of growth in TSB without apramycin. Cells from culture 6 were plated onto medium 65 and incubated at 30°C. After 5 days colonies were transferred and patched out over an area of approximately 1 cm<sup>2</sup> onto medium 65. After 3 days incubation at 30°C the patched cells were transferred to medium 65 containing apramycin at a final concentration of 50 µg/ml. Following 72 h incubation at 30°C, cells sensitive to apramycin were selected and the respective patches used to inoculate 50 ml flasks containing 10 ml TSB and grown at

30°C, 250 rpm for 4 days. Genomic DNA was prepared from each culture and analysed by PCR using oligonucleotides O/AGvar01bF and O/AGvar06r. PCR products of a size consistent with the deletion of the *actA* gene were generated. In parallel, analysis of fermentation broths by hplc demonstrated that these same samples did not produce actagardine.

#### EXAMPLE 4 - Heterologous Expression

This example illustrates the expression of actagardine from the SEQ ID NO:100 gene cluster in a host cell which is a non-producer cell, *S. lividans*. Such host cells provide an alternative means of generating active variants of these two peptides.

The cosmid CosAG14 and CosAL02 containing the biosynthetic gene clusters encoding the production of actagardine and deoxy-actagardine B do not possess an origin of transfer (*oriT*) necessary to facilitate conjugal transfer to a heterologous host. Using Redirect technology (Gust *et al.*, 2002) an *oriT* together with a phage attachment site *attP* and integrase (*int*) can be introduced into the SuperCos1 backbone of CosAG14 and CosAL02 replacing the neomycin resistance gene, *neo*.

#### Construction of vectors for heterologous expression.

The cosmid pMJCOS1 (supplied by the JIC, Norwich) is a derivative of SuperCos1 (Stratagene) in which the gene encoding for neomycin resistance has been replaced by a cassette (HEapra) which includes DNA encoding an *oriT*, *attP*, integrase (*int*) and apramycin resistance gene (*aac(3)/IV*). The cassette HEapra was isolated by digesting pMJCOS1 with *SspI* and recovering the DNA from an agarose gel. This cassette together with CosAG14 and CosAL02 were used to generate the cosmid CosAG14HEapra and CosAL02HEapra respectively following the Redirect protocol as described by Gust *et al.*, 2004.

The cosmid CosAG14HEapra was subsequently introduced into *S. lividans* via conjugation. Apramycin resistant exconjugants of *S. lividans*/CosAG14HEapra were isolated. Three exconjugants were used to inoculate TSB seed media. *S. lividans*, *A. garbadinensis* and *A. liguriae* were grown in parallel to provide controls. Following 48 h incubation the seed cultures were used to inoculate a range of four different production media namely, AAS1, GM1, GM3 and TSB. These cultures were incubated for a total of nine days at 30°C with 1.5 ml aliquots being removed from each flask after 5, 7 and 9 days incubation. The aliquots were centrifuged at 14000 rpm (IEC micromax benchtop centrifuge) for 10 minutes and the supernatants then decanted and used undiluted for bioassays and HPLC-MS analysis.

Zones of inhibition (haloes) indicative of the presence of a biological active compound(s) were observed around all of the wells loaded with supernatants of *S. lividans* containing the cosmid CosAG14HEapra (*S. lividans*/CosAG14HEapra) except for wells loaded with supernatant from fermentations in TSB where no haloes were generated. No biological activity was observed around wells loaded with supernatant from fermentations of *S. lividans* grown in any of the four media. Haloes were evident around all wells loaded with supernatants from cultures of *A. liguriae* and *A. garbadinensis* where growth was supported. All haloes were consistently generated from the first day of sampling on day 5 through to day 9 although a general reduction in the diameter of the haloes was evident.

HPLC-MS analysis of the supernatants from the fermentations of *S. lividans*/CosAG14HEapra confirm the presence of peaks with retention times and masses corresponding to ala(O)actagardine. These same peaks were absent from supernatants of *S. lividans* only.

Table 3 summarises the HPLC-MS analyses of supernatants from fermentation of *S. lividans*,

*S. lividans*/CosAG14HEapra, *A. garbadinensis* and *A. liguriae* following incubation for 5 days.

Table 3:

| Sample                           | Fermentation medium | Concentration of product (µg/ml) | Retention Time (min) | Molecular ion (m/z) | Identity                                 |
|----------------------------------|---------------------|----------------------------------|----------------------|---------------------|------------------------------------------|
| <i>S.lividans</i> /CosAG14HEapra | GM1                 | 83                               | 6.75                 | 981                 | Ala(0)Actagardine (M+2H) <sup>+2</sup>   |
|                                  |                     |                                  |                      | 991                 | Ala(0)Actagardine (M+H+Na) <sup>+2</sup> |
| <i>S.lividans</i> /CosAG14HEapra | GM3                 | 33                               | 6.75                 | 981                 | Ala(0)Actagardine (M+2H) <sup>+2</sup>   |
|                                  |                     |                                  |                      | 991                 | Ala(0)Actagardine (M+H+Na) <sup>+2</sup> |
| <i>S.lividans</i>                | GM1                 | Not Detected                     | Not Detected         | Not Detected        | Not Detected                             |
| <i>S.lividans</i>                | GM3                 | Not Detected                     | Not Detected         | Not Detected        | Not Detected                             |
| <i>A. garbadinensis</i>          | GM1                 | 58                               | 6.9                  | 945                 | Actagardine (M+2H) <sup>+2</sup>         |
| <i>A. garbadinensis</i>          | GM3                 | 24                               | 6.8                  | 981                 | Ala(0)Actagardine (M+2H) <sup>+2</sup>   |
|                                  |                     |                                  |                      | 991                 | Ala(0)Actagardine (M+H+Na) <sup>+2</sup> |
| <i>A. liguriae</i>               | GM1                 | Not detected                     | 7.06                 | 937                 | Deoxy-actagardineB (M+2H) <sup>+2</sup>  |
| <i>A. liguriae</i>               | GM3                 | Not detected                     | 7.06                 | 937                 | Deoxy-actagardineB (M+2H) <sup>+2</sup>  |

EXAMPLE 5 - Antibacterial Activities**MIC Determination**

A selection of the variants produced as disclosed herein above were tested further for activity against a range of bacteria. Minimum inhibitory concentrations (MICs) for all organisms with the exception of *Streptococcus pneumoniae* were determined by two-fold serial antibiotic dilutions in Mueller-Hinton broth (MHB) supplemented with calcium chloride dehydrate to a final calcium concentration of 400µg/ml. Minimum inhibitory concentrations (MICs) for *S. pneumoniae* were determined by two-fold serial antibiotic dilutions in Brain Heart Infusion (BHI) broth supplemented with 400 µg/ml calcium chloride dihydrate. Antimicrobial agent stock solutions were prepared and stored according to NCCLS standard M7-A6.

Actively growing broth cultures were diluted to contain 10<sup>5</sup> to 10<sup>6</sup> CFU/ml by adjusting to an absorbance of 0.2 – 0.3 at 600nm, equivalent to the McFarland 0.5 standard. They were then diluted a further 1:100 in broth. The assays were performed in duplicate in sterile 96-well microtitre plates in a total volume of 200 µl (160 µl broth, 20 µl antibiotic, 20 µl inoculum) in a concentration range from 64 µg/ml to 0.06 µg/ml. The 12th well of the microtitre plate contained no antimicrobial agent. Vancomycin was used as a reference antibiotic for quality control. Plates were incubated aerobically, shaking, for 18 – 20 hours at 37°C with the MIC defined as the lowest concentration of drug that produced no visible growth.

|                   | <i>E.faecium</i><br>19579 | <i>E.faecalis</i><br>29212 | <i>S.aureus</i><br>R33 | <i>S.aureus</i><br>SH1000 | <i>S.epidermidis</i><br>11047 | <i>S.pneumoniae</i><br>R6 |
|-------------------|---------------------------|----------------------------|------------------------|---------------------------|-------------------------------|---------------------------|
| Actagardine       | 4, 4                      | <4, <4                     | 16, 8                  | 8, 8                      | 8, 8                          | <4, <4                    |
| Actagardine       | 4, 4                      | <4, <4                     | 16, 16                 | 8, 8                      | 8, 8                          | <4, <4                    |
| Ala(0)Actagardine | 8, 8                      | 4, 4                       | 8, 8                   | 8, 8                      | 8, 4                          | <4, <4                    |
| Ala(0)Actagardine | 32, 16                    | 8, 8                       | <4, <4                 | 8, 8                      | 8, 8                          | <4, <4                    |
| Deoxyactardine B  | 16, 16                    | 4, 4                       | 16, 16                 | 16, 16                    | 16, 16                        | 8, 8                      |
| Deoxyactardine B  | 16, 16                    | <4, <4                     | 16, 16                 | 16, 16                    | 16, 16                        | <4, <4                    |

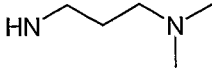
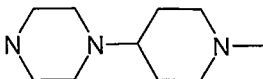
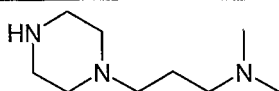
EXAMPLE 6 - NMR Analysis**NMR Studies on Actagardine and Deoxy-actagardine**

NMR spectroscopy (COSY, TOCSY, HSQC and NOESY) was successfully used to confirm the sequencing results obtained from producers of actagardine (*A. garbadinensis*) and deoxy-actagardine B (*A. liguriae*). Whilst the data obtained did not permit a completely unambiguous assignment of all residues, it was consistent with the structures shown in Figure 4 and sufficient to confirm that deoxy-actagardine B from *A. liguriae* has at positions 15 and 16 the residues Leu and Val respectively.

**EXAMPLE 7 – SYNTHESIS OF DERIVATIVES**

The following derivatives of deoxy-actagardine B were made, in which the groups Z and the C-terminal amide were as follows:

**Compound Structures**

| Compound | Z      | C-terminal amide                                                                  |
|----------|--------|-----------------------------------------------------------------------------------|
| I        | H      |  |
| II       | H      |  |
| III      | H      |  |
| IV       | D-Ala  | H                                                                                 |
| V        | L-Ile  | H                                                                                 |
| VI       | L-Val  | H                                                                                 |
| VII      | L-Phe  | H                                                                                 |
| VIII     | L-Lys  | H                                                                                 |
| IX       | L-Tryp | H                                                                                 |

5

The synthesis of the compounds I – XI was as follows:

**General procedure 1. Preparation of compounds I – III**

To a solution of deoxy-actagardine B (20mg, 11nmol), the appropriate amine (11nmol) and diisopropylethylamine (7.2μl, 70nmol) in dry dimethylformamide (0.8ml) were added 200μl of a solution of benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBop) (70mg, 134nmol) in dry dimethylformamide (1.0ml). The mixture was analysed by HPLC to follow the progress of the reaction, adding further aliquots of the Pybop solution until all the starting material had been consumed. HPLC analysis at this stage also showed variable amounts (5-20%) of the diamide. After completion of the reaction, the mixture was diluted with 30% acetonitrile in 20mM Kpi aqueous phosphate buffer, pH7 (10ml) and the monoamide was purified by preparative HPLC using the conditions described in Table 4. The appropriate fractions were concentrated to 25% of their original volume and desalted by loading on to a preconditioned C18 Bond Elut column (500mg) which was subsequently washed by sequential elution with two column volumes of 30, 40, 70 and 90% aqueous methanol. Evaporation of the appropriate fractions gave the desired products as white solids.

20

Compound I : Deoxy-Actagardine B N-[3-dimethylaminopropyl] monocarboxamide

Was obtained from coupling of deoxyactagardine B and 3-(dimethylamino)propylamine according to General Procedure 1. Yield 18mg, 85% yield.  $[M+2H\ 2^+]$  calculated 979.0, found 980.2

5      Compound II : Deoxy-Actagardine B N-[1-(1-methyl-4-piperidiny)l)piperazine] monocarboxamide

Was obtained from the coupling of deoxyactagardine B and 4-(piperidino)piperazine according to General Procedure 1. Yield 8mg, 37% yield.  $[M+2H\ 2^+]$  calculated 1019.5, found 1020.0;  $[M+3H\ 3^+]$  calculated 680.0, found 680.0

10      Compound III: Deoxy-Actagardine B [1-(3-dimethylaminopropyl)piperazine] monocarboxamide

Was obtained from the coupling of Deoxy-actagardine B and 1-(3-dimethylaminopropyl)piperazine according to general procedure 1. Yield 10mg, 46%  $[M+2H\ 2^+]$  calculated 1013.5, found 1014.0

15      **General procedure 2. Preparation of compounds IV-IX**

A solution of the appropriate Fmoc protected amino acid (34nmol) in dry dimethylformamide (0.4ml) was treated with a solution of benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBop) (11.4mg, 22nmol) and diisopropylethylamine (11 $\mu$ l, 68 nmol) in dry dimethylformamide (0.4ml). The mixture was then added to a solution of Deoxy-  
 20      Actagardine B (2mg, 11nmol) in dry dimethylformamide (0.5ml) The mixture was left at room temperature for 1h, after which time analytical HPLC (30-65% acetonitrile in 20mM Kpi aqueous phosphate buffer, pH7) showed complete conversion of the starting material. The reaction mixture was diluted with 40% aqueous methanol (20ml) and the mixture was passed through a C18 Bond Elute column (500mg) that had been preconditioned by washing with two  
 25      column volumes of 100% methanol followed by two column volumes of water. The column was eluted sequentially with two column volumes of 40, 50, 60, 70, 80, 90 and 100% aqueous methanol. The fractions were analysed by HPLC and the fractions containing the Fmoc-protected coupling product were evaporated to dryness. The residue was taken up in dimethylformamide (1ml) and piperidine (50 $\mu$ l) was added to remove the Fmoc protecting  
 30      group. Progress of the reaction was monitored by HPLC and after complete consumption of the starting material the solution was diluted into 30% aqueous methanol (20ml). The mixture was then eluted through a C18 Bond Elut cartridge (500mg) as previously described and the product obtained after evaporation of the appropriate fractions was further purified by preparative HPLC using the conditions described in Table 4. The appropriate fractions were  
 35      concentrated to 25% of their original volume and desalted by loading on to a preconditioned

C18 Bond Elut column (500mg) which was subsequently washed by sequential elution with two column volumes of 30, 40, 70 and 90% aqueous methanol. Evaporation of the appropriate fractions gave the desired products as white solids.

Compound IV : D-Ala (0)deoxy-actagardine B

- 5 Was prepared according to general procedure 2 from Deoxy-actagardine B and Fmoc-D-alanine in 74% yield.  $[M+2H\ 2^+]$  calculated 972.5, found 973.0  
043/188

Compound V : L-Ile(0)deoxy-actagardine B

- 10 Was prepared according to general procedure 2 from Deoxy-actagardine B and Fmoc-L-isoleucine in 27% yield.  $[M+2H\ 2^+]$  calculated 993.5, found 993.8

Compound VI : L-Val(0)deoxyactagardine B

Was prepared according to general procedure 2 from Deoxy-actagardine B and Fmoc-L-valine in 55% yield.  $[M+2H\ 2^+]$  calculated 986.5, found 985.9.

Compound VII : L-Phe(0)deoxyactagardine B

- 15 Was prepared according to general procedure 2 from Deoxy-actagardine B and Fmoc-L-phenylalanine in 22% yield.  $[M+2H\ 2^+]$  calculated 1010.5, found 1010.9.

Compound VIII : L-Lys(0)deoxyactagardine B

Was prepared according to general procedure 2 from Deoxy-actagardine B and Bis(Fmoc)-L-lysine in 45% yield.  $[M+2H\ 2^+]$  calculated 1001.0, found 1001.6

- 20 Compound IX : L-Tryp(0)deoxyactagardine B

Was prepared according to general procedure 2 from Deoxy-actagardine B and Fmoc-L-tryptophan in 55% yield.  $[M+2H\ 2^+]$  calculated 1030.0, found 1029.9.

EXAMPLE 8 – FURTHER ANTIBACTERIAL DATA

25 **MIC Determination**

Staphylococcus, Streptococcus, Enterococcus spp.

- Minimum inhibitory concentrations (MICs) were determined and antimicrobial agent stock solutions were prepared and stored according to the NCCLS reference microdilution broth method for aerobic bacteria (M7-A6, 2003). MICs were determined by two-fold serial antibiotic  
30 dilutions in Mueller-Hinton broth (MHB) or Brain Heart Infusion (BHI) broth (*S. pneumoniae*). Actively growing broth cultures were adjusted in sterile broth or by direct colony suspension (*S. pneumoniae*) to a turbidity equivalent to the McFarland 0.5 standard ( $1 \times 10^8$  CFU/ml), then further diluted in sterile broth for a final inoculum in sterile 96-well microtitre plates of approximately  $5 \times 10^5$  CFU/ml. The assays were performed in duplicate with *Enterococcus*

*faecalis* ATCC 29212 included as a reference control strain and Vancomycin as a reference antibiotic for quality control. Plates were incubated aerobically, shaking, for 18 – 20 hours at 37°C with the MIC defined as the lowest concentration of drug that produced no visible growth.

5           *Clostridium difficile*

Minimum inhibitory concentrations (MICs) for *C. difficile* were determined and antimicrobial agent stock solutions were prepared and stored according to the NCCLS reference agar dilution method for anaerobic bacteria (M11-A5, 2001). Two-fold serial antibiotic dilutions were prepared in Wilkens-Chalgren agar (WCA). Test organisms were selected from 48 hour  
10 growth on Braziers (C.C.E.Y.) agar, subcultured in Schaedler broth to a density equivalent to a McFarland 0.5 standard ( $1 \times 10^8$  CFU/ml), with a final inoculum onto WCA plates of approximately  $10^5$  CFU/spot. *Bacteroides fragilis* ATCC 25285 was included as a reference control strain and Metronidazole was used as a reference antibiotic for quality control. All manipulations were performed in duplicate in ambient atmosphere in pre-reduced media with  
15 only brief exposure to oxygen. Plates were incubated anaerobically for 48 hours at 37°C with the MIC defined as the concentration of drug where a marked reduction occurred in the appearance of growth on the test plate compared to growth on the anaerobic control plate.

*Propionibacterium acnes*

Test organisms were selected from 3-7 day growth on Wilkens-Chalgren agar (WCA)  
20 supplemented with furazolidone (1-2 µg/ml). Fresh Wilkens-Chalgren broth (WCB) was inoculated by direct colony suspension with single colonies of *P. acnes* and adjusted to a density equivalent to the McFarland 0.5 standard ( $1 \times 10^8$  CFU/ml), then further diluted in sterile WCB for a final inoculum in sterile 96-well microtitre plates of approximately  $10^5$  CFU/ml. Two-fold serial antibiotic dilutions were performed in sterile water with stock  
25 solutions prepared and stored according to NCCLS standards (M11-A5, 2001). The assays were performed in duplicate with Vancomycin and Clindamycin used as reference antibiotics for quality control. Plates were incubated anaerobically for 48-72 hours at 37°C with the MIC defined as the concentration of drug where a marked reduction occurred in the appearance of growth on the test plate compared to growth on the anaerobic control plate. All manipulations  
30 were performed in duplicate in ambient atmosphere in pre-reduced media with only brief exposure to oxygen.

Culture media were supplemented with calcium ions (as calcium chloride) at 50µg/ml except where higher concentrations are indicated. MIC values in µg/ml are shown in the following  
35 Tables:

Table 6: MIC values against Enterococci, Streptococci and Staphylococci

| Organism                                                    | Ala(O)-                |                        |
|-------------------------------------------------------------|------------------------|------------------------|
|                                                             | deoxyactagardine-<br>B | Deoxyactagardine-<br>B |
| <i>M. luteus</i> 4698 +200 µg/ml Ca <sup>2+</sup>           | 4                      | 8                      |
| <i>E. faecalis</i> 29212                                    | 16                     | 16                     |
| <i>E. faecalis</i> 29212 +200 µg/ml Ca <sup>2+</sup>        | 4                      | 8                      |
| <i>E. faecalis</i> 29212 +400 µg/ml Ca <sup>2+</sup>        |                        | 4                      |
| <i>E. faecium</i> 7131121 (VRE)                             | >64                    | >64                    |
| <i>E. faecium</i> 7131121 (VRE) +200 µg/ml Ca <sup>2+</sup> | >64                    | >64                    |
| <i>E. faecium</i> 7131121 (VRE) +400 µg/ml Ca <sup>2+</sup> |                        | 32                     |
| <i>E. faecium</i> 19579                                     | >64                    | >64                    |
| <i>E. faecium</i> 19579 +200 µg/ml Ca <sup>2+</sup>         | >64                    | >64                    |
| <i>E. faecium</i> 19579 +400 µg/ml Ca <sup>2+</sup>         |                        | 16                     |
| <i>S. aureus</i> R33 (MRSA)                                 | 32                     | 32                     |
| <i>S. aureus</i> R33 (MRSA) +200 µg/ml Ca <sup>2+</sup>     | 16                     | 8                      |
| <i>S. aureus</i> R33 (MRSA) +400 µg/ml Ca <sup>2+</sup>     |                        | 16                     |
| <i>S. aureus</i> SH1000                                     | 16                     | 16                     |
| <i>S. aureus</i> SH1000 +200 µg/ml Ca <sup>2+</sup>         | 8                      | 8                      |
| <i>S. aureus</i> SH1000 +400 µg/ml Ca <sup>2+</sup>         |                        | 16                     |
| <i>S. epidermidis</i> 11047                                 | 16                     | 32                     |
| <i>S. epidermidis</i> 11047 +200 µg/ml Ca <sup>2+</sup>     | 8                      | 16                     |
| <i>S. epidermidis</i> 11047 +400 µg/ml Ca <sup>2+</sup>     |                        | 16                     |
| <i>S. pneumoniae</i> R6                                     | 16                     | 16                     |
| <i>S. pneumoniae</i> R6 +200 µg/ml Ca <sup>2+</sup>         | 32                     | 6                      |
| <i>S. pneumoniae</i> R6 +400 µg/ml Ca <sup>2+</sup>         |                        | 4                      |
| <i>S. aureus</i> 12232 MRSA                                 |                        | 16                     |
| <i>S. aureus</i> R36 (MRSA)                                 |                        | 16                     |
| <i>S. aureus</i> R34 (MRSA)                                 |                        | 16                     |
| <i>S. aureus</i> R39 (MRSA)                                 |                        | >32                    |
| <i>S. aureus</i> R40 (MRSA)                                 |                        | >32                    |
| <i>S. aureus</i> W71 (MRSA)                                 |                        | >32                    |
| <i>S. aureus</i> W74 (MRSA)                                 |                        | >32                    |
| <i>S. aureus</i> W96 (MRSA)                                 |                        | >32                    |
| <i>S. aureus</i> W97 (MRSA)                                 |                        | >32                    |
| <i>S. aureus</i> W98 (MRSA)                                 |                        | >32                    |
| <i>S. aureus</i> W99 (MRSA)                                 |                        | >32                    |
| <i>S. epidermidis</i> 7755298 (MRSE)                        |                        | >32                    |
| <i>S. epidermidis</i> 7865688 (MRSE)                        |                        | >32                    |
| <i>S. epidermidis</i> 7753921 (MRSE)                        |                        | >32                    |

| Organism                              | Ala(O)-<br>deoxyactagardine- | Deoxyactagardine- |
|---------------------------------------|------------------------------|-------------------|
|                                       | B                            | B                 |
| <i>S. epidermidis</i> GRL05011 (MRSE) |                              | >32               |

Table 7: MIC values against fusidic acid-resistant *Staphylococcus aureus*

| Organism                                                             | Deoxy-Actagardine B |
|----------------------------------------------------------------------|---------------------|
| Fusidic acid-res <i>S. aureus</i> 8325-4                             | 8, 8                |
| Fusidic acid-res <i>S. aureus</i> CS1116                             | 32, 32              |
| Fusidic acid-res <i>S. aureus</i> CS957                              | 32, 32              |
| Fusidic acid-res <i>S. aureus</i> CS767                              | 32, 32              |
| Fusidic acid-res <i>S. aureus</i> CS858                              | 32, 32              |
| Fusidic acid-res <i>S. aureus</i> CS741                              | 32, 32              |
| Fusidic acid-res <i>S. aureus</i> CS1145                             | 16, 16              |
| Fusidic acid-res <i>S. aureus</i> CS872                              | 16, 16              |
| Fusidic acid-res <i>S. aureus</i> CS866                              | 32, 32              |
| Fusidic acid-res <i>S. aureus</i> CS607                              | 64, 64              |
| Fusidic acid-res <i>S. aureus</i> CS22                               | 16, 16              |
| Fusidic acid-res <i>S. aureus</i> 8325-4 +200 µg/ml Ca <sup>2+</sup> | 4, 4                |
| Fusidic acid-res <i>S. aureus</i> CS1116 +200 µg/ml Ca <sup>2+</sup> | 16, 16              |
| Fusidic acid-res <i>S. aureus</i> CS957+200 µg/ml Ca <sup>2+</sup>   | 16, 16              |
| Fusidic acid-res <i>S. aureus</i> CS767 +200 µg/ml Ca <sup>2+</sup>  | 16, 16              |
| Fusidic acid-res <i>S. aureus</i> CS858 +200 µg/ml Ca <sup>2+</sup>  | 16, 16              |
| Fusidic acid-res <i>S. aureus</i> CS741+200 µg/ml Ca <sup>2+</sup>   | 16, 16              |
| Fusidic acid-res <i>S. aureus</i> CS1145 +200 µg/ml Ca <sup>2+</sup> | 8, 8                |
| Fusidic acid-res <i>S. aureus</i> CS872 +200 µg/ml Ca <sup>2+</sup>  | 8, 8                |
| Fusidic acid-res <i>S. aureus</i> CS866 200 µg/ml Ca <sup>2+</sup>   | 8, 8                |
| Fusidic acid-res <i>S. aureus</i> CS607 +200 µg/ml Ca <sup>2+</sup>  | 32, 32              |
| Fusidic acid-res <i>S. aureus</i> CS22 +200 µg/ml Ca <sup>2+</sup>   | 4, 4                |

Table 8: MIC values against mupirocin-resistant *Staphylococcus . aureus*

| Organism | Deoxy-actagardine B |
|----------|---------------------|
| 8325-4   | 8, 8                |
| GISA-2   | 8, 8                |
| LZ6      | 16, 16              |
| LZ8      | 16, 16              |
| LZ9      | 16, 16              |
| LZ10     | 8, 8                |
| 420      | 4, 4                |

|      |        |
|------|--------|
| 1205 | 16, 16 |
| 1120 | 16, 16 |
| 1454 | 16, 16 |
| 1086 | 8, 8   |

Table 9: MIC values against *Propionibacterium acnes*

| Organism                                          | Deoxy-Actagardine B |
|---------------------------------------------------|---------------------|
| <i>Propionibacterium acnes</i> P37 (lab strain)   | 4, 4                |
| <i>P. acnes</i> AT1                               | 4, 4                |
| <i>P. acnes</i> AT26                              | 2, 2                |
| <i>P. acnes</i> 101897d                           | 2, 2                |
| <i>P. acnes</i> PF284 (tet res)                   | 2, 2                |
| <i>P. acnes</i> PF286 (erythro & clin res)        | 2, 2                |
| <i>P. acnes</i> PF289 (clin and co-trimazole res) | 4, 8                |

Table 10: MIC values against *C.difficile*

| Organism                   | Ala(O)-<br>deoxyactagardine- |   |
|----------------------------|------------------------------|---|
|                            | B                            | B |
| <i>C. difficile</i> 37779  | 4                            | 4 |
| <i>C. difficile</i> 19126  | 2                            | 4 |
| MIC <sub>50</sub> 10       |                              |   |
| <i>C.difficile</i> strains |                              | 2 |
| MIC <sub>90</sub> 10       |                              |   |
| <i>C.difficile</i> strains |                              | 4 |

5

Table 11: MIC values against *C.difficile*

| Organism                             | I | II | III | IV | V | VI | VII | VIII | IX |
|--------------------------------------|---|----|-----|----|---|----|-----|------|----|
| <i>C. difficile</i> 37779            |   |    | >8  | 4  | 4 | 8  | 1   | 4    | 1  |
| <i>C. difficile</i> 19126            |   |    | >8  | 4  | 4 | 8  | 2   | 4    | 2  |
| MIC <sub>50</sub> <i>C.difficile</i> | 2 | 2  |     |    |   |    | 2   |      |    |
| MIC <sub>90</sub> <i>C.difficile</i> | 4 | 4  |     |    |   |    | 2   |      |    |

Materials & Methods

The materials and methods used in Examples 2-7 above are as follows:

Media

All buffers, solutions and media were made up using reverse osmosis (RO) water and

5 contained per litre the following ingredients:

*AAS1*

|                       |      |
|-----------------------|------|
| Soluble starch        | 10 g |
| Glucose               | 10 g |
| Peptone               | 5 g  |
| Dry corn steep liquor | 1 g  |
| Yeast extract         | 2 g  |
| Adjust pH to 6.0      |      |

*GM1*

|                       |       |
|-----------------------|-------|
| Lablemco meat extract | 4 g   |
| Peptone               | 4 g   |
| NaCl                  | 2.5 g |
| Yeast extract         | 1 g   |
| Soy flour             | 10 g  |
| Glucose               | 25 g  |
| CaCO <sub>3</sub>     | 5 g   |
| Adjust pH to 7.6      |       |

*Mueller Hinton*

|                      |      |
|----------------------|------|
| Mueller Hinton broth | 21 g |
| For agar plates add; |      |
| Agar                 | 10 g |

*ABB13*

|                   |       |
|-------------------|-------|
| Soytone peptone   | 5 g   |
| Soluble starch    | 5 g   |
| CaCO <sub>3</sub> | 3 g   |
| MOPS              | 2.1 g |
| Agar              | 20 g  |
| Adjust pH to 7.0  |       |

*BHI*

|                      |      |
|----------------------|------|
| Brain Heart Infusion | 37 g |
|----------------------|------|

*GM3*

|                  |      |
|------------------|------|
| Arkasoy soyflour | 20 g |
| Mannitol         | 20 g |
| Adjust pH to 7.0 |      |

*LA*

|            |      |
|------------|------|
| Luria agar | 40 g |
|------------|------|

*LB*

|             |      |
|-------------|------|
| Luria broth | 25 g |
|-------------|------|

*SFM*

|            |      |
|------------|------|
| Soya Flour | 20 g |
| D-mannitol | 20 g |
| Agar       | 16 g |

*SV2*

|                   |      |
|-------------------|------|
| Glucose           | 15 g |
| Glycerol          | 15 g |
| Peptone           | 15 g |
| NaCl              | 3 g  |
| CaCO <sub>3</sub> | 1 g  |
| Adjust pH to 7.0  |      |

*TAE buffer*

|                  |         |
|------------------|---------|
| Tris             | 48.44 g |
| EDTA             | 3.72 g  |
| Adjust pH to 8.3 |         |

*TSB*

|                   |      |
|-------------------|------|
| Tryptic soy broth | 30 g |
| Adjust pH to 7.0  |      |

*2TY*

|                        |      |
|------------------------|------|
| Tryptone               | 16 g |
| Yeast extract          | 10 g |
| NaCl                   | 5 g  |
| Adjust pH to 6.5 – 7.0 |      |

*'65'*

|                   |      |
|-------------------|------|
| Glucose           | 4 g  |
| Yeast extract     | 4 g  |
| Malt extract      | 10 g |
| CaCO <sub>3</sub> | 2 g  |
| Agar              | 12 g |
| Adjust pH to 7.2  |      |

*Bioassays*

*Micrococcus luteus* was inoculated from frozen stock into 10ml Mueller-Hinton broth and grown overnight at 30°C with shaking at 200 rpm. 1 ml of this culture was used to inoculate 300 ml of Mueller-Hinton agar which was then poured into petri dishes. Wells (6 mm diameter) placed equidistant apart were made using a cork-borer and subsequently loaded with 50 µl of the respective sample. The bioassay plate was placed into a laminar air flow until the loaded samples had diffused, at which point the plates were transferred to a 30°C incubator and left overnight.

*Endonuclease restriction digestions*

Digestions of DNA with restriction enzymes were carried out in the supplied buffers and in accordance with the manufacturer's guidelines. Typically, for preparative digests 5 µg of DNA was digested with 12 units of enzyme for 3 h at the recommended temperature. For analytical digests, 0.5 µg of DNA was digested with 2 units of enzyme for 2-3 h again at the recommended temperature. The digested DNA was analysed by agarose gel electrophoresis.

Sub-culturing exconjugants

Agar plugs of patched exconjugants were used to inoculate 50 ml flasks containing 8 ml TSB and 2 glass beads. The cultures were incubated at 30°C, 250 rpm for 10 days then 100 µl were removed and added to 10 ml TSB in a 50 ml flask containing 2 glass beads. The flasks were incubated for 2 days then 1 ml was removed and used to inoculate a 50 ml flask containing 10 ml TSB. Using 1 ml inoculum a total of six successive rounds of growth were carried out each incubated for 2 days at 30°C, 250 rpm. Cells from the sixth round of sub-culturing were pelleted by centrifuging at 4000 rpm for 20 minutes (Heraeus Sepatech Megafuge) then sonicated (MSE Sanyo Soniprep 150, amplitude 10-15 microns) for 30 seconds in TSB to disrupt the mycelium. Serial dilutions ( $10^{-1}$  to  $10^{-5}$  in TSB) of the sonicated cells were plated onto medium 65 and incubated at 30°C.

Fermentation for heterologous expression

50 ml conical flasks each containing 2 glass beads and either 8 ml TSB or AAS1 media supplemented with nalidixic acid and the appropriate selective antibiotic were inoculated using agar plugs or 250 µl of a -80°C glycerol stock. Following 2 - 4 days incubation at 30°C, 200 rpm, 1.2 ml (3%) per seed culture was used to inoculate 40 ml of the respective production media in 250 ml conical flasks containing 2 glass beads. These cultures were incubated at 30°C, 200 rpm for 9 days. 1.5 ml whole broth aliquots were removed periodically from each culture for analysis by bioassay and/or HPLC-MS analysis.

Fermentation of *A. liguriae* for the isolation of deoxy-actagardine B

250 ml conical flasks each containing 2 glass beads and 50 ml SV2 media were inoculated with 500 µl (1%) of *A. liguriae* cells from a glycerol stock. Following 4 days incubation at 30°C, 250 rpm, 12 ml (3%) per seed culture was used to inoculate 400 ml of GM3 in 2L conical flasks. These cultures were incubated at 30°C, 225 rpm for nine days. The culture broth was harvested by centrifugation at 4000 rpm (Heraeus Sepatech Megafuge) for 30 minutes after which the supernatant was decanted from the pellet of cells.

Fermentation of *A. garbadinensis* for the isolation of actagardine and Ala(O)-actagardine

250 ml conical flasks each containing 2 glass beads and 50 ml AAS media were inoculated with 500 µl (1%) of *A. garbadinensis* cells from a glycerol stock. Following 9 days incubation at 30°C, 250 rpm, 12 ml (3%) per seed culture was used to inoculate 400 ml of AAS in 2L conical

flasks. These cultures were incubated at 30°C, 200 rpm for eight days. The culture broth was harvested by centrifugation at 4000 rpm (Heraeus Sepatech Megafuge) for 30 minutes after which the supernatant was decanted from the pellet of cells.

5           Isolation of Deoxy-actagardine B for MIC studies

Diaion HP-20 resin (50 g/L) was added and mixed with supernatant isolated from a fermentation of *A. liguriae* and left overnight at 4°C. The suspension was aliquoted into Bond Elut columns (60 ml) and the resin washed sequentially with four bed volumes of water followed by three bed volumes of 25, 50, 75 and 100% methanol. HPLC analysis confirmed the presence of Deoxy-actagardine B in the 50, 75 and 100% methanol fractions. These fractions were combined then concentrated to approximately a quarter of the volume of the starting pool. The concentrate from 1L of broth was loaded onto two C18 Bond Elut columns (5 g) that had been pre-conditioned by washing with two column volumes of 100% methanol followed by two column volumes of water. The columns were eluted sequentially with two column volumes of 50, 60, 70, 80, 90% methanol followed by two column volumes of 100% methanol. HPLC analysis confirmed the presence of Deoxy-actagardine B in the 80, 90 and 100% methanol fractions, these fractions were pooled and concentrated to a third of the starting volume. An equal volume of 40 mM potassium phosphate pH 2.5 in 50% methanol was added and the concentrate then loaded evenly onto three pre-equilibrated SCX Bond Elut columns (1g). The SCX columns were initially washed with 40 mM potassium phosphate pH 2.5 in 50% methanol and then eluted using 1.5 column volumes of 250 mM potassium phosphate pH 7.0 in 50% methanol. The eluent was desalted by loading onto a C18 Bond Elut column (5 g) that had been pre-conditioned with two column volumes of methanol followed by two column volumes of water. The column was washed with two column volumes of 50% and then 60% methanol. Deoxy-actagardine B was eluted following the addition of two column volumes each of 70, 80, 90 and 100% methanol. Fractions containing purified Deoxy-actagardine B as confirmed by HPLC and LC-MS analyses were pooled and evaporated to dryness.

30           Isolation of Ala(0)-Deoxyactagardine B from fermentation of A. liguriae

Diaion HP-20 resin (50 g/L) was mixed with supernatant from a four litre fermentation of *A. liguriae* and left overnight at 4°C. The suspension was collected into a glass sinter funnel and the resin was washed sequentially with four bed volumes of water followed by four bed volumes of 50% Methanol. Deoxy-actagardineB and Ala(0)-deoxyactagardine B were eluted from the resin by washing with five bed volumes of 100% Methanol. The 100% Methanol fraction was concentrated to a third of the original volume and was then diluted by addition of

water to a final concentration of 60% Methanol. The resulting solution was loaded onto four 10g C18 Bond Elut columns prior to washing with two column volumes of 50% Methanol. Deoxy-actagardine B-related components were eluted from the column using two column volumes of Methanol/0.5% Formic Acid. The resulting eluent was concentrated by evaporation to 40 ml and Ala(0)-deoxy-actagardine B was separated from Deoxy-actagardine B by preparative HPLC using the conditions described in the table below.

|                  |                                                        |
|------------------|--------------------------------------------------------|
| Column           | Capitol HPLC Ltd C18 – BDS – HL5 – 26052 15cm x 20mm   |
| Solvent A        | 30% ACN in 20mM Potassium Phosphate pH 5.0             |
| Solvent B        | 65% ACN in 20mM Potassium Phosphate pH 5.0             |
| Detection        | 210 nm                                                 |
| Flow Rate        | 10 ml/min                                              |
| Time (T) = 0 min | 100% A                                                 |
| T = 1 min        | 100% A                                                 |
| T = 29 min       | 35% B                                                  |
| T = 30 min       | 100% B                                                 |
| T = 33 min       | 100% B                                                 |
| T = 34 min       | 100% A                                                 |
| T = 35 min       | 100% A                                                 |
| Collection       | Start 10 min; End 30 min; 0.5 or 0.25 minute fractions |

- 10 Fractions containing Ala(0)-deoxy-actagardine B (as confirmed by HPLC and LC-MS analyses) were desalted using C18 Bond Elut columns as described above before being evaporated to dryness.

Ala(0)-deoxy-actagardine B was eluted from the column at Retention Time = 5.04 minutes.

- 15 MS analysis confirmed a species of 972.2 m/z (M+2H)<sup>2+</sup>.

Isolation of actagardine and Ala(O)-actagardine for MIC studies

- 20 Actagardine and Ala(0)-actagardine were purified using the method described for the purification of Deoxy-actagardine B from *A. liguriae* with the exception that preparative HPLC was required to resolve Ala(0)actagardine and Actagardine following the SCX Bond Elut step. Eluent from the SCX Bond Elut column was concentrated by rotary evaporation from 70 to 18 ml and the resulting concentrate was purified by preparative HPLC using the conditions described in Table 4 The respective fractions containing Actagardine and Ala(0)actagardine

(as confirmed by HPLC and LC-MS analyses) were desalted using C18 Bond Elut columns as described previously before being evaporated to dryness.

Table 4: Preparative HPLC conditions for the separation of Actagardine and Ala(0)actagardine.

|                  |                                                        |
|------------------|--------------------------------------------------------|
| Column           | Capitol HPLC Ltd C18 – BDS – HL5 – 26052 15 cm x 20 mm |
| Solvent A        | 30% Acetonitrile in 20mM Potassium Phosphate pH 7.0    |
| Solvent B        | 65% Acetonitrile in 20mM Potassium Phosphate pH 7.0    |
| Detection        | 268 nm                                                 |
| Flow Rate        | 10 ml/min                                              |
| Time (T) = 0 min | 100% A                                                 |
| T = 1 min        | 100% A                                                 |
| T = 19 min       | 25% B                                                  |
| T = 20 min       | 100% B                                                 |
| T = 25 min       | 100% B                                                 |
| T = 26 min       | 100% A                                                 |
| T = 30 min       | 100% A                                                 |
| Collection       | Start 8 min; End 20 min; 1 minute fractions            |

5

#### Agarose gel electrophoresis

Electrophoresis of DNA was carried out as described by Sambrook *et al.*, 1989. Agarose gels (0.7-1%) were prepared in TAE buffer containing a final concentration of 0.1 µg/ml ethidium bromide to allow visualisation of the DNA by UV light. 0.1 volumes of 10 x agarose gel loading solution was mixed with the samples. Samples were loaded onto the gel alongside a 100bp, 1 kb, or lambda DNA-*Hind*III digest ladders (NEB) and run at 1-5 V/cm. The gel was visualised at  $\lambda = 300$  nm and photographed using a UVP video camera.

10

#### Recovery of DNA from agarose gels

DNA was excised from agarose gels and recovered using a Qiaquick gel extraction kit (Qiagen) and eluted in either sterile reverse osmosis purified water, Tris-HCl (10 mM, pH 8.5) or TE buffer.

15

#### End-filling

Filling the recessed 3' termini created by digestion of DNA with restriction enzymes was done using *E. coli* DNA polymerase Klenow fragment. In a typical reaction 1 unit of enzyme was

20

added per  $\mu\text{g}$  DNA along with 250  $\mu\text{M}$  each dNTP. The reaction was incubated at 25°C for 15-30 min and stopped by adding EDTA to a final concentration of 10 mM.

#### Phosphorylation of DNA

- 5 PCR products were treated with T4 polynucleotide kinase at 37°C for 30 min, following the method described by Sambrook *et al.*, 1989. The enzyme was inactivated by incubating at 65°C for 20 min.

#### Dephosphorylation of linearised vectors

- 10 To avoid self-ligation of linearised vectors, 5'-phosphate groups were removed using shrimp alkaline phosphatase (SAP) following the manufacturer's guidelines. In a typical reaction 1 unit of SAP was added to the restriction mixture for the last hour of the DNA restriction reaction. The enzyme was inactivated by incubating at 65°C for 20 min.

#### Ligations

15 DNA ligations were performed as described by Sambrook *et al.*, (1989) using 1 unit (U) of T4 DNA ligase in a total volume of 15  $\mu\text{l}$  and incubating for 12-16 h at 16°C.

#### Maintenance of bacterial cultures

- 20 Viable cells were stored as glycerol suspensions by freezing 0.5 ml of the respective culture at -80°C with glycerol at a final concentration of 10%. Single colonies of *A. garbadinensis* and *A. liguriae* were obtained by streaking 50  $\mu\text{l}$  from a fermentation broth or glycerol stock onto either medium 65 or ABB13 plates.

#### Polymerase chain reaction

- 25 Polymerase chain reactions (PCRs) were performed on a Stratagene Robocycler Gradient96. In a typical reaction 100-200 ng template DNA was mixed with 20 pmol of each oligonucleotide primer and dNTP's at 250  $\mu\text{M}$  each. Thermophilic DNA polymerase buffer as supplied by the manufacturer and DMSO made up 10% (v/v) each of a final volume of 50 or  
30 100  $\mu\text{l}$  reaction mixture. A typical reaction began with an initial cycle of 1 min denaturation (94°C), 1 min, Y°C (annealing) and 30 seconds - 3 min extension (72°C), at which point 5 units of thermophilic DNA polymerase was added. This was followed by 30 cycles of 94°C for 1 min, Y°C (annealing) for 1 min and 72°C for Xmin and a final cycle of 72°C for 2X min. The extension time X, was 1 min per kb of product when *Taq* polymerase was used and 2 min per  
35 kb of product when *Pfu* polymerase was used. The annealing temperature Y was 55°C and

49°C in the generation of pAGvar1 and pAGvar2 respectively. The conditions used for the generation of SBdel-1 and SBdel-2 were as described in the Redirect protocol (Gust *et al.*, 2004).

#### Primers

| Primer name | SEQ ID NO: | Sequence 5'-3'                                                                                           |
|-------------|------------|----------------------------------------------------------------------------------------------------------|
| O/AGvar01bF | 303        | TTCTAGACGTTGTTCTCCCATTTTCAC                                                                              |
| O/AGvar02bR | 304        | AAGATCTTCGAAGGTGAGCTCGCCGAA                                                                              |
| O/AGvar03F  | 305        | GATCTTCGCGAGGACCGCACCATCTACGCCGCCAGCA<br>GCGGCTGGGTGTGTACACTGACGATCGAGTGCGGCAC<br>CGTGATCTGCGCCTGCTGAC   |
| O/AGvar04R  | 306        | CTAGGTCAGCAGGCGCAGATCACGGTGCCGCACTCGA<br>TCGTCA GTGTACACACCCAGCCGCTGCTGGCGGCGTA<br>GATGGTGCGGTCTCTCGCGAA |
| O/AGvar05F  | 307        | GCCTGCTGACCTAGGTGACGATCGT                                                                                |
| O/AGvar06r  | 308        | TGAATTCGGCTGCTCCCCGCGCGAAAT                                                                              |
| O/SB50F     | 309        | ATTCGCCCCGGGAAGTCCACCGAAAGGAAGACACACCAT<br>GATTCCGGGGATCCGTGAC                                           |
| O/SB51R     | 310        | GGGCGATGCCCGCCCCGGGCCGGAACGATCGTCGAT<br>CATGTAGGCTGGAGCTGCTTC                                            |
| O/SB52F     | 311        | AAGTATATATGAGTAACTTGGTCTGACAGTTACCAATG<br>ATTCCGGGGATCCGTGAC                                             |
| O/SB53R     | 312        | GCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCA<br>TGTAGGCTGGAGCTGCTTC                                           |

5

#### Preparation of plasmid DNA

Plasmid DNA was prepared on a small scale (less than 20 µg preparation) by inoculating 3 ml of sterile 2TY or LB containing the appropriate antibiotic with single colonies picked from 2TY (or LA) agar plates. The cultures were incubated overnight (12-16 h) at 37°C and 250 rpm.

- 10 The cells were collected by centrifugation at 12,000 xg for 1 min and plasmid DNA obtained using Wizard (Promega) Miniprep kits according to the manufacturer's guidelines. In the case of larger preparations of up to 100 µg of plasmid DNA, 30 ml of 2TY cultures were grown and plasmid DNA extracted using a Qiagen Midi-prep kit, following the manufacturer's instructions. All plasmid preparations were checked by a combination of restriction analysis and/or
- 15 sequence analysis.

#### Preparation of cosmid DNA

Cosmid DNA was prepared by inoculating 50 ml of sterile 2TY or LB containing the appropriate antibiotic with single colonies picked from 2TY (or LA) agar plates. The cultures were incubated overnight (12-16 h) at 37°C and 250 rpm. The cells were collected by

20 centrifugation at 4,000rpm (Heraeus sepatech Megafuge 2.0R) for 20 min and Cosmid DNA isolated using a Qiagen Midi-prep kit according to the manufacturer's guidelines.

Preparation and transformation of electrocompetent *E. coli* cells.

Electrocompetent *E. coli* DH10B were prepared by the method of Dower *et al.* (1988). Aliquots (60 µl) of competent cells were thawed on ice and 1.8 µl of ligation mixture or plasmid DNA added. The mixture was placed into an electroporation cuvette (Sigma 0.1cm) and transferred to the electroporator (Stratagene electroporator-1000). A potential difference of 1.8 kV/mm (25 µF, 200 Ω) was applied and 0.5 ml of 2TY or LB medium subsequently added. The cells were then incubated at 37°C for 45-60 min to allow expression of the antibiotic resistance genes, prior to plating on the appropriate selective medium.

Preparation of genomic DNA

Genomic DNA templates were prepared using the procedure described by Kieser *et al.* (2000).

Conjugation procedure for *Actinoplanes* sp.

Intergeneric conjugation between *E. coli* and *Actinoplanes* sp. was performed following the procedure described by Heinzelmann *et al.* (2003), except, the strain *E. coli* ET12567/pUB8002 (Kieser *et al.*, 2000) was used in place of the strain *E. coli* ET12567/pUB307 (Flett *et al.*, 1997). Exconjugants were transferred and patched out over an area approximately 1 cm<sup>2</sup> onto medium 65 or ABB13 containing 50 µg/ml nalidixic acid and the relevant selective antibiotic. These plates were incubated at 30°C for 4-7 days prior to being used as inoculum for broth cultures.

Conjugation procedure for *Streptomyces* sp.

Intergeneric conjugation between *E. coli* and *Streptomyces* sp. was performed following the procedure described by Kieser *et al.*, 2000. Exconjugants were transferred and patched out over an area approximately 1 cm<sup>2</sup> onto SFM containing 50 µg/ml nalidixic acid and the relevant selective antibiotic. These plates were incubated at 30°C for 4-7 days prior to being used as inoculum for broth cultures.

Table 5- Bacterial Strains

| Name                                        | Description/Use                                                                             |
|---------------------------------------------|---------------------------------------------------------------------------------------------|
| <i>Actinoplanes garbadinensis</i> ATCC31049 | Isolation of the biosynthetic gene cluster for the production of actagardine.               |
| <i>Actinoplanes garbadinensis</i> Δ         | <i>Actinoplanes garbadinensis</i> ATCC31049 in which the <i>actA</i> gene has been removed. |

| Name                                      | Description/Use                                                                                                                                        |
|-------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>actA</i>                               | Expression of variant <i>actA</i> genes                                                                                                                |
| <i>Actinoplanes liguriae</i> NCIMB 41362  | Isolation of the biosynthetic gene cluster for the production of deoxy-actagardine B. Expression of variant <i>ligA</i> genes.                         |
| <i>Escherichia coli</i> XL1-Blue MR       | Generation of a cosmid library.                                                                                                                        |
| <i>Escherichia coli</i> DH10B             | Routine cloning.                                                                                                                                       |
| <i>Escherichia coli</i> ET12567           | Isolation of non-methylated DNA.                                                                                                                       |
| <i>Escherichia coli</i> ET12567/pUZ8002   | Intergenic transfer of DNA via conjugation.                                                                                                            |
| <i>Escherichia coli</i> BW25115/pIJ790    | Strain containing the lambda red recombination plasmid pIJ790. Facilitates the targetted recombination of a cassette flanked by FLP recognition sites. |
| <i>Escherichia coli</i> DH5α/BT340        | Strain containing the plasmid BT340 facilitating FLP-mediated excision of disruption cassettes.                                                        |
| <i>Micrococcus luteus</i> ATCC4698        | Bioassay test organism.                                                                                                                                |
| <i>Streptomyces lividans</i> 1326         | Host organism for the heterologous expression.                                                                                                         |
| <i>Streptomyces coelicolor</i> B757       | Host organism for the heterologous expression                                                                                                          |
| <i>Streptomyces cinnamomeus</i> DSM 40005 | Host organism for the heterologous expression                                                                                                          |

### Antibiotics

Antibiotic stock solutions were prepared in water (unless stated otherwise) and filter sterilised by passing through a 0.22 µm Millipore filter. Solutions dissolved in ethanol were not sterilised (Sambrook *et al.*, 1989). All antibiotics were stored at -20°C. In media where apramycin was used, MgCl<sub>2</sub> was added to a final concentration of 10 mM (from a stock of 2.5 M).

|                      | Stock solution      | Working concentration |
|----------------------|---------------------|-----------------------|
| Ampicillin (amp)     | 100 mg/ml           | 100 µg/ml             |
| 10 Apramycin (apra)  | 100 mg/ml           | 50 µg/ml              |
| Carbenicillin (car)  | 100 mg/ml           | 100 µg/ml             |
| Chloramphenicol (cm) | 25 mg/ml in ethanol | 25 µg/ml              |
| Kanamycin (kan)      | 50 mg/ml            | 50 µg/ml              |
| Nalidixic acid (na)  | 25 mg/ml            | 25 µg/ml              |

15

Cassettes

| Name    | Size (bp) | Source                                                              | Description/Use                                                                                                                                                                                                                |
|---------|-----------|---------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SBdel-1 | 1462      | PCR using the primers O/SB50F and O/SB51R and pIJ773 as a template. | Contains an origin of transfer ( <i>oriT</i> ) and apramycin resistance gene flanked by FLP recognition target sites. The 5' and 3' regions are homologous to DNA flanking the <i>actA</i> gene from <i>A. garbadinensis</i> . |
| SBdel-2 | 1462      | PCR using the primers O/SB52F and O/SB53R and pIJ773 as a template. | Contains an origin of transfer ( <i>oriT</i> ) and apramycin resistance gene flanked by FLP recognition target sites. The 5' and 3' regions are homologous to DNA flanking the ampicillin resistance gene from SuperCos1.      |
| HEapra  | 5247      | pMJCOS1                                                             | <i>SspI</i> fragment isolated from pMJCOS1. Cassette consists of an apramycin resistance gene, origin of transfer ( <i>oriT</i> ), attachment site ( <i>attP</i> ) and $\phi$ C31 integrase.                                   |

Vectors

| Name    | Size (kb) | Resistance marker | Source      | Description/Use                                                                                                                                                                                                                                      |
|---------|-----------|-------------------|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| pAGvar1 | 3.1       | amp               | This study. | 449bp PCR fragment generated using the primers O/AGvar01bF and O/AGvar02bR and template pLITAG01 cloned into pUC19 previously digested using <i>SmaI</i> .                                                                                           |
| pAGvar2 | 2.8       | amp               | This study. | 91bp PCR fragment generated using the primers O/AGvar05F and O/AGvar06R and template pLITAG01 cloned into pUC19 previously digested using <i>SmaI</i> .                                                                                              |
| pAGvar3 | 3.2       | amp               | This study. | <i>XbaI</i> fragment (~450bp) cloned into pAGvar2 previously digested using <i>XbaI</i> .                                                                                                                                                            |
| pAGvar4 | 3.3       | amp               | This study. | Annealed oligonucleotides O/AGvar03F, O/AGvar04R ligated to pAGvar3 previously digested using <i>BglII</i> and <i>AvrII</i> .                                                                                                                        |
| pAGvarX | 6.3       | apra              | This study. | <i>XbaI</i> – <i>EcoRI</i> fragment (~650bp) from pAGvar4 ligated to pSET152 previously digested using <i>EcoRI/XbaI</i> . Variant actagardine genes can be assembled and introduced into the hosts chromosome via the attachment site <i>attP</i> . |
| CosAL02 | 47.2      | amp and neo.      | This study. | 40402 bp <i>Sau3AI</i> DNA fragment from <i>A. liguriae</i> cloned into SuperCos1 previously digested using <i>BamHI</i> .                                                                                                                           |

| Name          | Size (kb) | Resistance marker   | Source                            | Description/Use                                                                                                                                                                                                        |
|---------------|-----------|---------------------|-----------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CosAL02HEapra | 49.1      | amp and apra.       | This study.                       | CosAL02 in which the gene encoding neomycin has been replaced with the HEapra cassette.                                                                                                                                |
| CosAG14       | 45        | amp and neo.        | This study.                       | 38168 bp <i>Sau3AI</i> DNA fragment from <i>A. garbadinensis</i> cloned into SuperCos1 previously digested using <i>Bam</i> HI.                                                                                        |
| CosAG14ΔA     | 46.3      | amp, neo and ampra. | This study.                       | CosAG14 in which the <i>actA</i> gene has been replaced by the cassette SBdel-1.                                                                                                                                       |
| CosAG14ΔB     | 44.9      | amp and neo.        | This study.                       | CosAG14ΔA in which the cassette SBdel-1 has been removed by FLP-recombinase leaving an 81bp scar.                                                                                                                      |
| CosAG14ΔC     | 45.5      | neo and apra.       | This study                        | CosAG14ΔB in which the ampicillin resistance gene has been replaced with the cassette SBdel-2.                                                                                                                         |
| CosAG14HEapra | 46.9      | amp and apra.       | This study.                       | CosAG14 in which the gene encoding neomycin has been replaced with the HEapra cassette.                                                                                                                                |
| pIJ773        | 4.3       | amp and apra.       | John Innes Centre (JIC), Norwich. | Redirect template (Gust <i>et al.</i> , 2003) used to generate the cassettes SBdel-1 and SBdel-2.                                                                                                                      |
| pLITAG01      | 6.1       | amp.                | This study.                       | 3263bp <i>Nco</i> I fragment isolated from <i>A. garbadinensis</i> (19955-23217 CosAG14rc) cloned into pLITMUS28 previously digested using <i>Nco</i> I.                                                               |
| pLITMUS28     | 2.8       | amp.                | New England Biolabs (NEB).        | Routine cloning                                                                                                                                                                                                        |
| pMJCOS1       | 9.8       | amp and apra.       | JIC, Norwich.                     | SuperCos1 in which the gene encoding neomycin has been replaced by an <i>Ssp</i> I fragment consisting of an apramycin resistance gene, <i>oriT</i> , <i>attP</i> and $\phi$ C31 integrase. Source of HEapra cassette. |
| pSET152       | 5.7       | apra.               | NRRL B14792                       | Conjugative plasmid which can facilitate introduction of DNA into the host's chromosome via the <i>attP</i> site.                                                                                                      |
| SuperCos1     | 7.9       | amp and neo.        | Stratagene.                       | T3 and T7 promoter regions flanking a unique cloning site.                                                                                                                                                             |
| pUC19         | 2.7       | amp.                | NEB.                              | Routine cloning                                                                                                                                                                                                        |

High performance liquid chromatography

HPLC analyses were performed using a Hewlett Packard 1050 series HPLC system with the parameters as described below:

|    |                   |                                                             |        |      |  |
|----|-------------------|-------------------------------------------------------------|--------|------|--|
| 5  | Column:           | Zorbax SB-C18, 4.6x150 mm, 5 $\mu$                          |        |      |  |
|    | Mobile Phase A:   | 30% Acetonitrile in 20 mM potassium phosphate buffer pH 7.0 |        |      |  |
|    | Mobile Phase B:   | 65% Acetonitrile in 20 mM potassium phosphate buffer pH 7.0 |        |      |  |
|    | Flow rate:        | 1ml/min                                                     |        |      |  |
|    | Gradient:         | Time 0 min                                                  | 100% A | 0% B |  |
| 10 | Time 10 min       | 0% A                                                        | 100% B |      |  |
|    | Time 11 min       | 0% A                                                        | 100% B |      |  |
|    | Time 11.2 min     | 100% A                                                      | 0% B   |      |  |
|    | Cycle time        | 15 min                                                      |        |      |  |
|    | Injection volume: | 10 $\mu$ l                                                  |        |      |  |
| 15 | Detection:        | 210 nm                                                      |        |      |  |

High performance liquid chromatography –mass spectrometry (HPLC-MS)

HPLC-MS analyses were performed on a Hewlett Packard 1050 series HPLC system linked to a Micromass Platform LC (operated with MassLynx version 3.5 software) with the following parameters:

|    |                      |                                                         |  |  |  |
|----|----------------------|---------------------------------------------------------|--|--|--|
|    | Column:              | Agilent Zorbax SB-C18 150 x 4.6mm 5 $\mu$               |  |  |  |
|    | Flow rate:           | 1ml/min                                                 |  |  |  |
|    | Mobile phase:        | A 10% acetonitrile, 0.1% formic acid 90% water.         |  |  |  |
| 25 |                      | B 90% acetonitrile, 0.1% formic acid, 90% water.        |  |  |  |
|    |                      | Linear gradient A to B over 10 minutes, hold 1 min, B-A |  |  |  |
|    | Wavelength:          | 200 – 400nm                                             |  |  |  |
|    | Injection volume:    | 10 $\mu$ l                                              |  |  |  |
|    | Post column split:   | 1:10                                                    |  |  |  |
| 30 | Mass spectrometer:   | Micromass Platform LC                                   |  |  |  |
|    | Mode:                | Electrospray positive                                   |  |  |  |
|    | Nitrogen flow:       | 380 l/hr                                                |  |  |  |
|    | Capillary voltage:   | 40V                                                     |  |  |  |
| 35 | Skimmer lens offset: | 5V                                                      |  |  |  |

Deposit

NCIMB 41362 was deposited under the Budapest Treaty on 7 December 2005 at NCIMB Ltd, Aberdeen, AB21 9YA, Scotland, UK, by Novacta Biosystems Limited.

5 References

Altena, K., Guder, A., Cramer, C. and Bierbaum, G. (2000). Biosynthesis of the lantibiotic mersacidin: organization of a type B lantibiotic gene cluster. *Applied Environmental Microbiology* 66(6): 2565-71.'

- 10 Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (2002). *Current Protocols in Molecular Biology* (5<sup>th</sup> edition). Wiley Interscience Publishers.

Bierman, M., Logan, R., O'Brien, K., Seno, E. T., Nagaraja Rao, R. and Schoner, B. E. (1992).

- 15 Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* 116(1): 43-49.

Chatterjee, S., Chatterjee, S., Ganguli, B. N., Chatterjee, D. K., Jani, R. K. H., Rupp, R. H., Fehlhaber, H-W., Kogler, H., Siebert, G. and Teetz, V. (1992) Antibiotic, mersacidin, a process  
20 for the preparation thereof and the use thereof as a pharmaceutical. US Patent 5,112,806.

Dower, W. J., Miller, J. F. and Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Research* 16(13): 6127-6145.

- 25 Flett, F., Mersinias, V. and Smith, C. P. (1997). High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting *Streptomyces*. *FEMS Microbiology Letters* 155(2): 223-229.

- Gravesen, A., Kallipolitis, B., Holmstrøm, K., Høiby, P. E., Ramnath, M. and Knøchel, S.  
30 (2004) pbp2229-Mediated nisin resistance mechanism in *Listeria monocytogenes* confers cross-protection to class IIa bacteriocins and affects virulence gene expression. *Applied and Environmental Microbiology* 70(3): 1669-1679.

- Gust, B., Challis, G. L., Fowler, K., Kieser, T. and Chater, K. F. (2003). PCR-targeted  
35 *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *PNAS* 100(4): 1541-1546.

Gust, B., Chandra, G., Jakimowicz, D., Yuqing, T., Bruton, C. J. and Chater, K. F. (2004). A Red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Advances in applied microbiology* 54: 107-128.

5

Gust, B., Chater, K. F. and Kieser, T. E. (2002). Methods and materials for targeted gene disruption in actinomycete bacteria. Patent Application WO 02/103010 A1.

Heinzelmann, E., Berger, S., Puk, O., Reichenstein, B., Wohlleben, W. and Schwartz, D.

10 (2002). A glutamate mutase is involved in the biosynthesis of the lipopeptide antibiotic friulimicin in *Actinoplanes friuliensis*. *Antimicrobial Agents and Chemotherapy* 47(2): 447-457.

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. and Hopwood, D. A. (2000).

Practical *Streptomyces* Genetics. Norwich, John Innes Foundation.

15

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A laboratory manual*. New York, Cold Spring Harbor Laboratory Press.

Vertesy, L., Herbert, K., Schiell, M. and Wink, J. (2000). Lantibiotic related to actagardine, and processes for the preparation and use thereof. US Patent 6,022,851.

20

Sequence Listing**SEQ ID NO:1 Actagardine B:**

SSGWVCTLTIECGTLVCAC

**SEQ ID NO:2 Actagardine B variant VV:**

SSGWVCTLTIECGTVVCAC

**SEQ ID NO:3 Actagardine B variant LI**

SSGWVCTLTIECGTLICAC

**SEQ ID NO:4 Actagardine:**

SSGWVCTLTIECGTVICAC

**SEQ ID NO:5 - SEQ ID NO:10**

(Null)

**SEQ ID NO:11 Ala-Actagardine B:**

ASSGWVCTLTIECGTLVCAC

**SEQ ID NO:12 Ala-Actagardine B variant VV:**

ASSGWVCTLTIECGTVVCAC

**SEQ ID NO:13 Ala-Actagardine B variant LI**

ASSGWVCTLTIECGTLICAC

**SEQ ID NO:14 Ala-Actagardine:**

ASSGWVCTLTIECGTVICAC

**SEQ ID NO:15-SEQ ID NO:21**

(Null)

**SEQ ID NO:22 pre-pro-Actagardine B variant VV:**

MSAITVETTWNKNTDLREDLTAHPAGLGFGLSFEDLREDRTIYAASSGWVCTLTIECGTVVCAC

**SEQ ID NO:23 pre-pro-Actagardine B variant LI**

MSALAIEKSWKDVLDLRDGATSHPAGLGFGLTFEDLREDRTIYAASSGWVCTLTIECGTLICAC

**SEQ ID NO:24-SEQ ID NO:99**

(Null)

**SEQ ID NO:100 CosAG14 (38168bp)**

GATCGGCCCGCTCAGCTGGAGAAGCGCTGGCTGGCGCGGCTCACCCGCAACGGCCACCCGCATCCGACCCGGGCG  
 CTGGCGCTGCGGATGGGCGCGGTGGAGTTCGCGGGGCGAGCTGGCGCTGCAGCTCGCCGACATGCGCAAGCCCGGCA  
 CGCCGTCAGCGACCTGACCCGCTGCTCGACGCGAGCTCGCTCGCCGGTCCCGCGCCGCGCCGATCAACTCCGGCAGC  
 ACGGCGCGAGCTTCAATCAGCCGGCCCGGCGCACCCCTACGCCGGTGCAGCAGGCATGCCCCGAGACCAGCGCGG  
 CCGCAGAGGGCCGGTCCGGCGCCGGACCGGTACGCGCGCCGCAATCAGGCCGGCAGGCCCGCCGGAGCCGGACCG  
 GTCGGCGCGCCCGCAATCAGGCCGACAGGCCCGCCGGGGTCAGGCTGGTTCGGCTGCGCCGGCCGGCAGGCCGGC  
 GAGCAGCGCGTCGAGCGTCTCGGCATCCGCCAGCCGTTGACTGACACTGCGCCTGCCGCCGTGACCGTGACCTCA  
 AGCGGGTCCGGTGTACCGCCACGTGATCGGCGCGTCTCGATGCGGTGCGCTCAAGCGGATGAACTGCTGGTTTCG  
 TCGAGCCGACCCAGGGCCGCTCGTGGTCCAGCAGGTCTGCGGGAAGGGGCCGGCCACCAGCAGGTCCGTGGCGTC  
 GAGCAGTGCGGCGACGTGCGCCCGGCGCTCCGCGGCGAGAGCGCGCAGACGGGCGTACTCGTAGCCGGTGAACGTC  
 ATGACCGAGCGACCCGCTGCCGCACCCCGGCGGCCACCGCGGCGAGACCGGCCGCTGCTCGAACGGCTCGCCGC  
 CGAGCAGGCTGACACCCGTGGTGGTGGCGGAGCACTCGGTCCACCACTCGGCGGGTGGCGCCGGCACCCACC  
 CGGTACGCCGAACAAATGCGGATTGAAGCATCCGGCGCACCCGATGGTGCAACCCTGCACCCAGATCGCGGTCCGC  
 TCGCCCGGTCTTTCGGCGGTGGTCCGGTTCAGGAACCGGGCCACCCGCACGAGCGGCGGCTCAGACATCGACTATC  
 CGACCGAAAGGCGTCTGGTTCGCTTTCGGCGGCGGCTCGCACGACAGCCGGTCCCGGACCGCGACGAACCTCGTGTG

GTGCGAGCCCCGGCCACGGCCGCGAACTCGCCGAGACTGGCGAATCCGCCCCGCGCCTGCCGCGCCGCCACGACGTC  
CGCCACCCGCTGCGGGGTGAGACCGGGACGGCGGCCAACTGTTGCGCGTCCGCGGTGTTGACATCGAGCCGCTCG  
GGCGCCGGCTCCAGCAGCGAGCCCGGCACACCCCCAGCCGGCGAGCCAGCGCGGGCTCCAGCGGGCCGGGCGCGG  
GCCAGCGGGCCAGGCCGCTGTGGGCTCAGCCTGCCGGCCGGCTGCGGGCTCAGCCTGCCAGCCCGCTGCCGGCTC  
AGCCTGCCAGCCCGCTGCCGGCTCAGCCTGCCAGCCCGCTGCCGGCTCAGCGGGCCGGCCCGCCGCGCGGCTCAGCC  
AGCGGGCCCCCTCGCGGGCTCAGGCCCGGGAGTCCCTCCGTAGAACTCTTGCGGCGAAGGAACGACACCCTGTAGTT  
GCGGCGGCAGCGGAGTCGCCACGGGAGCGGGCGCGTTCGCCCCACGGCGCCTGCTGGGGCTGGGCCACACCATGAAC  
GTGAGCGGCACGCCAGCGCAGCCAGGAGGCGTTGATGACGAACGCGTGCGCGACGCTGATCGGCCACACCATGAAC  
ATGATCGAGAACGCGAGGTTCTGCTGGAGCGAGTCTCGGCGCCCTCACTGCCGAGAAAGAAGCAGAAGTTGGCCA  
CGACGGTGTAGAGGATCCCGCGATCCACCACACCGGCCCGCGGGCGCGGATGCCGACGTAGAGGAACCCGACCGC  
GGAGAAACAGCTGAAGGGCACCATCGGCGCGATCACCCAGGCGCTGTGCGCGAGGCGCCACGACAGCCGCGCCGGC  
CCGGTCCGCTTGCCCGGCTCGTGACCTGGATGCGCGCGATAGGACGGATACGGCTGCTGGTGAGCAGGCGGGAACG  
GCGGCAGCGGCTGCTGACCGGAAGCAGGCGGGAACGGCGGCAGCGGCTGCTTGCCGAACGACGGCGGGTACGCCGG  
CTGCGAGGCGGGATACGGGTCTTGCCCGGGTTAGGCGGTGTCCACGTACGGTTACGATCCGTTCCCGCGCGGT  
GTGCGCCTGACGCGTGTAACTCCGCGGTGTACGCACTGTGCCGACCTGCCCGCCGACGAGATCCTCCAGAACGGCC  
ACGTATCCAGGACGCGCTCGCCGAGAATGTCCACGGTCTCGGCACTGACCTGGCCGTGGGACCGACCGTGACGA  
CGATCCGAGGATTCTCCGTATGCCCGCTCCCTCTCCACCTCGAAGACCAGGCGCACACTGGCGTCTCTGGCCGACC  
GACTCCGACTCCAGCCGACGCCCGGCTCGGGCGCCTGCGCGCGCAGCCGCTCCAGGACGGCCTGTGGACCCGCC  
GCCCCGTACGCGGTGTGACGGTCTGTCATCAACTCGACAGCCCTGGCTGATCGACGCCGGTGACGTGCGCGGCCCA  
GATGCCGTCCGCGCCGCGGGTGAACGTGCGGGCGCTGCGTCCAGGTGCGCGAGATCGTGTGCGCGCGGCCGTC  
ACGGTGGCTCCGGTGTGCGCGCAGCGCCGCATCGAGCAGGGTGACATCGCGCATGCGGGTCTGCACCTGGCAGATCA  
GCCGGCCGTGCTCCATCCGGCCCGCGGCGGCCGTGAACCGCGGCCGCGCCGGCCATCGCCAACGGCACCAGCAGCAG  
TGAAACGCTCACCATGCCCCCGTTCTCGCTGTGGCCACACCTATCGGCAGGGTGCGACAAATCATCGGTTGGTG  
AGGTCCAGTCTGTCGGTGCCGTGGCCGACACGGCGGATTGCGGGCCAGCCGCGCAGGGCGTCGACCCGCTCGG  
CCTGCGTACGCTGAGCGGCACGATGCTCATACCGCGCGTACGAGATCGTTCGCGCCGAGTGGCGGCGGTTCCGA  
GAACCGCTCGAACAGTCCCGCAATGACCGCCTGCTATCTCCGCGCGGAGTAGCCCTCGGTACGCCCGGCCAGC  
TCGGTGAGCAGCTCGGCATCGACCCGCACTCGCCGCGGCACGCCGGTGCCGACGCGCCCGCCGAGGTGCAACCC  
GCCACACCGCGACCCGCTCGGACCGGCTCGGCAGATCCACGAAGAAGCTCTCGTCAAGCGTCCCTTGCGCAACAG  
CTCCGGCGGCAGCCCGTCAAGTCTGTGGCCGTGCGGATCACGAAGACCGGAGTCCGTTCTCCTGCATCCAGGTG  
AGGAAGGTGCCGAAGACCCGGGCACCGGTGCCGAATCACCCCGGTGCCGCGGCGAAGCCCTTCTCGATCTCGT  
CGACCCACAGGACGCACGGGGCGACCGCCTCCGCCGTACGCAATGCGGTGCGCATGTTGTGCTCGTGAACCGAC  
CAGGCCGGAAGAAGACACGGCCGATGTCAAGCGCAGCAGAGGCGAGTTCCAGGCGGTGCGGACCGCCTGGCGGTC  
AGCGACTTGCCGCGAGCCCGGCACGCCGGTGATCAGTACGCCGCGCGGCGCGGGCAGGCCGTACCCGGCCGCTCGT  
CCAGCCAGGATCCGTTGCGTTTGACAGCCAGGCCCTGAGGTTCTCCAGGCCCGGACGTCGTGAGCAGCGGTTCC  
GGCGGCCATGAAGTCCAGCAGCCCGATTGCGTACGGTCTGGCGCTTCTCCTCGTGCACGATCTCCAGGTGCGGC  
AGATCGAGCAGCGCTCGTTACCATGGCCGGGCTACGCGTCTCCTCGCTCCTGATGTTAGGCGCGCGG  
CCGTGACGAACCGCTCCCGGCTCTGCTCGTCTGAGTCCACCCGACGCCCGCGGACGCGGTGTTCCCGGGGTCACCAT  
GGCGTGCAGCAGAGCGCGCAACTCCAGCTGGCCGGGACGCGGGAAGTGCACGATCGTGACGTCTTGTCTAGCTCG  
ACCGGCAGGTCCAGTACCGGCGAGAGCAGCAGAGTGTCCGCGGGCTGCGGCGGGCCCGGAACGCCTGGGCGATGT  
CGCGCACCAGCCGACGACCTCCGGGCTCTGCGCGAACAGCGGGTGACGTTGCGGAAGACGAAGACAGCCGGTTC  
GTGATGCGCTGCACCGCCCTGAGCGCGTGGTGGCCGCTGCGCGCCCGAGCGGGCCTCGCCGTTCGGCTGGACC  
AGACCGGCGGTGACGACAGGTTCCAGACGGCACGTGGCAGCGGACGAGGTGCGCGTTCGCCGGCGATGCCGGCGA  
GGTGGTGACGCGCCGCTGCTCCTCGTACGTCTCGAGGTGGAGCACCGGGAACCGGGCCTTCAACAGCTGCGCGAA  
CGTCTCCGCGAACGACCTCTCCACCGCGCATCTACGGCCGCGCGGCGCGGCGTGTGAGAGCCGGACGGGGTGC  
CGGTGGCGTGCCTGCGTCCGTTCCGTCCTTCCCGCGCTGACGGGTTCGTTGGGTTGATCGGTATGGCATCCTCCTG  
CTAGTCCGGGCGGTAGACGGCGGTGTCCCGGACACGGCGCGGGTTCGTACCCGAGCGCGGTTTCTCCTCGGT  
GGTGTCTGCTCGTCAGCGCCGGCGCCTGTGCACTCCACCGGTTTGCGCAGTGACGCCGCTATGCCGGGGTTCCG  
TCCGGCTCGTCCGTGGGATGCCGCTGGAGTTGTTCTCGGTTCATCGGTTCATGAACGGCGGACTTCCCTGCAAGAGCG  
CGGACAAACGTGGTTCGTTTATCCCGCAGCCGGGCGGGCAGTCTCCACCATGGAACCCATCGGTGACCTGACGAG  
CAGACCGGCGACTTCGCCGAAGAGCAGCAACTGACCGAGGTGACAGCCGAGGAGGACGACGACGGCGAGCCGGAGA  
GCCCGGATCGCTGGACCGGCGGATGGAATCCGACGGTCCCCCGTGACCTGCGGCTACGATCCCTACAGGTGCCC  
GCCCCGCCGACGACCCGAGGTTCTGAGCCGTTGTCTCCATCACCGAGCTGACCGAGTCCGTGCGTGCCTTCGCGC  
GCGAGCGCAACTGGGAGCAGTTCCACACGCCGAAGAACCTCGCGATGGCCCTGGCCGGCGAGGTGCGTGAGCTTCT  
GGCCGAGTTCAGTGGCTGACGCCGGAACAATCAGCCGCGGTTCATGCGCGATCCCGATCTCGGCCCGCGGGTCCGG  
GCCGAGATCGGAGACGTACGATCTATCTCGTACGCTGGCGGACGTGCTCGGCATCGACCTGGTTCGAGGCGGCCA  
CCGACAAGTTGGCGGAAGCCGGCGGCTACACCGTGAAGCCGCCCGGCGACTCGGCCGCGCAAGATCGATCAGCT  
GTAGTTACGCGAGAAGTTGTTGACCGTGAAGTTGGACTGACCGCGGTTGCCGCTGATCTCGAAGCCGAAGTGGCC  
TCGCCGACCGTACGTACCCCAACAGCCGTTGGTGGCAGCCAGTTAGGATCGCCAGGATGTGACGGTGCCGG  
AGTTCTGTTGGTACGGATGAACGAGAAGACCGCGTTGGCGCGGTTGACCCGCGGTAGACGTTCCAGGTGTGCCC  
GCCGACGACAGGTTGCGGACGTTCCGACCGCGCGGTTGGCGTCTGACTGTTCCGGCGATCGGCCCGACCGCGCCC  
TGCTGGTTGGTCCAGATCATGACCTCGTAGGCGTGGTTGTTCCGCCAGATGTGCTAGGTCTGAGTAGTCCCGC

TGGACGGCACCGACACGTTGAAGGAGCTCGTCAGAGAATTGAGCGAGCTCAGCGTACGGTTGAGGGTCTTTCCGGT  
GTGGGGTAGGACTTGACCCGCTGGTGCGCGGGTGATTGCGGACGACGCCCCAGTTGGTGCCGCTGCGCGCCAG  
ATGGTCTGGGTGCCGGCGCCGAGCCCCAGATGTTGTTGTGAGGATGTATCCGTTGTTGGTCCAGTTCGCCCCTG  
TCCCAGATCGACCCAGACGGCGTCCGTCGGTACGCCGTTGGTGGGCGGCTGCGAGCTCGGGCTGGTGGGGCCAGCC  
GATCCGGTGACGGCGGTCCCGTTGAGCGTGAACGTGCCCGGTGCCGGGTGGCGGGCGCCGAGCCGTTGAATCCGA  
ACGACACGGTCGCGTTGGTGCCGACGGATCCGTTGTAGCCGGCGTTGCCGGGCGGTGACCTGCGATCCGCTCTGGGT  
CAGGCTGGTGTTCAGGCCTGGGTGACGGTCTGCCCGCCGGAGTAGGACCAGACCAACGTCCAGTTCGTGAGCGGA  
TCACCCAGGTTGGTGATGGTGACGTTGGCGCCGAAGCCCGGGCCACTGGCTGCTCACGGTGTAGGCGACGCGGC  
AGCCGGCCGCGGCGGACGCCGGGAGCGCGGTGACGACGCCGGCACCCGGCGATGAGGGCCGCCGAGGCGGCCAGGGA  
GAGCGCGAGAGTACGTCGCGCATGCCGTTCCAATCTCGGAGGGGGAACACACGACGCTCGTCGAAGCGCTTCGAT  
AATGAAGGCTGGGGATCTACGAGTGTGATGCGGGATGAGGCCGATGTTACGACCCGCCGCCGCGGCGAGGCAACC  
GCTACCTTGGGCGCATGACGATCACCGAGACCGATCTCGCCACCTGCGCCGATGCGTCGACCTGGCCCGCGAGGC  
CCTCGACGACGGCGACGAGCCGTTCCGTTCCGTCCTGGTCTCCGCCGACGGCAAGGTGCTGTTTCGAGGACCCGAAC  
CGGGTGAGGCACGGCGACGCCACCCAGCACCCGGAGTTCGCGATCTCCCGCTGGGCGGCCGAGCACCTGACCCCGC  
GGGAGCGCGCCAGCGCGACGGTCTACACCTCGGGCGAGCACTGCCCGATGTGCTCAGCGAGCCACGGCTGGGTCCG  
CCTGGGCGCGATCGTGTACGCGGCCAAGCCGGGACTCCACCACGGGATTGACCGGTCTGGGGCTCGTGAGGACG  
CCGCTCGCCCCGCTACCGATCACACAGTGGTCCCGCGCGCTCGTGAGGGGCCCGTCCAGCCTTCGAGGCGG  
AGCTGCGGGAGCTACATCGCGCCCGCTTACCCCCAGCGCAGTAGCCGCCGGACGAAACCCGACCTTCTCGCCAAT  
GAACCGGACCTGCGAGCGTGTCTGGATTGACCGCGGAGGCCCCGAACGCCACAGGCTCGATCTCCGTACCCGTCGG  
ACCTCAGGGGCGAGTTCGAGGATCTGCGCCTTGACGACCGAGTGCATGACCTCCCATCCAGGAACGAGCCAGCAG  
ATCTTGACAGTGTGCGAAAGGCTCAGTACCAGCTACCGAGTACGTCCTCGGCGCACATCCGTGACACCGTGGGCA  
TGGATTTCGAGACCGGTTCGAATGTGCTCGACGCGTTCGCAACGCGCTTCCCGCTGCCGAGATGACCGATGAGCG  
TTGAGCAGGTACAGCTGCCGATCGACGTAATCAACCGGCACACCACTACCATCGCCGCAACCGCCGATGCCGTACG  
GCAGGCCCCGCGCGCGGCGGAGGTGAGGACGCCATGGACTCGCAAGCGTACGGCCAGTTGTGTGCTGCTTCTGCCGACC  
CGGTTTCAGCGTGTGTTTCGAGAGTGCATGCTCGCCATGACCGGCGAGCGCCGAAGCGTTGCAGGAAACACCTTCA  
ACCTGACAGGATGCCGTCACGACGTTTCAGATACACCGACGAGTCCGCGCGCTCGCACCATCGCCTCCGAGGTTTCGCC  
GCCGTGACCACACCACTTGTGCGCCGAAGCCGGGACTCCACCACGGGATTGACCGGTCTGGGGCTCGTGAGGACG  
CCCACCAGATCGCAGAAGCCATCCGCGGGAACAGTTGGGTGACGCGCGTCTCGGCGGAGTCCGGGGCCAGCCTCGA  
CGGTCTGGCGCTGGCGATCGACCCGCTCGGCACCTTGGCCGCTGGGGCGTCCGCTGGCTCATCGAGCAGTCCAA  
CCGCTACAAGACGCCCTGGACTGGCTGGCCGGCGACGTCGACGAGATAGCCGCCAAGCCGCCACCTGGCGCAACG  
TCGCCGCTTTCACCGACAGCGCCAGCAGGATTACGCCGATCGGCTCCGCACCGAGGTTCGCCGGCTGGTTTCGCGC  
CTCCGGCGACGCATACCGGGCCCATGCCAGCGAACACTTGGCCGCACTGAAAGGCATCAGCACCGCAGCCGGCGGC  
ATTTCTGTCGCCGCTCGAAGGCGCGGCGCTGCTCGTCAGCCTGGTGCGCGGAATTGTCCGCGACCTGATCGCCAGT  
TCGTCGCCACTCTGGCCGTTTCGGCTGCCACAATGGCTCGCCGCCGAAGGACTCACTCTGGGCCTCGCCACCCCGT  
TGTCGCGAGCCAAGTTGCGCCCTCGTCGCCCGCGGGTCAACAAGATCCAGCACTTCATACGGGCGCTGCTCAAC  
AGCCTTCGACGGCTGATCCCCATCGACGCGGTGGTGAACTCCTGGAGCGGCTCCGATGCTCACCAGCCGGC  
TCGCAAGGTTCAGCCCCCTCCACCCGCCCGGAGCCGACACCCGCGCCGCTACTACGCGGGAACCGGACGATC  
GGGAAACAAGCCCGAGGGCGACCTCGAACCGAACGAACCGCGCCCGGCGGAGGCTGACGCCAGAGACAGTACGCT  
CAGGCGTTCGTGGACGAGGTCTGACCAACCCCGGTGAGTCGCGGACACTCCGCGCAGTCCATCGCGGACCAGT  
TCAACGCCCGCGGATACTCTGCGGTCTGTCGAGCAGAGCACCAGGAGCGGCACCTCGGGGAACGCCATCCAGGTACG  
CATTCACGCCACCCGGATATACCAACATCCAGGTGCATCCGGGAGGAGGACGGCACACGCCGAGGGGTCCCG  
TATTGGAAATTTTCGACGAATACGGTTCGGGAAGATCTGGATCATCCCCGAAATTTTCGGGGCGCCGATGAACGTC  
GTGGGAATGTGGTGCGCTATGACAAATAAGATCATCAAAGTGGATCGGGTGGGATCCGCGGGGGAAGCTGCCACGC  
TCGTCGCCCTGGGCGTCGACATCGTCGGGTGATCTTCGTCCTGATCCGCGATTCTGTTGATGATCGACGCTGGA  
CGCCGAACAGGTTTCGCCACATCCGCGAAGCGTTCGCCACACGACGCTGGCCGTACCATGGACACCGGCACCGAG  
CCGGACCATGTGCTCGAACTCGCAAGGCGAATCGGCGCGGATCTGGTCCAGTCGATCAACGGTGTCTACCCGCGC  
TCCCAGTACGCGTCGCCCTGCGCAAGGCCGACATCGCGATTGTCTATGCGGGCACAGAGATCTCCACGACGACGA  
TCCCAGGCTGGGTCTTCAGCGCGTACGACGACACTCCGGACCTGAACGCCGCCCTGTTTCAAGTGGAGGTGCTTCCG  
GAGTACCTCGATTCTGGGAGTTTCTGCGTGACCGCTCCCCGGAATATGCCGACGAGTTCAGATCGAAGATCTCG  
ACGACCTCGGACGCGCTCGCCCCCTTGGTTGCCGGCCTGAATCTCACGGCTCAGAACCTCGACGAGATCAGGGCAGC  
GCTTCCCCACGTCCGCGGTCTGGCGCTCACCTCGCGCACGAGGCCGCGAGGAGCGACGCGCGGTTCTTCATGTTT  
CCCCACGTAGTGACATCCTTTTCGGGAGACTCGTCCGGGTGACCGGCATCGCGCGGGAGGGGGCTGGTGTGCGAGA  
GTCAGGCTGACGGCATTCCGGCGGGCCTGACGGGAGGAACCTCATGACCGGGCTGTTTACCATGCGCAAGCACGAG  
GAGGGCGCCGGGGAGGTTTCGGATCGTCGTGAGCGCGGAGATCGACGGCGATGTGACGCGGGCGCTCACCTGTTCA  
TCGCCAACGCCCGGAGCAGGACGGTGTCCGATCGGTGCTGGTGGATCTGGAGCCGGTGTCTGCTGTTGCCCGCAGC  
CGGGATCCGGTCTGCTGAGCGGGCGTGAAGCGGCTCTGCTGAGGGCTGCTCCTACGAGGTGTCTGCTGCTCAACGTC  
CACGAGGTTCGCGGATTCACTGCACGCCCGCGCGTGGCAGATCTGCTGCTCACCTGCCGGTCCGCTGAGCCGCT  
CAGCTCGGGTCTGCAAGCCCTCGTCGGCGCGTTCGCGCGCAGGTGTGCTCGGCCCGCGCGCCTGGTACCCCT  
TCTCCGCTGCCAGCCCGCGTACATGGCCTGCACCTTCTCGTACGGCTGGCCGGGCGGCGAGCAGCGGGATCGCCGG  
ATCGACCACGGCCTCGATCAGCGTCGGGCGGTCCGCGGACAGCGCCGCTCCACGCGCCGGTTCAGGTTCGACGGG  
TCGGTGATCCGCACTCCGCGCAGGCCGAGCAGCTCGGCGTAGCGCGCTACGGGACGTTCGGGAGTTCCTGCGACG

TGACGAAGCGCGGCTCGCCCTCGGTCTCGCGCTGCTCCCAGCTGACCTCGGCCAGGTCCCGGTTGTTTACGACGCA  
CACCACGAAGCGCGGCTCCGCCAGTCGCGCCAGCGGGCGGCCACCGTGATCAGTTCCGCCATGCCCGCCATCTGC  
ATCGCGCCGTCGCGCGGCCAGCACCACGACCGGGCGTTGCGGCGCCGCCAGCTTGCGGGCGAGACCGTACGGCAACC  
CGCACCCCATCGACGCCAGCGTGCTGGACAGATGCGCCGGGACACCGCGGGGACGCCGAGGTGCCGGGCGTACCA  
GTAGACGACCGAGCCGACGTGACCGCGACCTGCGCGTCCGCGGGCAGGTGGTCGGAGAGGGAGCGAATCACGTGC  
TGCGGGTTGACCGGTTCCGCGGGGCGCCGCGGCCCGGGCCGCGCTGATCAGGCGCCAGCGCTGCACCCATTCTCCA  
CGCGGGCGGGCCAGGCGCCGCTTGCCCGGCTGGGCGAGCAGTCCGAGCAGCTCACGCAAGGTTTCCACGGCGTCGCC  
CAGGAGCGGGACCTCAACCGGATACCGTACGCCAGACGGCGCCCGTCCACGTGATCTGTACAGGCGCCGCGCTGC  
CCCGGCCGGGGATAGAACTCGGTCCACGGGTCTGTTCTGTCGGATCAGCAGCAGCGTGTGCGAGTGCCGCATCAGCT  
CGGCGCTGGCCGTGGTGCCAGGTGCCCATGACGCCGCTGTGGAACGGCAGGTCTCTGTCGAGGACCGGCTTGCC  
GAGCAGCGAGGTGGTCACGCCGACCGAGCCGCTGCGCCAGCTCGACGATCTCGCGCTCCGCGCCGTACGCTCCC  
TGCCCCACCATGATCGCCACCCGCTCGCCGGAGCGGAGCACCTCGGCCGCTCGCGCAGGTCTTCCGGGCGGGGCA  
CCACCCGGGCGGGGCGCAGCCCCGCGCTCGTCGCCAGCACACCGTGCTCATGCGCTCGGGGTCGGGCGCCGGGGC  
CGTCTGTACGTGCTGCGGCAGGACACGCAGGTCCGGGCTGCGGGTTCGCGAGCGCGGTGCGGAACCGCGGGTTCGAGG  
AGCATCGGCACCTGCTCGCTCGTGTGCGCGGCTGCACGAACCTGGGCGCACACGTGCGCGAAGAGCCGCACAGGT  
CGATCTCCTGCTGGTACGCGCTGCCGAGCACCGTGCAGACCTGCTGGCCGACGATCGCGACGACCGGTTTGGAGTC  
GAGTTTCGCGTCTAGAGGCCGTTGAGCAGGTGCACCGCGCGGACCCCTGGGTGGCGAGGCAGATCCGGCACCA  
CCGGTGTAATTGGCGTGGCCGACGGCCATGAACGCCGCGCCCTCCTGTCGCGGGGCTGCACGAAGGTGGGCGGGC  
CGGCCCGGCGCAGCGCTCCGATCAGCCGTCGATGCCGTCCCGGAGTACCCGAAGACCCGTTCCACGTCCACGC  
GGTCAGACGCTCCACGACCAGTCCGACACGCTCCGCTCCGTACCGCTGCCCTCCTCACGTACACCGGGGGCTAC  
CCGCTCAGCCGCGCGGTACTCGCCTTGCGCGAAATGCGACTCGCGTGCACCTCACGGGCGGGGCGGACGCGGCACG  
GATCTGAGCTGGTTCCCGCCCGTTCTGATCCCCTGTGATCGGGGTACACACGTGCGGTGAGCAGGGACAAGACGT  
ACCGGCCGGTCCGTCCCGACGGGATCCGGACCACTTCCACGACGGTGTACTCGGGCGCATGCGGATCCGGTGCCT  
GGGCGAGCCGCGTCCGGGCGTACCGGAGATCGTGATGATCCAGGATGACCGTGAGTGACTATCTGCTGCCGGGT  
CTGGGCGCGCTCAGCGCTTGGACCCGGGTGCACCTGGTGGAGCTGCCGGGAGGCAGCGGCAGCGGCCGCGCCGC  
ACGATCTCACGGTCGAGGAGTACGCGCGGGCGCGGCCGACTGGCTGTGCGCCAGCGCTTGGGCGGGATCGTGCT  
GGCGGCCATTTCGAGCGGCACGCAGGTGCGCGCGGAGACCGCGCTGTTGTGCGCGGACGAGGTGGCCGGAGTGGTG  
CTGGCGGGCCCGCGCATCGACCCGGTGGCCCGCGCGGCCCTGCGGGTCTTCGCGCGCTGGTGGATCGACCGGCGCG  
GCGACCCGAAGAGCCTGGACGAGGTGCACAAACCCGAACGCGAGCAGGTCCGGTTCGGGCGGCTGTTCCAGGTGCT  
GCGCGCCACCTGCGTCACGACCTGGAGAAGCCGGTGGTCCGGCTCTGCGTGCCGGTGGTGGTATCCGCGGCAGC  
GAGGACCGGCTCGGCACCGCGCGGTGGGCCGCGCGGCTCGCCGATCTGGCTGCCGTGGGCGGCCGGTACGTGAGG  
TGCCGGGCACCCACTCGTTCTGCTGGCGTTACCCGCGAGGCTGGTCCGCGCCGATCCGTGAATTGCGCGGATGGTC  
GGTATCGGTCTCCGGCACCTGAACGACGTTTCGGTGGCATCGCCGCCACCGCGGGCGCGCATCCGCATCCACA  
GTCCGGGGTGAGGTCAGCGCCGTTTCGGAGGCGGAGTGTGATGATCGGTCAATTGACGTGCTTCCATGAGCCGCC  
CTACGGTCCGACAGAGAAATCGTTATCCGGACCAACCATCGACATCCGAATCGGTGGCCCCGAGCCGATCATCGAC  
CGACCCCGGGCCACCCCGCTCTTGCCAGAGGGGGCGAGGCACATGACGGAAGCTGATCATCGTCCGACCGTTCTC  
AGGAGAATTTCAGGCTGTGCGCGTGCTTCTCATGGTGTCTCGAGTGGGCTGGTTCGGCGCCCCGACGTCGACG  
CCGGCGGCAAGCCGACCATCTCCAAGGAGCGCTTCGCGAGCGCTCGGCGGCAAGGCCGTCGACCGGTACACCTCAC  
CAACGGGCGCCTGCAGGTGCGGATCCTCACCTACCGAGGCATCCTGCAGACCATCACGTTCCCCGACCAACCGCGGC  
CGCAGGGCCAACGTGACGCTCGGATTACAGGACCTCGACGAGTACGTGACGACGAAGAACCCGGCGTACTTCGGCG  
CCATCATCGGCCGCTACGGCAACCGGATTGCCGACGGCCGGTTACCCCTGGACGGCACGACCTACCAGCTGGCGAC  
CAACAACGACCCGAACCACTGCACGGCGGGGTGCTCGGCTTCGACAAGCGGGTGTGGGACGCCACGCCGATCCGC  
GACGGCGACAGCGTCGGGCTGCGCCTGACCTACACGACCCGCGACGGCGAGGAGAACTACCCCGGAACCTGCGCG  
TGACGATGACCTACACCGTCACCCGGCAGATGGGTATCCGGATGGACTACCGGGCGACGACCAGACGACGAT  
CGTCAACCTGACCAACCACGCTACTGGAACCTCGGCGGCGAGGGCACCGGGACCATCGACGACCACCTGCTGAAG  
CTCAACGCCAACCCTACACGCCGTCGACGCCACGCTGATCCCCGACGGGGCGATCGACGCGTTCGCCGGCACAC  
CGATGGACTTCCGCCGGCCACGCCGATCGGCGCGCGGACCGCGACCCGTTCCAGCAACTGGTGTACGGGCGCGG  
CTACGACCACAACCTGGGTGCTGAACCGCGAGGACGCCAGTTCCGGCGTCTCGAGTTCGCGGGCCGGGCGGTGAC  
CCGGACAGCGGGCGGCGAGCTACCATCTACACCACCGAGCCGGGCATCCAGTTCTACGGCGGCAACTTCTCGACG  
GCACCTGTACGGCACCGAGCGGCGGGCTACCGTCAGGGAGACGGTTTCGCCCTGGAGACACAGCACTTCCCGGA  
TTCTCCGAACCAACGCAACTTCCCGTCGACCGTCTTTCGACCGGGACAGACCTACAACCTGACGACCATCTACCAG  
TTCGGTACCGCCGACTGATCGCCTCAGGCCGGCGCCGCGGGACGGCGCGCCGCGGGCGCGCCGCGGCGGCGAGT  
GGCGGCGACTGGAGGCACACATGCCACGCATCCATCCGAAAGTCGAGGAGGCCGCTCTCTACGCTCGACCTGAACCG  
GACGACACGACGCGCGGCTGTTGTCCGGAACCGGGTGTTCAGCGCCTCGCTCGCCGCGGGGCGCGCTGCTGAGCGCA  
TGCAGCGACCAAGACGACGGCCAGAACCAGACCGAGGGCGCCGGTAACCTTCCCGGACACCCCGAGTGGCGGTTCA  
CATTCGTCAACACAGTGACCAACCAACCGTTCTTACCCCGACGAGTACGGGATGGAGGACGCCGCGACGCTGCT  
CGGATCGCCAAGCCGAGTGGACCGGCTCGCAGAACTCGATCGTGCCGAAATGGTGAACGCGACGAACACCGCG  
GTCAGCGCAAGGTGGACGGAATCGCCATCGCCGTGGTCGACAAGGACGCGTTCCGGGGGCGGTCGACAGGCTC  
TCAACGCGGGGATACCGGTGGTGTGCTACAACGCCGACGAGCCGGGGCGCGCGGGCACGAACCGGCTGGCCTA  
TATCGGACAGGGCTGTACGAGTCCGGGTACGCGCTGGGCCAGCGGGCGTTGACGGTGTGACTCCGGCGAGGTG  
GCCGCTTCATCGCCACCCCGGGCGCGCTGAACATCCAGCCGCGCATCGACGGCGCGCAACAGGCGTTCAAGGACT

CCGGCAAGCCGATCACGTTACCGCGGTGCGCCACGAACGCCGACGTGACACGCGGTCTGTCCATCATCGACGCGTA  
CGCCAGGGGACGCCAACCTGGCCGGCATGCTGGCGGTGGACGCCGATCGACGTCGTCGGTGGGTGAGACCGTC  
AAGAAGTACAACATGCGCGGCAAGGGGCTCAAGGTGGCCGGCGGGTTCGACCTGATCCCGGAGACGCTGACCGGGA  
TCCAGGAGGGCAGCCTCGACTACACCATCGACACGAGCAGCCTTATCTGCAGGGATTCTGCCCCGTGCTGGCGCTGTA  
CTTCTACAAGGTGTCCGGCGGGCTGATCGCGCCGAGTGAGACGAACACCGGGCTGTGTTCGTACCAAGGACAAC  
GTCGCCCCGTACCAGAGCACGAAGAGCCGGTACGAGGGCTCCACGACCGACAAGGTTCTCGTACCTCGCAGCGGGC  
CGATCGCCCATGGATGACCGGATCTCGCCTGCGCCCGCGCAGGCGCCTTCCCTCGAGGTCGAGCAACGCGGTGGCC  
GGTGGCAGCCCGTACCGCGGCGCGCGGAAGGTCTCGACGCGCTTCCCTGCGGCGGCGTGAAGGCCAGCGTGTCTGCT  
CGTCGCGATCGGCTGATGATCTACTTTTGGGCGTCCAGCCCGGTCTTCTGAGCCGCGACAACCTGGTCAACATC  
GCCCAGGCGACCGCGCCCGTTCGCGATCATCGCGGTGCGCATCGTGTGCTGCTGGTCAGCGGCGAGATCGACCTGT  
CCGTGCGTATCGTGGCCGCGTGGCGCCGTTCTGTTCCTTCCACTTCGGGATCAACTTCTACTCGCTGCCGGTGGTGCC  
CGCCTTCGTGGTGGCGTGGCGATCGCGGCCGGAATCGGCCTGGTGAACGGGCTGATCGTCACGACGCTGCACGTG  
CCCTCGTTCTGTGACGACGCTGGGCACGTTCTTCGCGGTGCAGGGCATCCTGCTGATCACGTGCGACGCGCTATCCGG  
TGCCGATCCCCGACGCGGCCAAGGGCACGTTCCAAACCTGGCTCGGCGCCGGGCGGTGGGCCAGCATCACCTGGGC  
GCTGATCATCGTCGCGATCTTCCACACGGTGCTGACCCTGACCCGGTGGGGCTGCACACCATCTCGGTGCGCGGC  
AACCCGGTCGCGGCCACCGAGGCCGGCATCCGCGCCTCCCGCATCAAGATCGGCAACTTCGTGATCACCAGCACGC  
TGGGTGGGCTGGTCGGCATCATGGAGGCGTTCCGGATCAACACCATCGACCCGAACATCGGCGCGGCGACGACGCT  
GACCTTCTACGGCATCTCGGCGCGGTTCATCGCGGACCGCGCTCGCCGGCGGCTCCGGCACCATCGTGGCGCC  
TTCTGCGCGCCCTCGTGTGGCCGAGCTGCAGAACGGGTTCAACCTCATCGGCTACAGCGCCAACACCATCTTCC  
TCATCTGGGCTGGCCATCTCTGTCGATGATCGCCAACAGTACCTCTCCCGACTGCGCCGGGCGAGGCCGATC  
ATGACCGCGGAAACCGTGTCCGACGCGCTGCGCGTACAGAACATCGCCAAGAGATTGCGCGCTCTACCGCGCTGC  
AGGACGTGACCCTGCGCGTGCCTGAAGCGAGGTGCTGGGCCTGATCGGCGACAACGCGCGCGGCAAGTCGACCT  
CATCAAAATCATCTGCGGGTACCACCGGCCGACGCGCGGCGCATCTTCTGTCGCGCGGAGGAGGTACGCTGCGC  
AGCGTCGACCACGCCCGCTCGGTGCGCATCGACGCCGTCTATCAGGACCTGGCCCTGGTCAACGAGCTGTCCGTCT  
ACCACAACATGTTCTCAACCGGGAGCTCGTACGGTGGCCGTGCTGAACAACCGGGCGATGCGCCGCGTGGCGA  
GGAGCACCTGCGCGACATGGGAGTGAACCTGCCGGACGTGCGCGTCGAGGTGGCCAAGCTCTCCGGCGGCGAGCGC  
CAGGCGATCGCGGTGGCCCGTGGCTCTACTCCGACGCGCCGATCCTGCTGCTGGACGAGCCGCTAGCCGCGATGG  
GCGCGAAGGAGGGCACGATGATCTTCGACCTGATCCGCGACCTGAAGGCGCGCGCAACGTGTGATCATCATCAT  
CGCGCAAACTACGCGCAGGTGCTCGACGTGTGCGACCGGGTGAACCTGCTGCAGCACGGCCGGATCACCTTCGAC  
AAGAGGTGCGCGGACACGTGCTGGCCGAACGACCGAGCTGGTGGTGGCCGAGTACCGCACCGGCCGCGGGCGGT  
GACGCTCAGCGCTTGGCGCCGACTCCGCTGTACTGGCTGCGCGCGGTGCTCTGTCGCGGGGTGGGATGCCAC  
TGCGGGCTCGGCACGACGCGCGGGGACAGGTCCAGTCCCTCGAAGAAGCCGGCGATCTCGTCGGGCGTATGCA  
TCAGGTAGGGCACCGCACCGGAGGAGTGTAGGCGTCTGGGCCTGGTTGAGCGCCTCTCCCCGGCGACGCGCAC  
ACCGTCGTTGATCGACAGGTAGCTGCCGGGCGGCGAGCCCGGCCATCAGCTGCCGGACGATGTGCGCGACCTCCGCG  
GTGCTGGGGATGTGGCCGAGCACGCCGTTGAGGATCAGCGCCACCGGCCGGCCCATGTGAGCTCGCGGGCGGCGA  
CCGCCAGGATGTCCGCCGTTTCTGAGGTACCCCTCGAGGTAGACGTGCGACGACCCGGCCAGCATCAGGCC  
GTGCTCCACCGGTACGGGTCTTGTGACGTAGAGGACCCGGCGTCAAGGGCGACCCGCTGCGCGATCTGGTGG  
GTGTTGTGACGCGGTGGGCGAGGCCGCGCGGACGTGCGAGAACGCGCGGCTCGCGGCGAGGTAGCGCA  
CGGCACGCGACAGGTACCGCCGGGACTCGCGGGCGATCTCGTCGATGCCGGGGAACACGCGACGGTACTCGTCGCC  
GGCAGCGCGGTGACCGGGAGGTGTGCGGTGCCCGGAGCCAGTAGTTCCAGATGCGGGCCGACTGGGGCACCGAG  
GCGCCGACTCCATAGTCACGGATCGATAGCCTAGCCCTTGCAGCGCTGGGTGGCGACGCCCTCGACGAAGTAGCG  
CTGGCACAGGAAGAAGCGATGATCATCGGCACCGTGGTGTATCACCAGCGCGGCCAGGACGATCTCCAGCGGGCC  
TCGCCCCCTGGCCGAACGTTGTCGAGGATCGCCTTCATGCGCGCGGCGATGGTGAACAGCGCCGGATCCCGCAGAT  
AGATCAGCGGCTTGAGCAGGTGCGACCGAGCTGGCCCGCAGCTCGAACACGAACGCCACGATCAGCGCGCGCGGCA  
CAGCGGCACCGCGATGCGCCAGAACAGCGCCAGTACCCGGCGCGCTCGACGCGGGCGGCTCGAACAGCTACCGG  
GGCACGCGGAGGAAGAACTGCCGGATCAGGAAGATGTAGAAGGCGTGGCGAACAGGTTGCCGGCCAGAGCGGCA  
CCTGCGTGGCGGCCAGGCCGAGCTCGTTCCAGATCAGGTACGTGGGATCATGGTGACCGCGCGGCGAGCATCAT  
CGTGCCGACGACCGAGGCCAACAACGTTGCGCGCGGGGAAGCGGAAGTAGGCGAAGCCGACGCCACGACCGCG  
CTGGAGATCGTGACCGCGGCGCGCGGCGGAGTGCCACCACGACGCTGTTGAACATCCAGGTGACGACGGGCGCGG  
CGTCCCAGATGGCGGTGTAGTTCCCCGGCGCCACTCGGCGGGGAGCAGCCGGTTGTGGAAGACGTGGGGCGGGG  
CTTGAGGGAGGCGCTGACCAGCCAGACGAACGGGTACACGAAGATCACCAGCGCGGCGGCCAGGGCGGTCCAGAGC  
ACCCAGGGATGCGGGCGCGCGGAGGTGCGCGGTGGCATGGTCTGTCTCAGCCACAGCCACGCGCTCACTCGCTC  
TCGTAGTAGACGAATCGCCGGCTGAGCCGGACCTGCACGACCGTGATGATCAGGATGATCAGGAACAGCAGCCAGG  
CCAGCGCCGAGGCGTACCCCATGTGCAGGAACCTGGAAGGCGTGTGGAACAGGTAGACCACATAGAACAGGGCGGC  
GTCGTTGCCGTACGTCTGCTGGTTGGCACTGCCGTAGAAGGCGGTGTAGACCTCGTGAAGGTCTGCAGCGAGGCG  
ATCGTGTGATGATCAGCGTGAAGAAGAGCGCGCGCTGCATCATCGGCACGGTGACCGCGCGGAACCGCGCCACG  
CCGACGCGCGGTCCATCTCGGCGGCTCGTAGAGGTGCGGGGACGTTCTGACGCGCGCGGCGAGGTAGATGAC  
CGTGTGCGGAGGCTCCAGGCGCGGATCAGCACCGAGCCGGGCTTGATCCACGGGCGGTGACGCGCTCAGTTGCGG  
CCGTGATGCCGACCGTGCCAGCGCCTCGTTCACGATGCCGACCTGGCCGTGTAACAGCAGCAGGAACAGCACGC  
CCACGGCCACCTTGGGGGTGATCGTCGGCAGGTAGAAGATCGTGCGGAAGAACCCGGCGGAGCGGCGCGGACCCG  
GGCCAGCAGCATCGCCAGGGCCAGGGACACGATCATGGTGACCGGCACGTGCAGCGCGGTGTAGATCAGCGTGTG

GCGATGCTCGTACGCACCCGCGGGTCGGCGAGCATCTGCCGGTAGTTCTCGCCGCCGACGAAGTCGGTGCTGGTGA  
GCACGTCGTAGTCGGTGAACGACAGGACGAGGCTGGCCACCATCGGCCCGCCATGAAGACCGTGAAGCCGATGAG  
CCACGGCGACAGGAAACCGAAGGCGGCCAGGGTCTCCCGCCGCCGACGGCGGCACTCATCGTCAGCGTCCGCTGT  
TGGCCTTGTCCAGCGCGCCTGAGCCTGCTGCTGGGCCTCGGCCATGGCCTGGGCGGGTTTCTGCTTGCTTCCAG  
GACCCGGTTGACCGGTTCTGCCAGGCGGCCTTGAATTCCCTACCGGCGGCGAGCGCCGGCTCCGAGAAGCCGGAC  
TCCTGGGTCTGCAGGACGATCTGCACGTTGGCGTCGGGTTTACGACGTCCTGAAGATCACTTCGTGGCCTTCT  
TGTTGCCGGTGTAGACGCCGGCGAAGGGTTTGTCTCCTTCTTGCAGCTCGGCACGGGCCCCGCGGCGGCGAT  
CCAGGCTCTCGCTCGCGTCACTGCTTGTATCCACTTACAGGCCCTGCCAGGGTGCAGCTGTTCCGCGGATGGCC  
CAGGCGTTGCCCGTGATCCAGTCGATCGGCTGCCCGTCGACGCCGCGGAACGCGCCACGGTCACTTGACCTTCG  
GCGAGTTGTGCGCGAGCACGTTGAGGTAGAAGTCTCCATCGGGAAGGCGCCGAGCTGGTTGCTGGCGAAGTGGTT  
CTTGCGCCGAAGAAGTCCACGAGTCGCGGAACGACTTGAAGTTCGACCAGCCGCCCTGCAGCTTGATCAGGCCG  
ACGGCGTATTCGAGGGCCTCGACCACCTTCGGGTTGTTACGCTGGGCGGTGCGACCGTCGTGCTGACAGGGCGG  
CGCCGTTTCGCGCGCGCCAGACGGGCAGGAACCTCGGGCAGCTTCGGGTGCAACCGATCCGCTGGAGCTTGCCGCC  
GCTCATCTGGTGAGCCGGGCGGTGGCCGTCGACAGCGCCTGCCAGTCCCCCGTGTGCAACCCGGCGGGATCCAAG  
TTGACCTCGGCGAACGCGGCGTCTTGAGCATCAGGACGCGGTTGTTGTAGAAGTCCGGCAGGCCGTACAGCTGCC  
CGTTTCAGGGTCGCTCGGTGACCGCGGCCTCGCGGAACGGCTCATGTCGATCTTCTCCGTTTCGACGCGAGTCGCC  
GAGCGGAGTGAGCGCCTTCTTGCCCGCTAGGTGCCGAGCAGCCGCTTCCATGTACACAGGTCCGCGCGGGATC  
CTGAGGTCACCGCGACAGGAACGCCCTGGGCGTCGAAGCTGCCCTCGGACGCTTCGCCAGGCTGGGCGCGATCA  
CGGCGTTGGCCGCGTCGAAGCGGCTCTTGCGCATCTCGTCGCCGCTGCCGAAACCCATCATGGTCAGCGTCACCGC  
GGCGTTGCTGTGCGAGTCCGAATCGGATCCGCCCACGCCACCGCAGCCCGCCGCCAATGTCAGAACCAGGGCACCA  
CCCACCACAAACCGGATCGATCGGATGAACATCGTTGTTCTCCATTTTCACTTGTTCGCTGGGCGAAGCATAAG  
TTCACCTTCGATGCCCGCGCACTCGGAGCGGCTGCTGTCAACCGCCGACCGGTTGGCGGATCACCAGATCGGGTGG  
TGCGGCCGCTGCCGGGCGCGCTCCTGCTCCCGCCGCCACGCGGGTGGGCGAGATACCGATCGGCGTCGCTCTAATC  
GATGGCAGGGATACCGACGGTGACACCGTTTCTCCCGCGCTTCATCCCGACGTAATCGCGGGCATGGCGCCGTG  
GCCGCGTTTCTGTCGAAATCGCCGGGCTATTCGCCCGGGAAGTCCACCGAAAGGAAGACACACCATGTCTGCTCTCG  
CCATCGAGAAGTCTGGAAGGACGTCGACCTCCGTGACGGGGCGACAGCCACCGGCCGGCTCGGCTTCGGCGA  
GCTCACCTTCGAGGACCTTCGCGAGGACCGCACCATCTACGCCGCCAGCAGCGGCTGGGTGTGACACCTGACGATC  
GAGTGCGGCACCGTGATCTGCGCCTGCTGATCGACGATCGTTTCCGGCCGGGGCGGCGATCGCCCCCGGGGCG  
ATCCATTTTCGCGCGGGGAGCAGCCATGTCACCGGTTCTTCACTCAATTCCACCTCGGTACGCGACAGCGGATC  
TGACGAACGAACGGTGACCGGAGAAGACCAGCCGGCACCGGCCGCGCAGGCGCGGATAGCGTCTGGCGGATTC  
GGCGTTCTTCGACGACCGGTTCTCGACATCCGGCTACGTCAATGGGGAATCGACCGTGCACCTTCGGCCGGCTG  
CTCACCGACGACGACTTCACCGTTCCCGGCCGGCTGCTCGCCTGGGCGGACGAGCTGGCCACGCTGCTGGCCACCG  
ACACGACACCGGTCACCGGACTCGAGTTGTCCACCAAACTGTGGTTCACAGGGATTCGACCGGCTGCTTTCGCCGG  
CCTGCTTACCCGTTCTTCGCCCCACTACGAACAGCGGCTGCACGAGCGCGTGCCCCGGCCGATCGCCGGCAGCCTG  
CGCCGGCCGCTGCTGGAGTCGCTGGCCAACCGGCTGCTCGCCGTGCGGCGACGACCCCTGCTGCTGGAGCTCAACG  
TGGCCCGTGTGCACGGGCGGCTGACCGGCGACACCCCGCAGCAGCGCTACGACGACTACGACCGGCGGCTGCTCAC  
CGACCCCGCTTACCTGGCCGCCCTCTTCGAGGAATACCCGGTGCTCGGCCGTTGCCGTGCTGCAATGCGGCGCGGT  
TGGGTGGACCAACCGCCGAGCTGTTCAACCGGCTCCACGACGACGAGCCGAACTGCGCGCGGCGGCTGCTGCTGC  
CGCCGTGCGCGGAAGCGCTGCGCAGCGTACGGCTGGACCTCGGGGACCCGCACAACGGCGGCGGCTCGGTGGTGCA  
GCTGACCTTCGACGACGGCACCGGCTGCTTACAAGCCGCGGCGGTCGGATCCGAACGCGCCTACGCCGAGACG  
ATGGCCGCGCTGGCCCGCCACGGGCTGCCGGTGCCGGTGACCGCCCCGCGCGTGTGGACCGTGGCGGGCACGGCT  
GGTGCGAGTTCTGTCGGCCCGCGCCCTGCGCCGACGCCCGGAGCTGTCCCGCTTCTACCGGCGCGCGGATCGGT  
GCTGGCGGCCATGCTGCTGCTCGGTGGCGTCGACATGACATGGAGAACGTTCATCGCCGCGGGGTCGTGTTACCC  
CCGATCGACCTGGAGACCGTGCTGCACTCCGGAGAGCTCGGCGACGGCGCCACCGACGCGTACGGGCGGGCCCTCG  
ACCTGCTCAACCGCAGCGTGCTGGCGATCGGCATCCTGCCCGCCCGCGCGTTCCGTGGCCGGCAGCGAAAGAGTGT  
CGACGTCAGCGCCCTCGGCGGCGGCGAACCAGACCGCGCCCCCGGCGGTCGCCAGCATGCTCGACGCGTACACG  
GACACCGCAGGGCTGGAAGCGGTGAGGCCACCATGGCGGCCCGCCAGAACCGGCCGAGCCTGCCCGGCGCGGAGG  
TCCGCGCGTGGAGACGACCGCTGACGTGGTTCGCGGCTTACCGACGCTACGACATCATGCTGGCCCCACCGCGC  
CGACTTCGACCGGCTGCTGCGCGGCTTCCACGACGTGGAGGTGCGCTACCTGCCCGGCCCCACCGCTCGTTACAGC  
ATCTTCTGACCGAGAGCTACACCCCGACTACCTGCGCGACGCCAGCGACCGGGACCGGCTGCTCGACAAGCTGT  
GGACCGCGCGGACGCCCGCCCCGAGCTGATTCGATCATCGAGTCGGAGAAGCGACAACCTGCTCGCCGCTGACAT  
CCCGTGCTTCCGACGCTCGCGGGAAGCCGGCAGATCCGCAACCGCTCCGGCCGCTGCACCCGGAGTTCTTCACC  
GCGCGGCGCTACCGTACTGACCCGCGGCTCGGCGAGTTCGGACCGGTGACCGCGCCGCGCCAGGTACGCATCA  
TCCGCGACTCGATGGCCACGATGCCCGGCCCGCGGCCCGCCAGCCGTCCCCGACCGGGCGGCGGGCCCCCG  
GCCCCGCTACCGGCGCCGACCGGCCACGCTCGCCGACCGCATCGCCCGCAGGCTCGCCGACGAGGCGATCCTC  
GGCGACCGCGACGTGTCTGGATCGGCGTCAGCATCGAAGCGCTGCCCCAGGAGACCTACAGCTACAAGCCGATGG  
CGACCGGCTGTACGACGGCGTGGCGGCTGCGGCTCACCTTCGCTACGCGGCCGACCCCTCGGCGACGACCG  
CTACCTGGACATCCGACACCGAGCGGCCCGCGCTCGCGGCTACCTGCGCTACCTCGCCGAGCACCGCATCGTC  
GAGACCGTGGCGCCTACAGCGGCACCGCGGGCTGCTTACGCCCTGGACACGTGGCCACCGCACCGGCGACG  
ACTCTACCTCGACGCGGTCTCGGAGGCGGTGCCGTGGCTGCGCGAATGCGCCACCCGCGAGGAGTGGCCCGACCT  
GATCGCCGGGCTGGCCGGCTGCGCCCTGATCAGCCTGGACCTGCACGGGCGCCACCGGATCGACGGGCTGCGCGAG

GTGGCGGCCATCTGCGCCGAACGGCTCGCCGCCCTCGCCGTCGACGTCGACGGCGCCGCCGGATGGCCGGCCACTC  
CGGACGGACCGCTGCTCGCGGGTTCTCCACGGCGCGGCCGGCATCGCCTGGCCGCTGCACCGGCTCGCCACCGA  
ACTCGGCGACCCCTCGCTGCGCGAGCTCGCCCGCCGCGCGGTCCAGTTCGACCGCGACCTGTACGTGCCCGCCGCG  
GGCGCCTGGCGTGACCTGCGCCCCGAGATGGCCGGCACCGACTCCTACCCCGCGCTGTGGTGCCACGGGGCCGCCG  
GCATCGGCCTGTCCCGGTGCTCATCGCGCAGCAGCAGGACACGCGACTGGCCGCGGAGGCCACCGCCGCCCT  
CGACCTGGTCGGCGCGCACGGCTTCGGCCACAACCACAGCATCTGCCACGGCGACTTCGGCGCCCTCGCCCTGTTT  
GACCTGGCGGCGACGAACCGGCTTCGACCCCGGGCGCCACGACGCGGCCCGCGGCCGTGACCGCGACATCGCCG  
CCAACGGAGCCCGTTGTGGCCTGGTCGGCGACATCCACATGCCGGGACTCATGCTCGGCGCCCGCCGGCATCTGCCT  
GAGCCTGCTGCGCATCGCCCAACCCCGCGCAGGTGCCCGCGGTGGCCTGGCTGCAGCCGCCGCTGACCTGAGGAGGA  
GCCCATGCCGGAAGAGATCGTGCTCAGCGTGCTGGACCAGGTGCCGGTGTTCCTGACGGCAGCCCGCGGAGGCG  
GTGCGCGACGCGGTGCGCTGGCCCGCAGCGCCGAACAGTACGGCTACCACCGCTTCTGGATCGCCGAGCACCACG  
GCAGCGCGGCCAACGCGTGCGCCGCCCGGAGGTGGTGACGGCCGCGGTGGCCGCCGCCACGAGCCGGATCCGGGT  
GGGACGCGGCGCGCTGCTGCTGCCGCACTACAGCCCGTGAAGGTGGCCGAGACGTTCCGGGTGCTGGCGGCGCTG  
TATCCGGGCGCATCGACCTCGGTTTCGGCCGGGCTCCTGGTGGGCGCCCGCGCATGGCCGAGCTGCTCAACCCGT  
ACGCGGTCCGCACCGACGAGGCGTTCCTGGAGCAGATCGGCCGGTGTGGGGTTCTCGGCGACACCCGTACGGT  
CAGCCGCGTGTGGTGACGCCGAGGTGGAAGAGCCCCCGGTGCCGTGGATGCTCGGCGCCGGGACCGGCGAGCGCC  
CAGATGGCGGCGATGCTGGGGTTGCCGTTCTGTTTCGCCAGTTTATCGCGACCGAGGAGTGCCCGGAGGCGCTCG  
AGGCGTACCGGGATGCGTTCCGGCGCTGCCCTGGCTGGAGCGACCCAGCCGATGCTCGCCTTGCGGGTGCTGTG  
CGCCGACTCGGACGCGGAGGCCGAGGAAGTGGCCACGTGTTTCTGGATGTGCTGCACGACCGGCTGGCGTGCGCAG  
GTGCAGCTCACCGACGACTACCGTGGGGGCGCGCCCAATCTCGACGACGACGCGCGGTACCGGCTGACCGCGGAGG  
ACCTCGCGCTGCGGGAGAGCCGCGCGTTCCTGCAGATCTCCGGGACGCGCGCGCGGTGGGCAAGGAGATCCGCAG  
GTTGCAGGCGGTGTACGGGTATCCGAGGTGGTGCTCACCACGAAGTGGCCCGCCTGCCGCGCGCGCGCGCTCC  
TACGAAGTGTGCGCGGCGAGTTCGCATCCCCGCGCGCCTGACGGCCTCAGGTCCGCGGTAGCTCCGCGTAGGTCC  
CGCGCAATGCCGCGGACCTGACGGCCGTGAGGACAGGCGAGCGTCTGCTGGACAGACCCGACGCATACGCCAGG  
ACGGCGGCCTGCACGCGGTGGCCACGCGGAGCTTGCTCAGGATCACGCTGACGTGCTCCTTACCGTGGCGGCGC  
TGAGGAACAGCTGCCGGTGTATCTCCGCGTTCGTGAGGCGCGCGCGGAGCAGGCGCAGAATGCTCTGCTCCCGGT  
GGACAGCTGTTTGACCCGCTCGCAGGCCGCTGACCCGGCCCGCGCGCAGATCCCGGTGAGCCCGCACGACGAC  
ACCGACGACGCTCGGGCGCGCATGCTGCCCTCGGCCAGCGCCCGCACGGCGGCGACGAGCTGTTTCGGGCT  
GGCTGTGCGCAACAGGAAGCGCAGGCCCCACCCCGCAGCGAGTCGAGCACCAGCTCGGGTGGGGCGAGAGTGGT  
CAGCATCGCGATCGCCGGCGGGTGTCTCAGAGCCCGCAGCTGATCGAGCACCTCCAGCCCGTCTGCTTGGGCACTG  
TGGCCGTGAGCAGCACCACGGCAGGCCGGTGTGCTCGACGGCCGATCTCAGGTTCTCCCGGTCCGTCTGGCGA  
CCGCGAAGCCGCGGTGCTCTCAGGATCATCTTGATGCCGATGCTGACCAGTGCTTCGCCGTCCAGACAGTAC  
GTCCGTACGATCGGTCCCCGCGAGCCGCCGAACGCATGCAACTTGATCCGCATTCTGTGACATTACTTCACTTA  
TATGGTTCGAATCAATCGATTTTTCGCGTATAGTCATTAAACAACACACAGCGCGTCTGTCAGCGGGTCAGACCTCGG  
CCGGGGCGACCCGCGCCCGCCGCGCGCGCGGCCAGACTGGCCACCACCCCGGTGACACCGCCACCACAGCAC  
GGCGGCCAGCTGGCCCCACGGCAGGCGGATCACCGGATCGGCGGTACGCCCCACGGCCGCCGCCACCGCGGCCAGC  
CCCACCGGACACCGACAGGAGGCCGCGCACCGGTGCGGAGCAGCTGATCAGACCGCCTCGACCGCGCAGGATCG  
CCGCGCGCGCGCCCGCGGTGTGCCAGCGCCCGCAACAGCGCCATCTCGCGTACCCGCTCGACGACCGACAGGCC  
CAGCAGATTGCGGATGCCGAGCAGTGGCATCACCAACCGTACCGCCAGCATCGCCAGCGACGTCCCAACAGGATC  
GACAGCACGTTTCGCGATGTACCGCCCTCGGTGACCCCGCCACCGACCTGCACAGCGCGTCCGGGCGGCCACCG  
CGCCGACGGCGGCCGACAGCGCCTCCCGGTGCAAGCCGGGGTCAGCGACGCCCCACACCGTTGTGCGCACCGCCTC  
GATCCGTTTCGCCGCCAGGTTCCGCGCGTCAACACGCCCAGCAGCTGCCCGGTGGTGTCCGCCAGCCGCGACGCC  
CGGGCGATCAGGTTACCCGGCGCTCGCCGACCTCGACGACGATCGGGGCGCTGTCCGGCAGGCCAGCTCGGCCA  
GATAGGTGCGCGCACACAGCACACCGGCTCCCCGGCGAGCTCCGGTGCCACCCGGGCGGCCAGTTCGTGCGCGGG  
CGCGGCCAGCAGCGCGCGCTCGCCTTGCCGCGGTCCGGGAAACGGGCGCGACCGTGCCGACCGTGCCACTGGCC  
GACAGCTGTGGCACCGCGCGCAAGGCCCGCGGTGGCACCGCTGATCGGCTCGCGTGGTGGCAGGCCCGCGCG  
CCACCGGATAGCGCGCTCCAGGTGCGCGTTACCGTGGCCCGCCGCTGGTTCGCCGCCACCGCCAGGCGAGGTGAT  
CAGCGCGCGCGCGACACACCGCCATCGCCCGGAAGCGTCCGGCGCGCGTCTTGGCTCAGGCTGGTCCCGGCC  
AGGCCGGCGGCCACGCCGAACGCTCCAGCAGCGCGCGCGCGCGCGCGCGGCGCGAGGACAACCGCATCAGCCGCGGCAACG  
CCGCCAGCAACCCCGCGCGCAGCAGAAGCCCGCCAGCAGCGCCAGGGGCGAGCAGCGCGCGATCGCCGCCACCGC  
GAGCGCGCGCGCGCCACACCGGTACACCGGTGCCACGACGAGCCGGCGACCGCGCGGACCGTGCAGCGCGCGC  
GGCTCGTCGGCCGCTGCAACGCGCGCACCGGGCGATCCTGGTGGCACGCGGGCGCGCGCCAGGCGGCCACCA  
GCGTGGCCAGCACCCCGGCCAGCACACAGCCGGCCAGGGCGAACGGATTGACCCGCGAGCCCGCGCGCGCTGATGTC  
GAGCAGGTGCGCGCGGAGATAACCCAGCCCCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTTCGGCGATC  
GCCGCTCGGCCAGCACGACCGGCTCACCTGGGCACGATGACCGCGGACAGCCGCGAGCGGCGATCTGCCGGA  
TCCGCTGGACGATCACACGTTGAAGGTGTTTCGCGATGACAGCACCGCGCGGAGCGAGCGGCGAGCGCGCAACGC  
CAGCATGATCAGACAGCTGGCCGTGCGCGCGCGCAACCTCGCGCGCGCGCTGATCCGAGCGCGCGCGCGCTCG  
GTGCGCGAGATGCGCGGGGCCATGCTGCGCGCAACGCGTCAACGGTGGCGGTGAGCGAGGCGTGTGCGCGACAG  
TCAGCAGCGCGCGCGCGCGCACGTACCGGCGAAGAACGACGCGTGGCGTAGAACCGGAAGTCCGAGCCGCTCAG  
CGGCCGGAACACAGGTGGCGGCGACCGGTACCGGTGACCGGCTGCGGCGCGCGCTACCGTGGCGTACCGTCAGC  
GTGGCCCCGACGTCGATGCCGAGGTGCTGAGGGTGCCTGGTGGCGACGAGCTGTCCGGGCCGGTGGGCCACG

CGCCCCGGTCCAAGGTGAACCAGCGGACCTGCGGGGTGGCCGCGATGCTCTGCACGTTGGCCGAGCCGCGCCGCGA  
TCCGCCGAACACGCTGACCGTGCGCGCTACTGCGCGTCCACCGAGCGCACCCCCGGCACCGCGCCGCGGCGCTCG  
TACCAGGCCGGGTGCGCGACCGTGTCTGTCGCGCTCGAGCACGATGTGCGCGGTGGTCAACGGCGCGCGCGCGGTAC  
GCCGAGGCCCTCACCCGAGGTGGCCGCGAAGGTGCGGGTTCGCGGCCAGGAAACCGGTGCGCCAGCATCACCGCCGC  
GACGATCGCGAACAGCCGGGCGGGTAGGCGCGCAGCTGGGACCAGGCAAGCGCGAAGATCATCGTCAACCGCCA  
CCGACGCCATCGCCGCGAGGATCTGATCGCGGCCGGGACTGCGCAGTTCTTCACGGATCCGGCCGTGCGCCATCAC  
CAGGACCCGGTCCGCGTACGTGCGCGGCCCGGATCGTGGGTGACCATGATGATCGTCTGGCCGAGCGCGCTGCC  
GCGTTCGCGCAGCCCCGCGCAGCAGCGCGCGCGCGGTGGCGATGTCCAGCGCGCGCGGTGCGGTCTGTCGGCGAAGACCA  
CCGACGGCTTGGCCAGCAGCGCCCGCGCCACCGCCACCCGCTGCTGCTGGCCGCGCGGACAGCTCGGACGGACGATG  
CCCCAGCCGGTTCGGTGATCTGCAGTGAATCGGTGATCTTCCGCGAGCTCGCCCGGATCCACCCGCGCGCCGCGGAGC  
CGTAGCGGCAGCACGATGTTCTGCTCCGCCGTACGCTCGGCAGCAGGTTGAACGCCTGGAAGACGAAGCCGATCC  
GGTTCGCGCGCAGGTTCGGTCAGGGCCCGGTCTGTCAGACCGGCCAATGACCGCCCGGCCAGGCTGACCGTGCCCGC  
CGTTCGGTGTGTCCAATCCGGCCAGCAGGTGCATCAGCGTCTGTTCCCGGAACCGGACGGGCCCATGATCGCGGTG  
AACTGGCCCGCGGAGAAGCCCGCGGACACCCGTCGACGGCTGTACGGCAGCCGGCCCTGTTCCGTACGTCTTGC  
TGACGTCCCGGCAACTGACCATCTCGGCGTGTGTCGTCTGCGTTGTACCGCACCCACGCTAGAAATCCGCGCGGA  
CCGGCAGTATCCCAACAGGAGTCCCGGTTCGGCGGTGGACCGGATACCTTCGTATGACCGCCGGTATCGACCTCGG  
GCGCGGTGTTCCCGCAATCGATGATGCTGCTTAGCTGGGACCGTGTTGCGAGGGAACCGGTGCTCAGGATCGA  
CGCCCTCGTCGCGCGCGCGGTGGTTCATCGGCTGTCTGCTCTCGGGCTCGCCGGGCTGTCCGAGTGGTACTGGTTCG  
GCCGCGGTGGCCGTACCGTTGCTTCTCCGGCGCAGCGCTCCGCGCTGCTTCTGGCGCTGGTTCGCGCGGTCTCCG  
GCCTGCACCTGCTGGCCTCGCACAGCTTCATGTTCCCGCGCATCTGGTTCGCTCTGGTTCGCGGTGCACGCCGCGC  
GGCGCACGCGCCGGGCCGTGCCCGCCACGCCGACTGCTGCTCGGCGCCGCGGAGCTCTCGTGGTGGCCGCCCAG  
GCCCTGCAGGACCAGCGCCTCGGCTCGGCGCTGCCCGCGGTGCTGATCGTTCGCGAGCACGATGGCCGCTTGGTTCGA  
TCGGGCTGATGCAGCGCCAGCAGCGCAGCGCGGTGCTCGACGCCGAGCACCGCCCGCGGCTGGCCGAACAGGACAG  
CGCCATGCGCGCGCAGCTGGCCGTGCACGAGGAGCGCACCCGGATCAGCCAGGAGATGCACGACATCATCGCCAC  
TCGCTGGCCTCGATCATCGCCCAGGCCGAAGCGCGCCGGGTTCGCCGCCCGCGCCGACGCGCGGATCGCCGGGCCG  
TGTTTCGACCGGATCGCGGGCCTCGGCCGTCAAGCCCTCACCAGCTGAAACGGCTGCTACCGTCTGTGGACACGA  
CGACGAATGGCACGACGACGAGCTGGAGCGGCTGCCGCTGCTGCTCGCCGGAGTACCGAGGCCGGCTGGACGTG  
ACCGTGGACAGCAGCGGGGCGCGCAGCGCTCGCCGCGCGGATGGACCTCGCCGTGTACCGGGTGTACAGGAGT  
CGCTGACCAACGTGCTCAAGCACGCGCGCGCGCGCGGGCCTGCCTGCGGATGCGGTGGACGCCCGCGCTGCTCAC  
GGTCACGGTGAGCAGCCCGCTTCCCGGTGGCCGCGCGCGCGGCTGGTTCGAGGGGCGCGGCTGTCCGGGATCCGG  
CAGCGCTGCTCACTGTTCAACGGCGACTGCACCGTACCGCGACACGGAACCTACCGTACCAACACCTGGCCCC  
TCACCCCGGAAGGAGCGCGCGCATGACGCGGCCACCGATCGCCGTGCTGATCGCCGACGATCAGGAGTGGTACGC  
ACCGGTTTCGCGATGGTTCGTGACGCGCGCGCGGACATGCGGGTTCGTGGCCATCGCCGCGAGCGGCGCGGAGGCGA  
TCGAGCTGGCCGCGCAACACCGGCCGAGCTCATCCTGATGGACATCCGATGCGGGGACCGACGGGATCACCGC  
GACCAGCGCATCTTGGCCGCGCGCGCGGAGCGGCCACCGAAGATCATCGCGCTGACGACGTACGACAGCAGCGAC  
TAGCGCAGCGGATCTTACCGCGCGGGGCGAGCGGCTATCTGCTGAAGGACGCGACCGCGGAGGCTGACGGCGG  
CGATCCGCGTGGTCAACAACCGCAACGGCGGGCAGAAACAGGGCGCGTTCGACGATCACCATGCAGTGGGTACGGATCT  
CGCTGGCCTACTCGGCGACCAAGCCGAGGACGTTCATCGACGCCACCGAGGACGCCCCGAAGCGCAAGGTTCGCCGA  
GATGAAGTACGCGCTCGAGGTGGAGAAGCAGCTCAGCAAGGACAGATCCTGGAGCGGTACCTGAACATCGTGCCG  
TTCGGCAAGCAGACGTACGGCATCTACGCGGCCAGCGGGTCTACTTCAACAAGAAGCCCAAGGACCTACGATCG  
GCGAGGCCGCGCTGCTGGCCGCCATCGTGAAGGCGCGTCCGCGTACGACCCACCGACCGGACGGTTACGAGCT  
CATCCGGCAGCGGCGCAACGCTATGTATCCCGGCGATGGTGGAGATGGGCGCCATCACCCGGGCGCAGGCCGAC  
GCGGCGCTCAAGGAGGCCATCCCGCGCAAGGTACGCCGATGAGCAACGGCTGCGTGTGCGTGGCCAGAACT  
GGGGCTTCTTCTGCGACTACTTCTACCGCTGGTGGATGGAGCGCAAGGAGTTCGGGCCACGCCGTACGACCGGGA  
GCGCCGGCTGAAGAGCGGCGGCTACCGGATCACCAGCAGCTCAGCGTCAAGGCGCAGAAGCAGGCCCGGGATCGG  
ATCGCGACCTGATCTCCGAGAAGAACAAGACGCGTGTGCTGGCAGCGGTCGAGCCCGGACCGGCAAGGTAC  
GCATGCTCGCCGCCAACCGGTAAGCTGAGCAGTCCGGATGATCCGCAAGACGCGATCTCTCCGACCCGAG  
AAAGGCGCGCAAGGGCATCCGTGGCTCGTACCCGAACACGACGAATCCCTGCTGACCGGCGCGCGGACATCAC  
GGCTACAGGCGCGCTCGGTGATGAAGATGTTACCATCGTTCGCGCGCTGGAGCAGGGCTACCCGCTCGCTACA  
CGATCAGGACGAGAGCCGTACCGCTCCCGCTACATCATCGAGAGCAGCAACGACGCGGCTGCCCGGAACGCA  
CTTCTGGTGTCCAGCAACCGCGGTGGCGCGCGGAGGTGCTTCAACATGTGGACCGGCTGGGCGAGTTCGATC

AACACGTACTTCGTGCCGCTGGAGGAACGCGTCGGCGCGGAGAAGGTGGTCAGCGCCGCGAAACGCTTCGGCATCC  
AGTTCGGGGAGCCGGACGACGCACTGCTGGCCGAACCGGGCAACGCACACCAAGTGGGGCGCCTTCACCCTGGGCGT  
CTCGGCCACCACCCGCTGGACATGGCCAACGCCTACGCCACCCTGGCCGCCGACGGGATGTACTGCCCGCCGACC  
CCGATCGAGCGGATCGCCACCCGTGACGGCGACCAAGCTGGACGTCGGCCGGTCCCCGTGCGTACGGGCGACCGCCA  
AGGACGTCGCCCGCGCCGCCCTCGACGCGGCCCGCTGCCCGGTGGGCGACTCCGCGCAGCTCGGCCGGTGGGGGG  
AAGCACCGCCGGCATCACCCGGTCCGTGGTGGTGGGCGATCCGGTGTTCGGCAAGACCGGCACCAACCGACCGCGACCGG  
ACCGCTTCGTGATCGCCGGCACCAACCGCGCTGGTGGTGGCGGTTACCTGGTCAACCCCGACTACCAGAACCACC  
GCGACCGGCTCGACACGACCAAGGTCAACCCGGCCGTCTACCGCACGCTCGCCGACTACATGGAGGGCGAGGCCACG  
GGAGTCGTTCAAGCGGCCGTGAGCGGGCGGATCGCGTTCGGTGACCAAGCGCTCGATCCCGGACGTCGAGTGCAGC  
CCGATGCCCGCGCCCGCGACCGTCTGGAGGATGCCGGCTTCGACGCTTGCGGAGGCCAGGAAGTGGAGTCCGACT  
GTCCCGCGGGCACCGCGCCGGCACCGAGCCGAGCGGCCGGACCGTCAAGAACGGCGTGGTGGTATCCAGGTGAG  
CAAGGGCGCCCGGGGCGGTACCGCCGATCTTCCCGCCGATCGGGCCACCCCGCTGACCGGCTGACGGCGTAGAC  
GTCGCCGGCGTTCGCTCGTGGCGAGCCCGGCCAGGTGCGCGAGCATCTCGTGGCGGGCGGGCCACCGGCCGAAG  
GTGCGGGCCATGGTCTCGCCAGGGAGACGTGGAAGCCGTGGAAGCCAGCTCGCCGGCACGGTTCGAGGCGAGCGCC  
GGGCCACGTCCCGGGCGATCGTCTGAACTCGAAGCACAACGCCGGACACCGCGGCTCAGACCGGCGAGCACCGC  
GTCCTCGTACCCTTCGACGTCGATCTTGATGAAGGCCGGCAGGCCGAATCGCTGGACGAGCGTGTCCAGGGTGACC  
GCAGGGACGGTGTCTGCTGGTCCCAGACCTGTTCTTCCAGCCGCGGAGCCGGTTCGCCGCGGTTCAGGAAGCGCA  
CCGAGTTCGTGGAGACCGTTCGGGTTCGCCGAGTGTAGTAGAGCGGCACGCTCCCGCGGGCGGGCCGACCGCGC  
CTCGACGAGCGCCACCCGGTTCGTGGTGGGCGTAGAGCGCGCGAGCGCCCGCATGCACAGCGGCTGCGGTTCGACG  
GCGACACGCGGGCGCCGAGCCGCCGAAGCAGGCGACCCGGTACCGACGTCGCGCGCGATGTGGAAGACGACGT  
CACCCGGCCGGACGAAACGCCGTACATCGCGTCCATCGCGGCTCCCGGCCGGCTACCGTAGTAGTAACGCAG  
CGACCGGGTCAGCGGCGCCATGGCCGGTTCGACTCTCAGTTTGGCGAGGTATCGGTACGGCGGGGGCATTTCCC  
CGATCAGGCGATCCATGCATCGCCGGCGGTGCGCAGGTAGGCGGCGATCGCCCGGGCGGAACGCACGCCCCGAGGT  
GGTCATGACGTTGTGCGCGCGCGCCGGGACGGTACCGACACGCGCGCGGTATCGCCGCCACCTCGGCACGCCAG  
CGCAGCGGCGCCACCGGATCAAGCGACCCGCGGAGCACAGCGGGGTACGGGCAACCGGCGAGATCCTCCTCGA  
TCGCGTTGCGTACCGAGTGGCCGACCGTGGCCAGAACCCGCCACGGCTTGGCGTCGGCGATGTGCGGACCGAC  
CGGCGCCTGCCACGGCGCCTCGATCCACAGGTCCACCGCCACCGGCCGATCAGCGCCCGCGGGACCGGGCCCGC  
GGTTCGGTGGTTCGGGCTGGCCAGCACGCGCGCGGACAGGTCCGGCCGAGCACCGCGCGAGCACCGCGGGCCACCT  
CGGCGCGAACGAGTGCCTGGCGATGCAGGCCCGGGGACGCGCGAGCACGTCCAGCCAGGCCGCGAGGTGTTCCGG  
GTGCCGGCCACGTCGTACGCGCGGCGCGGCTTGTGCTGAACCCGAACCCGGGAAGGTCCGGAACGAGGACCGGA  
TGGCGGCGCCGAGCGGTGGCGGTGGGCATCAGGTAGCGGTGCGAGACGGCGAGCCCGTGTACCAGCACCGG  
GAAGGCCGCGCGGTGGCACGCGTGGCGGTGCTGCGTACGCACTCCACCGACCGTACCAGGCGGTGTGCAATCG  
GGGCACCGCGGAGATGTACCCCGTCTTCCCAATCTTTCCCGCGCGGTTAGTTTCGAGTTGTGTCGCCGCCCTTC  
CGGCTCGACGAGGCGGTGCGCGACGTCTACGGCGACCTGCGCCTGGCGCCCGTGTGTCGGCGGTGCTGCGGCACA  
GCGGCGCCTGACCGGCTCGGTGGCCGGCAGCGTGTGCTGATCGACGCGGACCGCGGCTGTACGTCAAGGCCGC  
CGAGTACGGCGCGAAGTGCAGCTCGGACGCTTTTCCCGCTCGACGAGGGCGCCACCGGCCGCGCCTTCGGCAGC  
CGCGCGCCGTGGTGTATCCCGACTACGGTACGTCGCGGCGGTTCATCTCGCGCGGCGCATCCGCGCCGGAAGG  
GCCCCGCGTGTGTCGCCATCTGGTGGCGCGGCGAGTATCGCGGTGAACGTGCGCTTCGCGCGGCGCTTCTC  
CCTGGGTGGTGTGACGAAGTGGAGGCGGTGACCCAGTCCGCGGCGCGCGATCGTCCGCGAGCGGGGTGTGCGG  
GTCCGCGCGGACCGCGGTACGCGGCTCCGGCCGACCGTTACCCCGCGCGAGGCCGAGGTCTCGATCTGCTCC  
GGCAGGGTCTGACCGACCGCGAGATGGCCCGCGGCTCGGCCTGTCCGCGAAGACCGTGGAGAAGCACGTGCGCGC  
GATCAGGCGCAAGACCGGGACCTCCAACCGTACGGCGCGGTGCTGACGGCCCTGGACAACGACTGGGTGGGGAAT  
CTTCCCCATACCGCGGAGCACACCGCGGTCTTGAAGTGGCGGCGATGCCCGTGTCTCCATGAAGGACATCCTCGA  
CCGTGCGCTGGCCGAGCGGTACGGTGTGGCCGCGCTTCAACATCGTCAACGACCTGACCGTTCGAGGCCGTGCTCGCC  
GCCGCGGCGGAGGAACGCGCCCCGCTCATCTGCGAGCTCGGTCAAGACCGTCCGGATGTACGGCCGCCCCCGGC  
TGTACGAGATCGTCCACGCTTCGCCCACGACGCGCCCGTCCCGGTGACCTGCACCTGGACACTGCCCCGAGCG  
GTCGGTTCATCTCCGACTGCTCGCCGGCGGCTGGAACCTCGGTGCTCTTCGACGCGCACGAGCTCGACGTGGCCGAC  
AACCTGCGCCAGACACCGAGGTGGTGGCCGAGGCCGTGCGCGCGGCGCCACGTTCGAGGGCGAGATCGAGGGCA  
TCCAGGGTGTGAGGACGACGTGCGCAACGACTACGCCCCGATGGTGCAGAGCCTGGAGGTGGCGGTGCACTTCAT  
CAAACGCACCGGCGTCACTGTTTCGCGCGCGGCCATCGGCAACGCGCACGGCCAGTACAAGCAGGCGCCCGTGTCTC  
AACACCGCGCGGGTCAGCGACCTCGTTGCGGCCACCGGCATCCCGATGGCCCTGCACGGCGGCACCGGCTCTCCG  
ACGAGCAGTTACCGACCTCATCGCCCGTGGTGGCGGAAGGTCAACATCTCCACGGCGCTCAAGGAGTCGTTTCAT  
GAAATCCGGCCTGGAGTTCTGCGCGAGGCCGATGAGCGCGGCAATGGGATCCGCCGTGCTGTTCGGCGATCAG  
CGGCGGGCGGTGCTAGAGATGGCCCGGCGACATCCGGCTCTTCGGCGGATCGGGGCGCGGTGGTGAACGCCCT  
GGTCTTCGACTGCGACGGCGTGTGGCCGACACCGAACGGCACGGCCACCTGCCCGCGTTCAACGCCACGTTTCGAG  
CAGTTCCGGGTGCCCCGTGCGGTGGAGCGAGGAGGAATACGGCGAGAAGCTGCGCATCGGCGGCGGCAAGGAGCGGA  
TGGCGTTCGCTGATTCGCCGATCCCGCTTCGCGCGGCGGCGGCGGCGACACCGACCGTACGGAACCTG  
GCACCGCGCAAGACCGCGGCTTTACGAAGCTGGTGGCGGAGGGCGGATTTCGGGCCGTCCGGGCGACGCGG  
ATCATACGCGAGGCACTCCGGGCGAGGTGGACGGTGGCGGTGCTTCCAGCTCGGCGGAGGATTTCGGTACGCGCAG  
TGCTCGTCAACGCCGTGGGAGCGACGACTGCCGAGCGGATCCCGGTGTTCCGCGGAGACGTGTCGCCCGGAAGAA  
ACCGGACCGGCGATC

**SEQ ID NO:101 orf1 Hypothetical protein 146aa**

MPRRSTRCS PACRPGAADQPDFGRACRPDCGRADRS GSGRACRPDCGRAD  
 RSGADRP SAAALVSGMPRRTGVGGAPGR LIEARAVLP ELIGAARDRRASC  
 VEQRVRS LTACRACACRR AAAPAPRT PPRPSAAPAGSDAGGRCG

**SEQ ID NO:102 orf2 Hypothetical protein 149aa**

VSVSLLL VPLAMAGAAVQAAAGRMDDGR LICQVQTRMRDVTLLDAALRD  
 TGATVTAAGDTISATWTQSAATFTRGADGIWAAHV TGVDQPGAVELMTTV  
 DTAYGRRVQQAVLERLRQAPEAGLRLESESVGQDASVRLVFEVERERA

**SEQ ID NO:103 orf3 ATPase AAA involved in cell division 518aa**

VERSAETFAQLLKARFVVLHLETYEEQRALHHLAGIAGDADLV RVPRAV  
 WTWSLTAGLVQPNGEARSGAQRATDALRAVQRIDEP AVFVFRDLHPLFAQ  
 SPEVVRLVRDIAQAFRAGRSPTLVLLSPVLDLPVELSKDVTIVDFPLPG  
 QLELRALLDAMVRGNTASGRLRVELDEQSRERFVTAAAGLTMQEAENAYA  
 RAMVNDAVLDLADLEIVHEEKRTVRKSGVLEFMAAGTVLDDVGGLLENLK  
 AWLVKRNKSWLDEAAGYGLPAPRGVLITGVPGCGKSLTAKAVATAWNLP  
 LRFDIGRVFSGLVGSS EHNMR TALRTAEAVAPCVLWVDEIEKGFAGGTGG  
 DSGTGARVFGTFLTWMEKRTPVFVIATANDFDGLPPELLRKGRFDETFE  
 VDLPSRSERVAVVRVHLGRALRHRAAGELRVDAELLTELAGLTEGYSGA  
 EIEQAVIAGLFDASFERRPLRRDDLVRVMSIVPLSVTQAERVDALRGWA  
 RNRAVSATGTDWDLTNR

**SEQ ID NO:104 orf4 Sugar hydrolase 203aa**

VGELDQQRIHPHNNIWGSGAGTQTIWARS GTNNGVVANHPRTSGVKSYPN  
 TGKTLNRTLSSLNSLTSSFNVSVPSG DYSTTYDIWANNHAYEVM IWTNQ  
 QGAVGP IAEQYDANGAVPNVRNLSVGGHTWNVYRGSNGANAVFSFIRTNT  
 NSGTVDILAILNWLRTNGWWGDVTVGEAQFGFEISGTAGQSNFTVNNFSL  
 NYS

**SEQ ID NO:105 orf5 Endoglucanase 219aa**

MRRRTLALSLSAASAALIAGAGVVTALPASAAAGCRVAYTVSSQWPGGFGA  
 NVTITNLGDPLTNWTLVWSYSGGQTVTQAWNTSLTQSGSQVTARNAGYNG  
 SVGTNATVSFGFNGSGAATPAPGTFTLNGTACTGSAGPTSPSSQPPTNGV  
 PSDAVWVDSGQWANWNTNGYILTTTSGARAPAPRPSGRAAAP TGASSRIT  
 RAPAGSSPTPTPERPSTVR

**SEQ ID NO:106 orf6 Cytosine/adenine deaminase 161aa**

MTITETDLAHLRRCDLAREALDDGDEPFGSVLVSADGKVLFE DRNRVRH  
 GDATQHPEFAISRWA AEHLTPRERASATVYTSGEHCPMCSASHGWVRLGR  
 IVYAASSAQLTAWYKEWGIPAGPVAPLPITTVVPGAVVEGPVPAFEAE LR  
 ELHRARFTPAQ

**SEQ ID NO:107 orf7 Unknown 413aa**

VTTPLVAQAGDSTTGTLGLGLVEDAHQIAEAI RGNWSVVDGVLGGVGASLD  
 GLALAIDPLGTLAAWGVAWLIEHVQPLQDALDWLAGDVDEIAAQAATWRN  
 VAAFTDSAQQDYADRLRTEVAGWFGASGDAYRAHASEHLAALKGISTAAG  
 GISSAVEGAGLLVSLVRGIVRDLIAQFVATLAVRLPQWLAAEGLTLGLAT  
 PVVASQVAALVARGVNKIQHFIRALLNSLRRLMPMIDRLGEVLERLRMLT  
 DRLARSSPSTRPEPTPGPATHAGTENASGNKPEGDLEPNEPRPAEADARD  
 STPQAFVDEVVSNPRSVAGHSAQSIADQFNAAGYSAVVEQSTRSGTSGNA  
 IQVRIHGHPDITNIQVHPGGGRHTPEGSPYWKISTNTVGKIWIIPENFRG  
 ADELRGNVVRYDK

**SEQ ID NO:108 orf8 Unknown 217aa**

VGKPCPDLVEVLSREIQAGNQGASASEVVEIFDLELVGIFRGAVTQKLPG  
 IEVLRKHLHLKQGGVQVRSVVVRAEDPAGIVVVDLCARIDNRDVGLAQQ  
 DAYWERRDDTVDRDLQIRADSPCEFD DMVRLGAGVHGDGQRRVRQRFA DV

ANLFGVQRAIINESRIRTKIDPDDVDAQGDERGSFPRGSHPIHFDDLICH  
SAPHSHAVHRRPGNFRG

**SEQ ID NO:109 orf9 Pyruvate oxidase 592aa**

VSDVVVERLTAWDVERVFGYSGDGIDGVIGALRRAGRPTFVQARHEEGAA  
FMAVGHAKEYTGGAGVCLATQGGPAGVHLLNGLYDAKLDSKPVVAIVGQQVS  
TVLGSAYQQEIDLVRLEFGDVCAQFVQAHAHTSEQVPMMLLDRAFRTALATRS  
PTCVVLPDHDVQTAPADPDQAHEHGVLAATSAGLRPARVVPRPEDLREAAEV  
LRSGERVAIMVGQAYGAEREIVELAQRLLGAGVTTSLGKPVLDENLPFH  
TGMVGHLLGTTASAEMLMRHCDTLLLLIGTNDPWTEFYPRPGQARAVQIDVDG  
RRLGVRYPPVEVPLLGDVETLRELLGLLPSRASGAWGARVEEWVQRWRLI  
SAARAAAPAEFVNPQHVRSLSDHLPADAQVAVDVGSVVYWARHLRLPR  
GVPAPHLSSTLASMGCGLPYGLAAKLAAPQRPVVVLADGAMQMAGMAELI  
TVAARWRDWADPRFVVCVLNNRDLAEVSWEQRETEGEPRFVTSQELPDVP  
YARYAELLGLRGVRIITDPSDLTGAWEAALSADRPTLIEAVVDPAPILLPP  
GQPYEKVQAMYAGLAAEKGDQARRAEHLRRERADEGFDDPS

**SEQ ID NO:110 orf10 Hydrolase or acyltransferase 266aa**

VSRDKTYRPPVRPDGIRTTFHGVLGRMRIRCLGEPRPGVPEIVMIQGMTV  
SDYLLPGLGALSATRVHLVELPGSGSGRPPHDLTVEEYARAAADWLCA  
QRLGRIVLAGHSSGTQVAAETALLCPDEVAGVVLGPAIDPVARGGRLRVF  
ARWWIDRRGDPKSLDEVHKPEREQVGRRLFQVLRHLRHDLKPPVGLC  
VPVLVIRGSEDRLGTARWARRLADLAAGVGRYVEVPGTHSFCWRYPQAWS  
APIREFAGWSVSVSGT

**SEQ ID NO:111 orf11 Aldose epimerase 396aa**

MTEADHRPTVLRRLRLCAVLLMVLGVGLVGAPTSHAGGKPTISKEAFGS  
VGGKAVDRYTLTNGRLQVRILTYGGILQTITFPDHRGRRANVTLGFRITLD  
EYVTTKNPAYFGAIIIGRYGNRIADGRFTLDGTTYQLATNNDPNHLHGGVV  
GFDKRVWDATPIRDGDSVGLRLTYTSPHGEENYPGTLRVMTYTVTRQMG  
IRMDYRATTDPRPTIVNLTNHAYWNLGGEETGTIDHLLKLNANRYTPVDA  
TLIPTGAIDAVAGTPMDFRRPTPIGARNRDPFQQLVYGRGYDHNWVLNRE  
DGQFRRLEFAARAVDPDSGRQLTIYTTEPGIQFYGGNFLDGTLYGTSGRA  
YRQGDGFALETQHFDPSPNHANFPSTVLRPGQTYNSTTIYQFGTAD

**SEQ ID NO:112 orf12 ABC sugar transport periplasmic component 378aa**

MPRIHPKVEEAVSTLDLNRTRRRLLSGTGLFSASLAAGALLSACSDQND  
GQNQTEGAGNFPDTPPEWRFTFVNHVTTNPFPTPTQYGMEDAATLLGIAKP  
QWTGSQNSIVAEMVNATNTAVSAKVDGIAIAVVDKDAFRGPVDQALNAGI  
PVVSYNADGARGAPGTNRLAYIGQGLYESGYALGQALQVLDSGEVAEFI  
ATPGALNIQPRIDGAQQAFKDSGKPIFTTAVATNADVTRGLSIIDAYAQG  
HANLAGMLAVDAGSTSSVGQTVKKYNMRGKGLKVAGGFDLIPETLTGIQE  
GSLDYTTIDQPPYLQGFPLVLAIFYKVSGLIAPSETNTGLLFVTKDNVA  
PYQSTKSRYEGSTTDKVLVPRSGPIAHG

**SEQ ID NO:113 orf13 ABC sugar transport permease 352aa**

MDDRISPAPAPAPSLEVEQRRGRWQPVTAAGRKVLDAFLRRREASVLLVA  
IGLMIYFRASSPVFLSRDNLVNIAQATAPVAIIAVGIVLLLVSGEIDLSV  
GIVAALAPFLFHFGINFYSLPVVPAFVVALAIAAGIGLVNGLIVTQLHVP  
SFVTTLGTFFAVQGILLITSHAYPVPIPDAAKGTFTQWLGA GPWASITWA  
LIIVAIFHTVLTLTRWGLHTISVGGNPVGATEAGIRASRIKIGNFVITST  
LGGLVGIMEAFRINTIDPNIGGGTTLTFYAISAAVIGGTALAGGSGTIVG  
AFLGALVLAELQNGFNLI GYSANTIFLIILGLAILVSMIANQYLSRLRRAG  
RS

**SEQ ID NO:114 orf14 ABC transport protein ATP-binding 253aa**

MTAETVSDALRVQNI AKRFGALTALQDVTLRVAEGEVLGLIGDNAGAKST  
LIKIICGYHRPDAGRIFVGGEEVTLRSVDHARSVGIDAVYQDLALVNELS

VYHNMFLNRELVRWFLNNRAMRRRAEEHLRDMGVNLPDVGVEVAKLSGG  
 QRQAIIVARCVYSDARILLLLDEPLAAMGAKEGTMILDLIRDLKARGNVSI  
 IIAHNYAQVLDVCDRVNLLQHGRITFDKRSADTSLAELTELVVAEYRTG  
 RGR

**SEQ ID NO:115**      *orf15 Hypothetical protein. Methyltransferase*      **255aa**

MESGASVPQSARIWNYWLGSTDNLFPVDRAAGDEYRAVFPGIDEIARESRR  
 YLSRAVRYLAGEAGVRQFLDVGAGLPTVDNTHQIAQRVAPDARVLYVDKD  
 PYAVEHGRELLAGSSDYLEGDLQKPADILAVAARELDMGRPVALILNGV  
 LGHIPSTAEVRDIVRQLMAGLPPGSYLSINDGVRVAGEEALNQAQDAYNS  
 SGAVPYLMHTPDEIAGFFEGLDLVPPGVVPSQWHPDPGDETTGASQYSG  
 VGRKR

**SEQ ID NO:116**      *orf16 ABC transport permease*      **280aa**

MPPRTSRGPHPWVLTALAAAIVFVYPFVWLVASLKPDPDVFDRLLP  
 AEWAPGNYTAIWDAAPVLTWMFNSVVVALAAAAAVTISSAVVAFGFAYFR  
 FPGRNVLFALVVGTMMLPGAVTMIPTYLIWNELGLAATQVPLWAGNLFGS  
 AFYIFLIRQFFLGVPRELFEAARVDGAGYWRLEFRIAVPLCRPALIVAFV  
 FELRASWSDLLKPLIYLRDPALFTMPRGMKAILDQFGQAGEARWEIVLAG  
 AVITTVPMIIAFFLCQRYFVEGVATQARKG

**SEQ ID NO:117**      *orf17 ABC transport permease*      **301aa**

MSAAVRRRETAAFGFLSPWLIGFTVFMAGPMVASLVLSFTDYDVLSTSD  
 FVGGENYRQMLADPRVRSIANTLIYTALHVPVTMIVSLALAMLLARVGR  
 RSAGFFRTIFYLPTITPKVAVGVLFLLLFNGQVGIVNEALGTGIDGPNW  
 TVDGPWIKPGLVLIGAWSLGSTVYIYLAALQNVPRDLYEAAEMDGASAWA  
 RFRAVTVPMISGALFFFTLIINTIASLQTFDEVYTAFYGSANQQTGNDAA  
 LFYVVYLFQQAQFLHMGYASALAWLLFLIIVITVVQVRLSRRFVYYES  
 E

**SEQ ID NO:118**      *orf18 ABC transport substrate binding*      **445aa**

MFIRSIRFVVGGALVLTAAAGCGVGSGSDSDSNAAVTLTMMGFGTGDE  
 IAKTRFDAANAVIAPSLAKASEGSFDAQAFLSAVASRTTPDLVYMERRLL  
 GTYAAKKALTPLGDCVEREKIDMSQFREAAVTEATLNGQLYGLPDFYNNR  
 VLMLNDAAFAEVNLDPAFGDTGDWQALSTATARLTRMSGGKLQRIQFDPK  
 LPEFLPVWARANGAALVSDDGRTAQLNNPKVVEALEYAVGLINAQGGWSN  
 FKSFRDSWDFFGAKNQFASNQLGAFPMEDFYLNVLADNSPKVKVTAPFR  
 GVDGQPIDWITGNAWAI PANSAPGQACKWIKTMTASETWIAAARARAEL  
 RKKENKPFAGVYTGKKADEVIFRDVVKPDANVQIVLQTOESGFSEPALA  
 AGEFFKAAWQNAVNRVLEGKQKPAQAMAEAQQAALDKANSGR

**SEQ ID NO:119**      *orf19 actA lanA structural gene*      **64aa**

MSALAIKSWKDVLDLDGATSHPAGLGFELTFEDLREDRTIYAASSGWV  
 CTLTIECGTVICAC

**SEQ ID NO:120**      *orf20 actM lanM modification gene*      **1053aa**

MSPVPSLNSTSVRDSAYLHERTVTGEDQPAPAAQARIASWRDSAFLLDDR  
 LDIRLRQWGIDRATFGRLLTDDDFTVPGRLLADELATVLTATDTPVTG  
 LELSTKLWSQGFDRLLFAGLLHPFLAHYEQRLHERVPRPIAGSLRRPLLE  
 SLANRLLAVAARTLLLELNVARVHGRLTGDTPQQRYDDYDRLLTDPAYL  
 AALFEEYPVLGRCLVECGRRWVDHAAELFNRLHDEPELRAAGLLPPSAE  
 ALRSVRLDLGDPHNGGRSVVQLTFDDGTDLVYKPRPVGSERAYAETMAAL  
 ARHGLPVPVTAPRVLDRGHGWCEFVRPAPCADAELSRFYRRAGSVLAA  
 MLLGGVDMHMENVIAAGSSFTPIDLETVLQSGELGDGATDAYGRALDLL  
 NRSVLAIGILPARAFGGRQKRSVDVSALGGGEPQTAPRPVPRIVDAYTDT  
 ARLEAVEATMAGAQRPSLPGAIEVRPWEHTADVAGFTDAYDIMLAHRAD  
 FDRLLRGFHDVEVRYLPRPTRRYSIFLTESYHPDYLRDASDRDRLDKLW  
 TAADARPELIPIIESEKRQLLAGDIPCFRSVAGSRQIRTASGPLHPEFFT  
 APAVTVLTRRLGEFGPVHRAAQVRIIRDSMATMPGPRPAAQPSPDRAAGP

RPRVTGADPATLADRIARRLADEAILGDRDVSIGVSIIEGVAQETYSYKP  
 MATGLYDGVAGLALTFAYAAARTLGDDRYLDLAHRAARPVAGYLRYLAEHR  
 IVETVGAISGTAGLLYALDHVAHATGDDSYLDVAVSEAVPWLRECATREEC  
 PDLIAGLAGCALISLDLHGRHRIDGLREVAAICAERLAALAVDGDGAAGW  
 PATPDGPLLGGFSGHGAAGIAWPLHRLATELGDPISRELARRAVQFDRDLY  
 VPAAGAWRDLRPEMAGTDSYPALWCHGAAGIGLSRLIIAQHDQDTRLAAE  
 ATAALDLVGAHGFHNHSICHGDFGALALFDLAARTGFDPRHDAAAAAY  
 TADIAANGARCLVGDHMPGLMLGAAGICLSLLRIAHFAQVPAVAWLQP  
 PLT

**SEQ ID NO:121    *orf21 actO monooxygenase*    341aa**

MPEEIVLSVLDQVPVFRDGSAPAEVRDAVALARSAEQYGYHRFWIAEHHG  
 SAANACAAPEVVTAAVAAATSRIRVSGGGVLLPHYSPLKVAETFRVLAAL  
 YPGRIDLGFGRAPGGPPAMAELLNPYAVRTDEAFLEQIGRLLGFLGDTRT  
 VSRVSVTPQVEEPPVPMWLGAGTGSARMAGMLGLPFCFAQFIATEECPEA  
 IEAYRDAFRPSPWLERPQPMALRLVLCADSDAEAEELATCFWMSCTTGWR  
 AQVQLTDDYRGGAPNLDDARRYRLTAEDLALRESRPFLQISGTPAAVGKE  
 IRRLLQAVYGVSEVVLTTNCPGLPARRRSYELLAGEFASPA

**SEQ ID NO:122    *orf22 actR response regulator*    231aa**

MRSAARGGPIVTDVLVVDGEALVSIKIMILESTGGFAVATTDRENLRSA  
 VEQHRPAVVLDDGHSAQSDGLEVLQDLRALSSPPAIAMLTTLAPPELVLD  
 SLRGGACGFLLRDSQPEQLVAAVRALAEGSIVLAPEASSVVVRAGSRGSA  
 AGAGSPACERVKQLSDREQSILRLLAGLITNAEISRLFLSAATVKEHVS  
 VILSKLGVANRVQAAYLAYASGLSSDDVCLS

**SEQ ID NO:123    *orf23 actT ABC transporter associated permease*    812aa**

MIFALAWSQLRAYPARLFAIVAAMLATGFLAATATFAATSSEGLRRTAA  
 APLTTADIVLDADDTVRDPAWYEAAAAPGVRSVDAQYARTVSVFSGSRR  
 GSANVQSIATPQVRWFTLDRGAWPTGPGQLVADQRTLLDLGIDVGATLT  
 VRHGEAAPQPVTVTGAADLGFRPLTGSDFRFYADASFAGDVPPAALLTV  
 ADDASLTGTVDALRRSMGPGISATDASAAADQAAARFAGGNGQLVVMILA  
 FAAVALLAAVLVIANTFHVIVQIRIQIALLLRVGGHRAQVSRVLAEEA  
 IAGTAGGLVGAAAGVGLGYLGADLLDISGGGLRVNPFALAGCVLAGVLAT  
 LVAAWAPARRATRIAPVRALQAADEPPAGTVRGGRLVVGTVVTVVGAAA  
 LVAAGAIGASLPLALLGGLLLAAGLLAALPRILIALSLAPAARLLERFGVAA  
 GLAGTSLSQNARTASAAAMAVVGAALITCLAVAATSGRATVNADLEARY  
 PVAAGLRTDGEPISGATAGAFAAVQLSASGTVGTVAAARFPDGGKATPRL  
 LAAPGDELAARVAPELAGEPVVLVPATYLAELGLPDSAPIVVEVGERRVN  
 LIARASRLADTTGQLLGVVSARTLAANRIEAVPTTVWGVADPGFDREALS  
 AAVGAVAARDALVQVGGGVTEGGDIANVLSILLGLSLAMLAVTVVIALLG  
 IANLLGLSVVERVREMLLRALGTRRARLRAMLAVEAVVITLLGTVAGLV  
 VGVVPVGLAAVAAVGRADPVIRLPWQGLAAVLVAVLTGVVASLAPARR  
 AARVAPAEGLTR

**SEQ ID NO:124    *orf24 Hypothetical protein*    244aa**

MIVWPERRAASRSPSSARPVAMSSAPVGSSAKTTDGLASSARATATRC  
 WPPDSSDGRCPSSVICSESVIFRSSPGSTRRPASRSGSTMFCSAVSVGS  
 RLNAWKTKPIRSRRSVRARSSRPANDRPARLTPAVGVSNPASRCISVV  
 FPEPDGPMIAVNWPAEKPADTPTAVTAAGPVVYVLLTSRQLTISACVVC  
 VVTAPTLEIRADRHIPPRDPGRAVDRLRMTRRYRPRAPCSRNR

**SEQ ID NO:125    *orf25 Response regulator kinase*    367aa**

VLRIDALVAAAVVIGCLLLGLAGLSEWYWSAAVAVPLLLRRSAPRCFLAL  
 VAGVSGLHLLASHSFMFPGDLVALVAVHAAAHAAPGRARHAGLLGAAGA  
 LVVAAQALQDQRLGSALPAVLIVASTMAAWSIGLMQRQQRSAVLDAEHRR  
 RLAEQDSAMRAQLAVHEERTRISQEMHDI IAHSLASIIAQAEGRVAARA  
 DARIAGPVFDRIAGLGRQALTDVKRLLTVVVDHDDHDDGLERLPVLLAG  
 VTEAGLDVTVDSGAPQPLAAGMDLAVYRVIQESLTNVLKHAPARRACL

MRWTPALLTVTVSSPLPGGRGAGLVEGRGLSGIRQRCSLFNGDCTVTATT  
ELTVTTTWPLTPEGARA

**SEQ ID NO:126 orf26 Response regulator sensor 220aa**

MTRPPIAVLIADDQELVRTGFAMVVDAAPDMRVVAIAASGAETELAAEH  
RPDVILMDIRMPGTDGITATSAILAAGGERPPKIIALTYYDSSDYATRIL  
TAGASGYLLKDATAEGLTAAIRSAYHGGSVIAPTTTRNLVAARAEPPPPA  
RDPAPLDTFTARERDVFDLIVAGANNAEIAARLHLAEVTVKTHVGRVLAK  
LGVDRDLNVVVWAYRNGAVG

**SEQ ID NO:127 orf27 Penicillin binding protein 782aa**

MLICGLIAGVVAAAAFFFAAMSGLAAGQQTFAASLPSELKAFRSPQIS  
RIYAADNRTQVAQFYDEFSDVPLKEMSPFMRDAMVAAEDRQFYQHHGVD  
LKGAARALVNNRNGGQKQGASTITMQWVRISLAYSATKPDVIDATEDAP  
KRKVAEMKYALEVEKQLSKDQILERYLNIVPFGKQTYGIYAASRVYFNKK  
PKDLTIGEAALLAIVKAPSAYDPTDPDGYELIRQRNAYVIPGMVEMGA  
ITRAQADAALKEAIPRKVRPMSNGCVSVAKNNWGFCDYFYRWWMERKEF  
GPTPYDRERRLKSGGYRITTTLDVKAQKQARDRIGDLISEKNKNALLAA  
VEPGTGKVRMLAANRRYKLDDPDDFQNAISSDPKARKGIRGSYPNTTNP  
LLTGGGDITGYQAGSVMKMFTIVAALQGGYPLAYTIRTQSRYSRYIES  
SNDAAACPGTHFWCPSNAGGGGEGVFNMWTGLGRSINTYFVPLEERVGAEK  
VVSAAKRFGIQFREDDALLAEPGNAHQWGAFTLGVSATTPLDMANAYAT  
LAADGMYCPPTPIERIATRQDQLDVGRSPCVRATAKDVARAALDAARCP  
VGDSQQLGRCGGSTAGITRSVVGHVPVFGKTGTTDRDRASLIAGTTALVV  
AGYLVNPDYQNHDRDLHDQVNPVYRTLADYMEGRPRESFKRPSSGRIA  
FGDQRSIPDVECDPMPRARDRLEDAGFDVWRGQEVESDCPAGTAAGTEPS  
GRTVKNGVVVIQVSKGRRGASPPIFPPIGPPR

**SEQ ID NO:128 orf28 Methyltransferase 253aa**

MAPLTRSLRYYYGDAGREAAAMDAMYRRFVRPGDVVFDIGAHVGDRVACFR  
RLGARVVAVEPQPLCMRALRALYAHDDRVALVEAACGPAGGSVPLYINSA  
NPTVSTNSVRFLTAATGSRGWENEVWDQQITVPAVTLDTLVQRFGLPAFI  
KIDVEGYEDAVLAGLSRGVRALSFEFTTIARDVARRCLDRAGELGFDGFD  
VSLGETMARTFGRWAARDEMLAHLAAGLPHEANAGDVYAVSRGAWPDRRE  
DRR

**SEQ ID NO:129 orf29 Hydrolase 261aa**

VPRFDSRLVTVGVRTHDRHACHAGGLPVVLVHGLAVSHRYLMPTAHALA  
GRHPVLVPDLPGFGFSDKPRRAYDVGRHAEHLAAWLDVLGVPRACIAGHS  
FGAEVAARLAVLRPDLVAAVVLASPTTDPAAARSRRALIGRWAVDLWIEAP  
WQAPVLVRDIADAKPWRVLATVGHSVRNAIEEDLRRLPVPPVLVGGSLDP  
VAPLRWRAEVAAMTGGVSVTVPAAAHNVMTTSGVRSARATAAYLRTRRC  
MDRLIGGMPPP

**SEQ ID NO:130 orf30 Response regulator 244aa**

VSPFFRLDEAVADVGLRLAPVLRRLLRHSGRLTGSVAGSVSLIDADRG  
CYVKAAEYGANCQLGRSFPLDEGATGRAFGSRRPVVIPDYGQLRAGHLAA  
AHPARKGPAVAVPIWWRGDVIAVNVAFAFAPFSLGGVDELEALTQSAAAAI  
VRSRGVRVRADPPYAAPAPFTPREAEVLDLLRQGLTDREMARRLGLSAK  
TVEKHVGAIIRKKTGTSNRATAAVVTALDNDWVGNLPHTAEHTTGS

**SEQ ID NO:131 orf31 Fructose biphosphate aldolase 281aa**

MKDILDRALAERYGVAAFNIVNDLTVEAVLAAAAEERAPVILQTSVKTVR  
MYGRPRLYEIVHAFADAPVPVTLHLDHCPERSVISDCLAGGWNNSVLFDA  
HELDVADNLRQTTEVVAEARRAGAHVEGEIEGIQGVEDDVGNDYAPMVQS  
LEVAVDFIKRTGVDCFAPAIAGNAHGQYKQAPVLNTRVSDILVAATGIPMA  
LHGGTGLSDEQFTDLIARGCAKVNISTALKESFMKSGLEFLREADERGKW  
DPPSLFRHQRAAVVEMARQHIRLFGGSGRAW

**SEQ ID NO:132 orf32 Hydrolase 192aa**

MGSAVAVPASAGGRRDGPAAHPALRRIGARVVNALVFDCDGLADTERH  
 GHLPAFNATFEQFGLPVRWSEEEYGEKLRIGGGKERMASLFADPAFAAAA  
 GDTDRTELLRTWHRAKTAAPTKLVAEGRIPARPGTARI ISEALRAGWTVA  
 VASTSAEDSVRAVLVNAVGAATTAERIPVFAGDVVPAKKPDPA

**SEQ ID NO:133-SEQ ID NO:199**

(Null)

**SEQ ID NO:200: CosAL02 (40402bp)**

GATCAGGTCGACCGCGGAGCTGATCTGGTCACGGATCGCGCGGATCGGCAGGTCCATGCCGGCCATCAG  
 CACCATCGTCTCCAGCCGGGACAGCGCGTCCCGCGGCGTGTTCGCGTGCAGGGTGGTCAGCGAGCCGTC  
 GTGGCCGGTGTTTCATCGCCTGCAGCATGTTCAGCGCGGCGGCGTTCAGGACCTCGCCGACGACGATCCG  
 GTCCGGGCGCATCCGCAGCGCGTTCGGGACCAAGTTCGCGGGTGGTGATCGTGCCCCGGCCCTCGGAGTT  
 CGGCGGCGCGGACTCCAGGCGCACCACGTGGTCTTGACCAAGTTGCAGCTCGGCGGCGTCTCTCGATGGT  
 CACGATGCGCTCGTCGGCCGGGACGAAGCTGGAGAGGACGTTTACGATCGTGGTCTTGCCGGAGCCGGT  
 GCCGCCGCTGACCAGGATGGTGCGCCGGCGCGCACGCATCCAGGAACCTCGGCGGCGTCCCGGGT  
 GAGCGTGCCGTACTGAGCAGGTACCGACGCGTGGAGCGGACCGCGGCGAAGTTGCGGATGGTCAGCGT  
 CGAGCCGTCAGCGCATCGGCGGACCGCGCGTTCGACCCGGCTCCCGTCCGGCAGCCGGGCGTCGAC  
 GGTGGGCTGGCCTCGTCCACGCGGCGGCGGACCGCGGAGATCCGCTCGATGATCCGGCGCAGGTG  
 CTGCTCGTCTGTCGAACCTCGGCGGCGACCTTCTCGAGGCGGCGAACCCTCGACGTAGACCGAGTACGG  
 CCCGTTACCATCACCTCGGTACCGACGGGTTCGCGCAGCAGCAGCTCGATCGGTCCGTGGCCGAGCAC  
 CTCGTGGTCACCTCGCGGGTGATCCGGGCGCGTCCGCGCGGAGAGCGGGTCTCTCCCGGCGGAG  
 CAGGTCTGGAGCGCGTCCCGGACCGGCGGTCCAGGTCTCTCGCTCTGGCCGGTGGTGTAGAGCGTCGG  
 CCCGAGCTCGTCGCGGAGCCGCGCTGGATCCGCAGCCGCACCTCGCCGACCTGGTCTCGCGCGGCG  
 GCCGGCGGCGCGGTAGCGGTTCGCCCGCGGTGGTCTGGTGGGCGGTGGTGGGGCGGTGCCGGCGCGCT  
 CAGGCGGCTGGAGAGGCTCACGGGAACAACCTCCCGAGGAACCGCGGCGCGCTGGGGTGTGGTGCCG  
 ACGCCGGCGATCCGGTCACCGAGCTCCCGGATCGCCCGGCTGACCGGGTGCAGCGGGTCCGGTGACGGCG  
 ATCGGCTCGCCGTGGTTGACCGAGACCGTCACGTCCCGGCTGGCCGGCACCTGCACCGCGAACGGCATG  
 CCGGCCGCGACCTCCACCTCGGACGGGCTCAGGCCGACCTGGGCGCCGGCCCGGTTGAAGACCAGCAGC  
 CGCTTGTCTTGGGGTAGTCGAGCAGGTGGAACATCTCCGCGGTGAGCCGGACCGACTTGAGCGCCGGA  
 AGATCCGGGGTGACGATCGGGATGAACAGTCGGACATGTCCAGCGCGGCCAGCACCTGGTCCGGTGACC  
 ACGGACGGGGTGTCCACCACGATGAAGTCGTAGAGCGGGCGGGCCACGTCCAGCAGCTCGACGACGAAC  
 TCCCGGCGGACCTGCTCGCCCTCGGCCGGGCTGGCCGGCGCCAGGAGGGCGTCGAGACTTTGCCGGTAC  
 GTCGTGACGATCGACCGCAGCCCGGGCTCGTCCAGCCGACCGGCCATCTGCAGGCCCCCGGCGATGTTG  
 CGTTCCGGGGACAGCTTCATCATGATCGCCACGTCCCGAAGTGCAGGTCCAGGTCCATCAGCAGCACC  
 CGGCGTCCGGCACCGGCCAGCGCAGCGGCCAGGTTACCGCCACCACGCTGCGCCCGCAGCCGCCCTTG  
 CCGGCGAAGACCGTGACCACACTGGCGTACGGGGCGTTGTGCGCGGTCCCGCGCATGGTGGTGTCTCAGC  
 GCCTTGGACAGGTGACCGACCGGACCGTTCGCGTTCGCGATGCCCTCCTCGTCTCTCGGCGACCGAGG  
 TCCCGGATGCCGGACTGCAGGCAGCGCAGCACCACCGCGGCTCGATCGCGTGGCGGATCAGCACCACC  
 CCGATCACCGGGCGCTGCAACCGCTGGTACCGCGTGAACATCAGCGCTCGTCCAGGTCCACGACCGCG  
 CCGATCACACGAGCAGCGTCTCCGGGTAGCGGGGAGATACGACTCCAGCTCGCTCATGCTCTTGAGC  
 ATCGTGCCGCGCATCGCCCGGAAGTCGTTGCCCGAGTGCCTGCGGCGCTCGTCTCGGGCTCGACGTAC  
 ACATAGAGGCTCATCGCACGACCCAGCTCGTGTGATCGCCGGGCGGCTGCTCGCGGTGGCGTTCCGGC  
 CGAGCAGCGCCAGGTAGAGCGATCCGCTCTGGGCGGCGTGCACGAGTCGCAGGGCGTCCGTGTCCCGGA  
 CGGCCAGGGTCACGACGTACCGCTGGATCTCCTTGACCGCGGCGGTGCTGTGCGCGGCCGAGGCGGACG  
 GACCGGGGGTGGGCGAGGGCGACGGCGTACGGGGGCTGACCGGTTCCGCCTCCCGATGGTGATCA  
 CCCGTGCCTTCGGCAGCAGGAGCTCGGTCTCGGGATGGTCTCGTGGTCTCCGCCATCAGGACCTTCG  
 CCTGGTACGTGTAATAGACCGCGACCTGGTTCGCCGGGCGTGATGTTGCCGGCTACCTGCGGGGCGACGT  
 TGAGCGCCACCGAGACCGCGAGCGACCTACGGGGCACCGGGATCCGCCCGGTGTGCGACGGCGCGGTTG  
 ACGCCGGGACGAACAGCGTCCGCATGAGGAGTGCCTGGGGTGCAGGTCCCGCGGAGCCGAGCGGGT  
 CGAGCGCGCTGTCCAGGTGGTCAGCGCCCCGGCGGGCACGGTGACGGTGGGGACCAGGAGCCGCTCGG  
 TCAGGCCCCGGGCGCGGATCTCGGCGCGCTGGTCCCGGACGGGATGTGCTGCCCCGGCGACCGAGGATCC  
 AGGTGCCCTCCCGGCGGCTGAGCGCGCGCGGTTCGGCCGACCGGGCGTAGGACAGGACGGCCGCCCGC  
 TGATCCCGGCCAGCACCAGCGCGGCCAGCAGGATCAGGATGCGGCGGCGCATGACTTCACCTTCTCAGG  
 CACCGGTCTCCCGTACGGCCGATCACCATCGCGCCGTAATACCGCGAGGTGGCGAACCGCGTCCGGTAC

GTGGTCCGTGACCAGCGCTCTGGTGAAGTAGCCGTAGAGGCAGGACTGCGCCGCGCACAGCGCTCTCTG  
CGCCCCCGGCACGAGCGAGGGGACCGCGGATCCGGGTGACACCGGGCCGCTGACCGGGCTCTCGTAGCC  
GGTGAGGACGAGCGGCGCGAAGCCGGCGATCCGGTAGTACGACGGCAGCGTGCCGATACCGCCACCTG  
CCTGTGCAAGATCGGCACCAGCACCGGCTGCCCCAGGTGATCAGGACGTTACGCCGGTCGAGGCAGGC  
CGTGGCGGCGGAGGCGTTGCCGGGGCCGATGGTGAATCCGCCGACCCAGCTGTCCGCCGGCCCGTCGGT  
CGCCGGGATGCCGGTCAGCTCGCAGGAGGCGTCCGGTGCGCTGACCGGCAGCAGGCCCGGCAGCGCCGG  
CGGGCCGTCCGGCTGCCCGAGCCACGTGTAACCGGCGACGGTCTCGCCGGCGTCGGCCGGGCAGGACGT  
GGCGAGGATGCCGTGCTGATCGGGATGTAGCCGGCCGTCGGGTGCGAGAGGCCGAGCAGCGGGTGGAC  
GCCGGTCTGCGGGTTCGGGGCCGGTGGGCGGCGGCGGGGCGAAGAACCCTCGTGTAGTCGCCGGTCAGCCG  
CAGGAAGTCGCACCGGGAGACGCCGAGCGCGAGGACGTCCGTGACCGCCGGAGCTCCCCAGGCGACCCG  
GCCGAGGCGCCCATCTTCGCCCCGTGGTACGCCGACCCGGCCAGACCGCGCCCGAACAGCGGGCGCAC  
CACCGAGGTGTTGTGCTGTTGCGGGTGGACGTCCGGACCTCGACGTACCGGTACGGCCCCGAGGCGCT  
CGGCGGCGCGGGGCACGTGACCGGGGTGTTCCAGGAGCTCGGGCAGCCGGCGTCGGCGGTGGTGACGCC  
GGCGGCGGAGACGGTGGTGATGCAGAACTGGACGTCCGAGACGAGGTCTTGGCGTTCCGGTCGGCGTA  
CCGCTGCGCGTTGTCCCGCTGCGCGGCGACGGTGCAGGTGGCGGACGTACAGTCCCGGCCGGCCGTGCC  
GGCGCAGGCCTGCGCGACCTTCCACGACGCCGCGTCGGCCCCGCTCTGCAACTGCTCCCGCTCGGCGTA  
GAGCGCGCCGATGTGATCACCAGCGCCGCCATCCCGAGCAGGACGCCGGCGCCGGCCAGGACGGCGAC  
CAGCGCGGTGATCAGCCGTGCTCGCCGCGGGGCGGGAACAGGGCGTGGATCAGCCGACGCATGACATC  
ACGCCGGTCGCGCTCATCGTGATCGTGCCCATGGTCTGCCCGCCGACCGAGCCTGACCAGCGCCACCATC  
GGCGTGATCGGCTGGTACCGGCGGGTCAGCACGACCTCGGCGTCGGCGCCGGCGAGCGAGGTGGCCGAA  
CAGGTCTGTGACGGTCTGGGTACCGACGGCGCCCCGGTGGCGCCGACGATGCCGTTGACCTTGGCGTGC  
ACCGCCGCGGCCGTGCCGTTCCAGGGCGCCGAGCCGGGCGCCCTCCCGGGCCGCTCGGTGAGCTGGATG  
TACTGCTGGAGCAGCCGGCCCCATGTGATGATCCCGAACAGGATCAGCAGCAGCACCGGCATCACGATG  
GCGGTCTCCAGCGCCGCGGCGCCCCGGTCCCGGTTATCGTGCGGAGCTCTCCGCTCCCTGGGCGAGAAC  
CGTACGCAACATCCCGGCCAGCCGGTCCGCCGTATACGGCTTCGCCACGAATGGCGACCCGGCACGGAT  
CAGCCCTTTCTTGATGGCGATCTCCTCCGGGACCCCGGAGACGTAGACGATCTTCATGCCCGGCCGGAC  
CTCGGAGGCCGAGCGGGCCAGCTCCCCGCCGAGACTCCCGGCAGGCCCAGGTCCGTGAGCAGCACGTC  
GATCGCTCCGCTGTGCACCCGGCACGTGATGATGGCGCTGACCGGGTCTTTCGCCACGAGGACGACGAA  
CCCGCGCATCTCCAGCATCTGTGCGGCGAGCTCACGCAGGTCTCTCGTCTCTCCACCAGGAGAACGAC  
CGGCCCTGCCGCTCGGTGCTTTCCACATCATTCCTCTCCCGGTGCACGAGGAATCCGGCGACGACTC  
TTCCCGTCGCTGCCTCTGCTCACGCTATCCACCGGCCTGCCTGGCCCCGGTGGCATTCGCGGAAATGTG  
ATGCCCTCAAATCAGCATTTGTCCGGCGGCACTGTCCGTGTTAACGCTCACTACGTTGCGCTGACGTCA  
CGTTTCGTCCCGGTAACAGTTCCGGGTCTTGACAGATCACGAGCGTCGATGGGTGAATGTGTTCAATCC  
TGATCGGACGAGTGACAGGATAAACAACGTTTGCCTGTGAAGCCGTTTGGCTTTGAGGGGGCTCCATGTT  
CAATTACGTAGTTTCGTGCGCCCTAAAACCTTTGCCGCCACCTTCGCGGCAGCCGCCCTGCTGCTCGG  
CTCGGGCGCCTGCGCGAAGAGCGAGGACTCCGGGGACACCGTGCGGCGGGTCCCGCCCCGTGCGCGGC  
GCAGGTGGTCCAGTCCGCCTCGGCCGGTCCGCGACCTGTGCTTGGATCAGTACGGCGCGTCCAAGCT  
CGACCTGAAGACCGCCTCCGTGCGGCTTCTCCAGTCGGAGAAGGAGGCGAACCCTTCGGATCGCCGA  
GACGAAGTCCATCAAGGACGAGGCCGCGAAGCTGGGCATCACCAACCTCAAGACCTCGAACGCGAACTC  
ACAATTCAACAAGCAGATCGCCGACGTGAGCAGATGATCGATGCGGGCGTGACGTGCTGGTGATCGC  
GCCGCTCAACTCGGACGGCTGGGACTCGGTGTTCCGCAAGGCGACGGCGAAGCACATCCCGATCATCAC  
GATCGACCGGAAGATCAACGCGACCGCTGCAAGGACTACCTGACCTTCATCGGCTCCGACTTCGCCGA  
GCAGGGCAAGCGGGCCGCGACGCGCTGGCCAAGTCGCTGGGCAACAAGGGCGAGGTGGCGATCCTGCT  
GGGCGCTCCCGGCAACAACGTACACGCTGCGGACCAGCGGCTTCAAGGACGAGATCGCCAAGGTGCG  
GCCGGACATCAAGATCACGTTCGAGCAGACCGGCAACTTCTCCCGGAGGACGGGCAGAGGTGCGCGA  
GCAGCTGCTGCAGTCCAAGCCGAACATCAACGGCATCTACGGCGAGAACGACGAGATGGCGCTCGGCGC  
GATCACCGCGCTCAAGGGCGCCGGCAAGAAGCCGGCGACGTCAAGATCGTCTCGATCGACGGCACCAA  
GGGCGCGGTGCAGGGCATCGTGACGGCTGGGTCTCCGCGGTGATCGAGTCCAACCCGCGCTTCGGGCC  
GCTGGCCTTCGACACCGCGACGAAGTTCTTCGGCGGCGAGCCGGTTCGGCCAGGACATCGTCATCCAGGA  
CCGTGCCTACGACGAGTCGAACGCCAAGACCGACATCGGCAGCGCGTACTAGAGAGCGCTCCCAATCGG  
GTGTCCGGGGATGAACCGGGCACCCGCCGGCACGGAGGAGGGCGGCTCATGCTGCTGGAAGTCTCCGGC  
GTCTCCAAGACCTTCCCCGGCGTACGCGCCCTGGACGGGGTGTCTTACCCTGAACCCGGGCGAGGTG  
CACGCGCTGGTGGGGGAGAACGGCGCCGGCAAGTCGACGTTGATCAAAGTGCTCACCGGGGTCTACCAG  
CCGGACAGTGGCGAGCTGCGCTACCGCGGCGAGCCGGCCCGGTTCCGCCACCCGCTGGACGCCAGCGG  
GCCGGTATCTCGACCATCTATCAGGAGGTCAACCTCGTCCCGCTGATGAGCGTGGCGCACAACTGTTC  
CTCGGCCGGGAGCCGCGCAACCGGTTCCGGGTGCTGGACGAGGCCCGGATGGTGCAGGAGGCCACCGAG

ATCCTGGCCGGTTACGGCGTACGCACCGATGTCCGCCGCCGCTCGGCACCCTGGCCCTGGGCGCGCAG  
CAGATGGTCGCGCTCGCCCGGGCCGTCATGGTCGACGCCCGGGTCGTGGTGATGGACGAGCCACCTCG  
TCGCTGGAGCCGCGCGAGGTGGAGACCCTGTTCCGGGTGATCCGCGAGCTGCACACCGCGGGCATCGGC  
ATCGTCTACGTCTCGCACCGGCTGGACGAGCTCTACCGGGTGTGCGACGCGGTACGATCCTGCGCGAC  
GGCAAGCTCGTGACACCGGCCGGATGGCCGATCTCGACCGGCGCACGCTGGTCTCGCTGATGCTCGGC  
CGCGAGTTCGGGGCGGACTTCACCAGCTTCTCCGAGTCACCGCAGAGCACCCCGGAGGGCGAGCCGGTC  
CTGCGGGTGTCCGGCCTGACCAGCCGCCCCGGCTCGACGACATCAGCTTCGACGTGCGCCCCGGCGAG  
GTGGTCGGCCTGGGCGGCCTGCTCGGCGCCGGCCGAGCGAGACGATCAAGGCGATCGGCGGGGCGTAC  
CCGATCGACTCCGGCGTGATCGAGGTGCGCGGCGTCCGGCTCGGCAGGCCAGCACGGTACGGGCGGTG  
CGCGCGGGCGTGGCCACCCAGCCGGAGGACCGCAAGGCCGAGGGGATCGTCCCCGGCCTGTGATCCGG  
GACAACATCGCGCTCGCGATCCTGCCCGGGATGGCCGTTTCGGACTGGTCAGCGACAAGCGGATCGAC  
AGCATCGTCGCCACGTACATGAGCCGGCTGCGGATCAAGGCGTCCGGTCCGGACCAGGCGGTGCGCGAT  
CTCTCCGGTGGCAACCAGCAGAAGGTGCTGCTCGCGCGGCTGCTCGCCACCGGCCCGAAGGTGCTGCTG  
CTCGACGAGCCGACCCGGGGCATCGATGTGCGCGCCAAGGCCGAGGTGCAGGCGCTGATCGACGAGCTG  
GCGAAGGAGGGGCTCGGTGTGCTGCTGCTGCTCCTCGGACGCCGAGGAAGTGGTCGAGGGGCGCGGACC  
GTGGTGGTGTGCTGCGCGACGGCGCGGTGCTGCGCACCCCTACCGGCGACCGGGTGACCACCGAGGCCCTG  
ATGGCCACGATCGCGGAGGGCCGCGGATGAGCACTGAGACCCTGACCCGCCCGCGGATGACGTTCAACCC  
GGCGTGGGCGGCACGCTACGGCGTCTACGCGGCGATCGTTCTGCTGATCGTCTCAACATCGCCCTCAC  
GCCGTACTTCTGACCCTGAGCAATCTGCGGATCCAGCTGATCCAGGCGGCGCGGTTGGTGATCGTCTCGC  
GCTCGGCATGGCCCTGGTCATCGGCACCGAGGGGATCGACCTCTCGGTGGGTTCCGGTCATGGCGCTCGC  
CGCGGCCCTTCATACCCCTCTATCTGGGGTACGGCGTGACAGCCGCGATCCTCGTCTCGCTCCTCGCGGG  
TGTGGCGGTGCGGGCTGATCAACGGTGTCTGCTGCGGAAGGCCGGCCTGCAGCCGATCGTGGCGACGCT  
GGCCCTGTTCTGTCGGCGGTGCGGGGCTGGCCGTGGTGATCTCCGGTGGACAGCTCAAGGACGTGCGCAA  
CGCCGACCTGCTCTACCTGGGCTCCGGTGACCTGCTCGGCGTGCCGGTCTGGTCTGGATCGCGGCGCT  
GCTGGTGTCTGTTGGTGGCGTTCTGTTGCTCCGGCGTACCGTCTTCCGGCCGGCGCCTGCTGGCCGTGCGCGG  
CAACCGGCCCGCGCGCGAGCTCGCCGGCCTCCCGGTCAAGCGGGTGTGATCGGCGTCTACGTCTTCTG  
CGCGGTGCTCGCCTCGATCGCCGGCCTGCTCTCGGTGCGCGCATCCAGTCCAGCGACGCGTCCGCGGT  
CGGCCTGCTGATCGAGCTCTCCGCGATCACCGCGGTGGTGTGTCGGCGGCACCCCGCTCACCGGCGGCCG  
GGTCCGGGTGCTCGGCACCGTGGCCGGGGCCCTGCTCATGCAACTGGTGGTGGCCACCATGATCAAACA  
CGATCTCCCGCCGTCCACCACCGAGATGGTGCAGGCCGTGATCATCTGGTTCGCGGTCTACGTGGCCCG  
GGAGAGGAGGACCCGGTGACCATGTGATCCCGGTGCCCGCTTTCGGGAACGGCGGCTTCGTGCAGCGC  
CAGGGCGCGCTCGCGGTGCTGGTCACCGTGGTGGCGATCAGCCTGGCCGCGTTCGCCGGCTTCGCGAGC  
GCGGACAACGCCGGCACGATCCTGGTTCGCCGCGGCGCCACCGATGCTGATCGCGCTCGGCATGACCTTC  
GTGATCATCACGGGCGGGATCGACCTCTCGGTGCGCTCGCTCTACGTGCTCGGCGGGGTGGTTCGCCGCC  
TGGGCGTTCGAGTGGGGAGTGTGTCGGCGGCTGGCCGCGCGCTGCTCCTGTGCGGGGCCATCGGCGTA  
CTGAACGGGATCCTGATCTCGAGAACCGGGATGGCGCCCTTCATCGTCAACCTCGCCGCGCTGCTCGGC  
GCCCCGTGGCCTCATGCGCAGCATCAGCGATGAGGGCTCGACGACGTACCTGGTTCAGGAGCGACGTCTTC  
CACGAGCTGGGCACGGGATCCCTGCTCGGCGTGGTCTGCCGGTCTGGTGGCCGCGGTTCTGGTTCGGC  
GCCGGGATCCTGGTGTGAACCGGACCCGGTTCGGGCACGCCGTGCAGCCATCGGCGGGAGCGAGGAC  
GCCGCCGCGCTGATGGGGCTGCCCGTACGGCGAATAAAGGTCTGGGTCTATCTCCTCTCCGGGTGCTC  
GCGGACTCGCCGGTGGGATCAACGCGGCCAAGCTCGGCTCCGGGGTACCGTGTCTCGGCAGTGGCATG  
GAGCTCGACGCCATCGCCGCCGTGCTGATCGGTGGCACGCTGCTCACCGGCGGCTCCGGGTGATCGCC  
GGCACGGTTCGCGGGGGTGTGCTGCTCGGCGTGATCCAGAACCTGATCAACCAGGTTCGGCAACGTCAAC  
AGCAACTGGCAGCAGGTGATCAGCGGTGGGTTCTTCGCCGCTGTGGTGTGTCGCGCAGACAACCTCTCGTA  
CGCGCAAGAAGATCTTGACGTGGAGGCCCGCCGGGGACCTGCCGCGGGCGGGCAGGCGTAGGTTGCC  
CGCATGACGGAGCATCGGGTCTCCCCGGCCGGGCGGGCGCTCGCGTGGCTGATGCTGGTCTCCGGGTTC  
GCCGCGGTGCTCGGGTCTTCTGCCGTGGGCTGGGTGCACTTCGGGAGCAGGGCAGCAGCCAGGTG  
TCCGGGAGCCTCGTGTCCGACCTGTTACCGCGTTCCTTCGGGGTGTGCTCGCCGGGTACGCGGTGCC  
GCCGTGCGCGGCGAGCGCCTGCCCGGCTCGAGACGTTCTTCTCGGTGCGCGCGGCTGGTGTGTTTC  
ACGCTCGGGGCTGGGACAGCCGGCACCTGACCTCGATGACCGCCCGCTGGTTCGCGGGCGACCCGGC  
ACGACCATCGGGCCCGGTCTCTGGCTGTGATGATCGGCGGCTGCTGGGCGCGCTCGCGGCGGCGGCC  
CTCGCCGTCCGGGGACGCCGGCCGGTCTGTCAGCTCCCGGGTGTGACCAACAGTCCGGGTGGTTTCGGCA  
TCGGAGGGGTGCTGACGCCGAGGAGCCAGGAGGTGAGGAACGGCCGAGATCCTGGCGGGCGACCCGGC  
AGGCGAGCGTGATGAAGTCACCGGTGAGGCGGAGCGGCCGCGGTACGTGCTCAGCCAGGCGCGTTTCGA  
CCCCTGGAACGTGCGCGCGCGATCTTCTGCCGAGCGCGTAGAGGACCAGCGCGCCCGCGGCGTAGA  
CGTTCCGGGTGAAGACGTCCACGTGGTTCGCCGCTGAGCGGGGCGGCGACCGGACCGTACCTGGCGC

GGAAGATGTCGCCGGCGGGCTAGACGGCCTTGAAGAACGCCTCGCGGTGGGCCAGGCCGGTGTACTGCG  
GGAACCCGCCGGTCTCTCCGACCAGAGCATCTCGTACCAGGTGGCGTGGCCCTCGTTGAGCCAGACGT  
CGCTCCACGAGAACGGCGAGACGCTGTGCGCGAACCACTGGTGGGCCAGCTCGTGGGTCTCCAGGGCGT  
CGCGGGTGGTCTCGGGGCCGGTGAACAGGGCGGGCGCGGTACAGCGAAAGGGTCTGGGTCTCCAGGGCGT  
AGCCCAGGTCGTCTCGATCACCAGGGAGCCGTAGTCTCTGAACGGATAGGGCCCGGCCTGCTTCTCCA  
TCCAGGCGAGCTGCTCGCGTTCCCCGGCGATCGCGGGCAGCAGCGTGCCGGCGAGGCGTCGCGGCACCA  
CGTCGCGGATCGGGGTGCCGCCGGCCGCCGGGCGGCGCTCGACCACGAAGTCGCCGACGGCCGTCTGCA  
CCAGCTCGGTGGCCATCGGGGCGGACTCGCGGTAGACCGAGCTGACGTGGCCGTCTGTCTCGGTCTGTGC  
TCACCAGCGTGCCGTTGGCCGTCCCGGTCCAGCCGGCCGGCACGGTGAGCGTGATCGTGAAGGTCGCCT  
TGTCCCGTGGATGGTCTGTGCGGGGAACAGCAGGTGCGCCGAGCCGGGCTGGGCAGCGAGCACCGTCC  
CGTCCGGCGTTCGCCACGAGAGCGGGCCGGGGCGGCGGTGAAGTTCGCGACCGTGACCCGGAACAGCCGCC  
CCCGGGGAAGGGGGCGCCGCGGGATGACGGTCAGCTCGTTCGCCGTCCCGGGCGGGCCGCGCGGGCTGCC  
CGTCGACCAGACGCGCCCCGATCGAGGCGCCGCCGAAGTCGAGGTCGAAGCGGGACAGGCTCTGCCCGG  
CGACGGCGGTGATCGTCACATCGCCGGATACGACCTGCTTCGGGTCTCTCCGGGTAGCGCAGGCGGGA  
GGTCTGTAGTCGAGCACGTCTGTAGCCGCCGTTGCCGAGCAGCGGGTAGAGGCGGTGCGCGAGGCGGCGG  
ACCCGGGTGACGGCGAGGGCCCGGCCTGCGCCGGTGCGGGCGGTGCGGCCAGGGCGAGGACGACGGTGG  
CGCCGACGGTGAGGACGCGGGTGCCCGCGGACGTTTCGATGGCTTCCGACGCTAGGCGACACCACCGGA  
AGCCATCGACCCATGGGCATGTATGAAATCCCAATCTCCGATGTGGACGGACGGATGGTGCGCAGTA  
TGCAACTTGAGATTTTCGGGACATTCCTAGCCCGCGGTTGCCGTGTGCACCCGTTTCGACCACTGTCC  
GGAACCGCGTGGTGAGCAGGGTGAACGTTTCATGCGGCCCGGGCCGCCGGAGGTCCCGTGACCGTTTCGCA  
TTGCCGTGCGCCGGTGAATCACTCCGCTTTACGACTTCGCGACAAGTGGTTAATGTGCGGCGAAGCGATA  
TATCGTCCAGTTATTTCGGGGGACTGCGGTGCATGAGCAAAAAATCACTCCGGCGATTGCCAGCGGATCA  
CCGCAAATTGAATTGCTGCCCGTCAATTCCTTCGAGTGATGGATTTCGCCGCGACTGAACGGTGAGGAC  
CCGCGGCACACCGAGGTGCTCGCCGGCTTCGGGGCGGAGCTCCCGCCGATCGTCTGTGCACCGGGCGACC  
ATGCGGGTTCATCGACGGCGCGCACCCGGCTGAGCGCGGGCCCGGTGCGCGGCGACGACCGGATCAGGGCG  
GTGCTGTTTCGACGGCACCGAGCAGGAGGCTACGTGCTGAGCGTCAAGGCCAACGTGACGCACGGGCTG  
CCGCTGTCCGCCCGCGAGCGCACCCGTGCGGCGGAGCGCATCATCACGATGCATCCTGACTGGTTCGGAC  
CGGATGATCGCGGCGTCCAGCGGGCTGGGCGCGCGGACCGTTCGGCGGCCCTCCGCCGGCGGCGCGCGGCC  
TCCGGCGAGAGCCCGGCGGGCCTCCGCTCGCGCGCCGGCCGGGACAGCCGCGTCCGGCCGGCCGGCAGC  
ACCGCGGGCCGGCTGAAAGCGGTGCACTACCTCCAGGACCGGCCGGACGCCCTCACTGCGGGAGATCGCC  
CGGCACGCCGGCGTCTCTCCGTGACCGCCCGTGACGTCCGGGACCGGTGCACCGCGGCGAGGACCCG  
ATCCCCGCCACGCAACGCGCGGCGGCACGCCCCGGAACGATTCCCCGCCGTGCGGTGCTGTTCCAG  
GGCTTGGCGAGCGACCCGTCTGTGCGATTTCAGCGAGTCCGGGCGCGATCTGCTGCGCTGGTTGATTGCC  
CACGCCGTTTCAGGACGGCGAATGGAAGGGCTGGTTCGACACTATTCCGGCGCATTCGCGCAGGCGCTG  
GCGAAAATCGCGCGCCATTGTTTCGCGGGAATGGCGTGAGTTTCGCGGACATCCTGGAGAAGGACGCCGCG  
TAGGCCACCGGCATTCTTCCGGCACCCGGTCCGCCCTCCGTACGCACAGGTGAGAGTACTGATGGCG  
GCTGTCAACCCGCCGATCTGCGGGTAGGTGCCGACGTCTTCCGCCGACCGTCTGCCGAACGCATCCGG  
TCGGTGCCCGATGTGCGGGTAGGCATTACCGATCAGGGTTGCTCCAATGTCTCTCGTCGATACCGCGG  
CTGATACCTCCCGGTACTCGGCAGAACACTCCGTGGCTATGGCGCCGTGGTCCCGGAGCCCTTTCCTGA  
AGTACTTCGACGTACCTGTCCGGAGTCCCCACAGAAGGAGACATCATGTGCGCCATCACCCTGGAAA  
CCACCTGGAAGAACACCGACCTCCGGGAGGACCTCACCCTCACCAGGCGGGCCCTCGGCTTCGGCGAGC  
TGAGCTTCGAGGACCTGCGTGAGGACCGCACCATCTACGCGGCCAGCAGCGGCTGGGTCTGCACCTGA  
CCATCGAGTGCGGCACGCTCGTCTGCGCCTGCTGACCAACGGTCATCGCACTGCCGGGACCGGGGCTGA  
CCCCGGTCCCGGCACCTCTCCCCGGCAGGAGCCTGCATGTCTATCCTTTGCGATCGCCGCTTCGCCGGC  
CAGCGCGTACCTGCACGAGCGCTCCGCCGGCCCCGGCGGGCGACCCGTGGCGGAGCACGAGCGGGTCA  
GTCGTGGCGCGAGTCCGCGTTCTTGACGACCCGGTCTTGACATCCGCTGCGCGAGCTCGGACTCAG  
CCGGGCCGAGTTTCGGCCGGTGTGACCGACGGCGCGTACGACGCCGGGAGCACGGCCCTCGACTGGGC  
CGGCGAGCTGGCCGCCGTGCTGGCGACCGGGACCGGCGCGGTACCGGACTCGCCCGGTTCGACCAAGCT  
GTGGGCGCAGGGCTTCGACCGGCTGCCGTTTCGCCGGGCTGATCGAGAGGTTCTTCGCTACTACGAGCC  
GCGGGTGC CGCGCACCGCCGGGACCGTCCGGGTGTCCCTGCTGGAGAGCCTCGCGAACCAGCCTGCTCAC  
GGTTCGCGACCCGGACCTGCTGCTGGAAGTGAACGTGGCCCCGGGTGCACGGCCGCTGACCGCGGCCAC  
CCCCGGGGAACGCTACGACCACTACGACCGGGTCTGCTGACCGACCGGACTACCTGCGCTCGCTCTT  
CGGCGAGTACCCGGTGTCTCGGCCGGGCCATGGTTCGAGTGCGGCCGCCGTGGGCGTGGCGATGGCCGA  
GCTGTTCCAGCGCCTGGACGCCGACCGTCCCGCCCTGCACGCCCGGGGCTGCTGCCGGCCGGCGCGGG  
CGAGGTACCCGCGCTCCGGCCCCGACCTCGGAGACCCGCACAACAGCGGCCGCGCGGTTCGCGATCCTGAC  
CTTCCGGTCCGGTGCCAGCTGGTGTACAAGCCCCGGCCGGTTCGGGCCGAGCGCGCTACGCCGAGAC

CGCCGCGGCCCTGAACCGGCACGGCCTGAGCCTGCCGCTGACCGCCGTGGACGTGCTCGACCGCGGCGC  
CTACGGCTGGTGGAGCTGGTCCGGCAGAGCCGTGCGCCGACCGCGCCGACCTGGACCGCTTCTACCG  
GCGTACCGGCGCGGTCTTGGCCACCACTCTGCTGCTCGGCGCCGTCGACGTGCACATGGAGAACGTCAT  
CGCGGCCGGCTCGTCCTGCATGCCCATCGACCTGGAGACGCTGCTGCAGCCCCGGGTCCCCTCCGGTGA  
CGCGACGGACGCTACACCCGGGCGCTCGACCTGCTCAACCAGAGCGTGCTGGCGATCGGGATCCTGCC  
CGCCCGGGCGTTTCGGCGGCCGGGAGCGCAAGAGCGTCGACGTGACGCGGATCGGCGGCGGCGAGGCACA  
GACCGCGCCCCGGCCGGTGCCGATGGTTCGTGGAGCCGTTACCCGACGTGGCCCGGATCGAGGCGGTGGA  
AGCGACCATGCTGGGCGCGCAGAACCAGCGCCGGTGCTGGTTCGGCGCCGAGGTGCGGCCCGAGGAGCACAC  
CGAGGCGGTGGTGGCCGGCTTACCGAGGCGTACGACCTGATCGTCCGGCACCGCGAGGACTTCGCCGA  
CCTGCTGGCCGGCTTCGGTGACGTGAGGTGCGCTACCTGCCCCGCCGACCCGCCGGTACAGCATGTT  
CCTGACCGAGAGCTACCAACCGGACTACCTGCGCGACGCGCGTGACCGCGACCGGCTGCTCGACAAGCT  
GTGGACCGCCGCGGGCGCCCGTCCCGACCTGATCCCCATCATCGAGTCGGAGAAGCGCCAGCTGCTGGC  
CGGCGACATCCCCTGCTTCCGCGCGCTGGCCGGCGACCGGGCGATCCGCACCGCGTCCGCCCCGGTGGC  
GCCGACTTCTTCGACGCCCCCGGCATCGAGGTGCTGGCCGGCCGCTGCGGCAGTTTCGGGCCGGTGCA  
CCGGGCTGCCAGCTGCGGATCATCCGCGAGTCGATGGGAACCATGCCCGCGCCCGGCCGATCGCCGG  
GACACCCGCGCCATCCTCCGAGCGGCGTGCGGGCTCGACCCCCGTGAGGCCGCCACGCTCGGGGACCG  
GCTCGTGCGGGAGCTCGCGGACGAGGCGATCCTCGGGGCCGACGACGCCGGCTGGATCGGAGTCAGCAT  
CGAGGGCCTCGACCAGGAGACGTTACGCTACAAGCCGATGGCGACCGGGCTGTACGACGGCATCGCGGG  
AATGGCGCTGACGTACGCGTACGCCGCCCGCACGCTCGGCGACGAGCGCTACCTCGACCTGACCCGCCG  
TACCGTCAAGCTGGTCTCCGGCTATCTGCGGTACCTCGCCGAGCACCGGATCGTGGAGACGGTTCGGGGC  
GTACAGCGGGATGGCCGGCCTGCTCTACACGCTGGATCACGTGCCCCATGCCACCGGCGACGCGTCTGCT  
GCTGGGCGAGATCGAGGCGCGCTGCCCTGGCTGCGGGAGTGCGCTACCCGCGAGGAGTGCCCGGACCT  
GATCGCCGGCCTGGCCGGTTGCGCCGTCGTGCGCTGTGCTGTACCGGCGGCACGGCATCGCCGGTTA  
CCGCGAGGTGCGGGAGATCTGCGGCCGGCGACTGGCCGGCACCGCGGTGACGTGAGGGCGCCGCGGG  
CTGGGCCGCGACCCGGACCGGCGTGATCCTCGGCGGTTTCTCGCACGGCTCGGCCGGGATCGCCTGGGC  
CCTGCACGAGCTCGCCGCCGAGTTCCGGTGACCGGGACCTGCGCGAGCTGGCCGACCGGGCGGTGAGTT  
CGACCGGCGGCTGTACGTTCCGGCCGCCGGCGCCTGGCGCGACCTGCGGCCCGAGATGGCCGGCACCGA  
CGGTTACCCGGCGCTCTGGTGTACGGCGCCGCCGGCATCGGCCTGTCCCGGCTGCTGATCCACCGGAT  
CCGGCCGACGAGCGGCTCGCCGAGGAGGCCCGCGCCGCGGTGGCGCTGGTCCGGCGGCACGGCTTCGG  
CCACAATCACAGCCTCTGCCACGGCGACTTCGGCGCGCTGGCCCTGCTCGGCCTGGCCGACCGCGCGTG  
GCCCCGGTTCGGGCGGCCACGACGAGCGTGCCGGCGCGGTGCTGCGGGACATCGGCGAGACCGGTCTGCG  
CTGCGGGCTGGGCAACGGCATCCGGATGCCGGGTCTGATGCTCGGCGCCGCGGGCGCCGGACTGAGCCT  
GCTCCGGCTGGCCGCGCCCGCCGACGTGCCCGCGGTACCTGGCTGGAACCGCCGCGGGGCACGCATGT  
CTGAGACGGCCGGCTGCTGCGGCGGAGCCTGCTCGATACCCGGGGCAAGCTCGCCCGCGTGGCCGGTC  
TCGCCGTTCGCCGGGTTCGGCTGCCAGCTGGGCCAGCCGTTCTCATCCGGCGCGTGCTCACGGCGGTGC  
AGTCCGCACAGCCGTACCGCCAATTGGCTCTCGCCGTTCTGGCGGTTCATGGTTCGTGCGCGCGCGCTGG  
GCGCCGTCCAGCAGTTCTGCTGCAGCGCACCGGGAGGCGATGGTCTTACGGTGCGCCGCACGCTCG  
TCGCGCACCTGCTGCGGCTGCCGGTCGCGGCCTACGACGAGCGGCAGTCCGGCGATCTGGTGTCCCGGG  
TCGGCGCCGACACCGCCCAGGTGCGCTCGGTGATCACGTCCGGGTGGTTCGACCTGGCCGGGGCGTCC  
TGCTCGTTCGGCGGCTCGATCGCCGGCATGATCATATTGATCCGGTCTGCTCGGCGTCAGCCTGGCGC  
CGGTGCTGTGCGGCGCCGCGGGGTCCGGCTGGTTCGGCCGTGCGTTCGCCCGCTCAGCTCGGCGGTCC  
AGGAGTCGATCGGCGCGCTACCCGCTCGACCACCCGGGCCCTCGGCGCGATCCGCACGATCCGGGTTCG  
CCGGCGCCACCGAGCGCGAGACCGCCCTGATCGTTCGGGAGGCCGACCGCGCCCGGGCGGCGGGCGTCC  
GGCTCGCCCTGGTTCGCGGCCAGGCCGGCCGATCGTCCGGCTCGCCCTGCAGGGCGCGTTCGTTCGCGG  
TGATCGGCTTCGGCGGCTACCGGGTCGCGAACGGCGCGGTGTCGGTTCGGCGACCTGGTTCGCGTTACGCG  
TGCTGCTGTTACGCTGGCGCTGCCCTGGCCAGCTCGCCGAGGCGGCCACCCGATCCAGACCGGGC  
TGGGCGCGCTGACCCGGATCGAGGAGATCCTGGCCCTGCCGACGAGGACAGCGCGCTCGGCGTCCGTG  
CGAGGACGCCCCGACGGTCCGGCACGATCCGGTGTGCTGGAGTTCGATCACGTGTCGTTCCGCTATC  
CCACCGGCGGGGAGATCCTGCGCGACGTACGTTCCGGGTGCCGGCCGGCAGCACACCGCTCGTTCG  
GGCGTTCGGGGCCGGCAAGTCGACGATCCTGGCACTGATCGCCCGGCTCTACGAGGTCCACGGTGGCC  
GGATCCTGCTGCACGGCCGCGACATCCGCGACTACCGCTCGCCGAGCTGCGGGCCGCCCTCGGCTACG  
TCGAGCAGGAGGCCCCGGTGTGTCGGCGGACCGTCCGGGACAACCTGACACTGGCCGCGCCGGACGTCG  
CGGAACACGCGATCCGCCACGTACCCGATCCGTTCAACCTCGACGACCTGCTCGCCCGTACCCCGCG  
GCCTGGACGCGCCGGTTCGGGGACGGCGCGGTGCTGTTCTCCGGCGGTGAACGGCAACGGCTCGCCGTG  
CCCGGACCCCTGCTCGCCCCGGGCGAACTGCTGCTCTTCGACGAGCCGACCGCCACCTGGACGCCCCGA  
ACGAGCAGGCGCTGCAGCACGGCCTGACCGCGCACGCCGCCGGCGACGCTGGTGGTGGTTCGCGCAC

GCCTGGCGACCGTCGCGCACGCCGACCAGATCCTGGTGATCGACGACGGTCGTTCCGGTCGCCGCCGGGCG  
GGCACGAGGAGCTGCTCGTCCGCGACCCGACCTACCGCGAGTTCGCCACCCGTCAACTGCTGACCTGAA  
TCCCCAGGTGTACGCCATGCTCAGTGTCTGGACCAGGTGCCCGTGTTCGTGGCGACGACCCCGCGGA  
GGCGGTCCGCGAGGCCGTGCGGCTCGCCCGGGCCGCCGAGTCGCTCGGCTATCACCGGTTCTGGATCGC  
CGAGCATCACGGCAGCGCCGCCAACGCGTGCGCGGCCGCCGAGATCGTGGCGGGCGGCCGTGCGCGGAGC  
CACCGAACGGATCCGGGTGGGCACCGGGGAGTGCTGCTGCCGTACTACAGCCCGCTCAAGGTGGCGGA  
GGCCTTCCGGGTGCTGGCGGCGCTGTATCCGGGGCGGATCGACCTCGGGTTCGGGCGGGGCGAGGGGCGG  
GCCGGCGGTGATGGCCGAGCTGCTCAACCCGTACGCGATCGCGACCGAGGAGGCGTACGCCGAGCAGGT  
CGGCCGGCTTCTCGCGTTCTTGGGCGACGCGCGGACCGTCAGCCGGGTGTCGGTCACGCCGGCGGTGCA  
GGACCCGCCGTTGCCGTGGCTGCTCGGCTCCGGCGTCGGCAGCGCCCGCCTCGCCGGGATGCTCGGCGT  
GCCGTTCTGCTTCCGCGAGTTCATCGCGACCGAGGAGTGCCCGGAGGCGATCGCGGCGTACCAGGAGTC  
GTTCCGGAGCTCGCCGTGGCTGGACGAGCCGAGGCGATGCTGGCGTTGCGGGTGCTCGCGGCCCGGCAC  
CGCCGAGGACGCCGAGGAGCTGGCCACCGGCTTCTGGATGTCTGTCACCACCGGATGGCGGGCCAGGT  
CCGGCCGACGACGACTACCGGGGTGGCGTGCCGAACCTGGCGGACGCCAGCGGTACACGCTGACCGA  
GGAGGACCTGGCGATGCGCGCGAGCCGGCCGTACCTGCAGATCTCCGGCACGGCGGAGACGGTCGCGCA  
GGAGATCCGGCGACTGCGCAAGGTGTACGACGTGGCCGAGGTATGCTCACCACGAAGTGTCCCGGGGCG  
GGCCGCCCCGGCGCCGGTCTTACGAGCTGCTGGCCGCCGAGCTCGGGCTGACCGCGCCGGCGTGACCCG  
GGTCAGGCCAGCTCGTCCGTGGTGAGGCCGGCGGCGTACGCCAGCACGGCGGCGCTGCACCCGGTTGGCC  
ACACCGAGCTTGCGCAGGATGACGCTGACGTACTCCTTGACCGTGAGTCTGCTGATGAACAGGCGCCCG  
GCGATCTCCGCGTTGGTCAGCCCTGCCCGATCAACCCAGAACGGCCTGTTCCCGGTTCGGAGAGCTGT  
TTGACCTCGTCGACCGCGTTCTCGGCGGCCGGAGTGCCCCGGCAGGCCGCGCCAGCATGATCGACGAG  
GCCTCCGGCGCCAGCACGGTCACCCCGGACGCCAGCGCGCGGACCGCCGCGACCAAGTTGCTCCGGCTGA  
CTGTCTTGAGCAGGAAGCCGCAGGCGCCACCGCGCAGCGACTCCAGCACGGTGCTGGACGCCGCCAGC  
GTGGCCAGCACCGCCAGCGCCGGCCCGGACTCCAGGTGCGCGCAGCCGGGTGAGCAGCGGCACGCTCTCC  
GGCAGGTCGCGTGCGCGTCCAGCAGGACGACCGCGGGCCGGTGCTCGCTGACCGCGGCGAGCGCGCTG  
TCGCGGTGAGCGGCGACGACGGAGAAGCCACCCATCGTCTCCAGGATCATCTTGATGCCGATCGAGACC  
AGGGCCTCCTCGGCCACCACCAGTACGTCCGCCATGCCTCCCCCGGGCAATGTAGCCGTACGAGATTTG  
ATAGTGCACAAAACCATCGAATTACTGCTCTTTCGAATCATACATGTATTTATGCAAAATTTCTGGTCA  
GGAAACGGGGACGATCCGAGACGCCCCGGGCCGGCGTCTCGGATCGTCCCTTACCGCGCACGGCTATCGC  
ACGAGCCCTTCCGCGGGGGCGATCCGCGTGGCTCGCCGAGCCGGCGCCAGGCTGGCCAGCACACCGGTG  
ACCGCGGCGGCCACCAGCACGAGACCCAGCTGCGGCCACGCCAGCATGATCACCGGTTCCGGCTGCCGG  
CCCACGGCCCGCATCACGCCGACAGCCGACCGGCACCCCGATGACGATGCCGGCCACCGTGCCGACC  
AGCGTGATCGTGACCGCCTCGACGGCGACCATCGCGCGCAGCCGGCTCCGGCGGGTGCCGAGCGCCCGC  
AGCAGGGCCATCTCCCGGGTTCTTTCGATCACCGAGAGGCCGAGCAGGTTGGCGATGCCGAGCAGCGCG  
ATCACACGGTGACCGCCAGCATGCCAGCGACAGTCCGAGCAGGATGGACAGGACGTTTCATGATGTCG  
CCGCCCTCGGTGACACCGCCGCCGACCTCCACACCGGCGTTCGCGTGCCGCCACCGCGTTGACGTGCGCG  
GCCAGCGTCTCCCGGTTCGAAGCCGCCCGGCGCGGTGCCCCAGACCGTGTTGCGCACGGTCCGGACCCCG  
GCGGCGGTGACACCGCAGGTCCCGCCCCCGCGCGGTGACCGTGAGCGGAGCACCGTCCGTACGCCCGCGC  
GCCGTACGGTAGGAGGCCGGCACAGCAGCACCGGGTCCCGGTGGACGCGCTCAGCTCCGGCGCGAGG  
CGCCCGGCCACGTCCGTGGGAGCGCCCGCATCCGGGCCGGCGTGGGTTTGCCCCGCGGCCGGGAACGTC  
GCCGCGCTCGTCCGACCGTGCGCGTGGTCAGCCCGGTGATCCCGGACAGCGCCCGGACCGTGCGGTGCG  
TCGATCGCCGCGCCGTCCGTGTGCACGCTGACCGCCACCGGATAGCGCGCCTCCAGGTCCGGCTCCACC  
GTGGCCCGCCCGCTCGCCGAGGCCACGGCCAGGCCGTTGATCAGCGCGGCCCGGACGACCACGGCCATC  
GTCGCCGACGCCGTCCGGCGCGCGTCTGCGCGAGGTTGCTGCCGGCCAGGCTCGCCGCCACCCCGAAC  
CGTTCCAGGCCGCGCGCCGCCCGGCGGCAACAGCAGGGCGATGCCAGCGGCAGCGCGGTGAGCAGCCCG  
GCGGCGAGCAGCACCCCGCCGACAGCGCGAGCGGCAGGGAGGTGCCGATCGCGGCGAGACCCAGCACG  
CCCACGGCCAGGCCGATCAGGATCAGCCCGGCGACAGGCGGCGCCCGCCGCGGACCTGTGCGGGAAG  
GCGTCCGGCACCTCTGACGCGCGCGCACCGGGGAACCCGGGTGCGACGGCGGGCCGGCGCCAGGCC  
GCGACGACCGTGCGCGCCACCCCGGTGAGCACACAGCGCCAGCACGATCCGGTTGACCGCCAGGCCG  
CCCGCTTGATGTGAGCAGGCCGCCCCGAGATAGCCGAGCCCGACACCGGCGACCGCGCCGATCACC  
GCGCCGATGCTCCCGGCGATCGCCGCTCCGCCAGCACACCCGGCTACCTGGCGGCGGTGCCCGCG  
ACCAGCCGACGAACGCGACCTGGCGGACCCGCTGGGAGACGATCACCTGGAAGGTGTTGCGGATGACC  
AGGATCGACGCGAGCAGCGCCACCGCGGCGAACGCCAGCATCAGCACGACCGAGCTGGGTGTTGCCACCG  
GCGAACCGGCCGGCGGCCGCGTCCGGCGGCCGCGACGCGTCCGTGCGGTCGCCCCGGCGGCGAGGGCC  
CGGCCGACCGCGTTCGACGGTCTCGGCCAGCCGGTCCCGGTCCGTGACGGTGAGCAGCGCGGCCGGCGG

GTGTCCCCGGCGAAGAACGACGCGGCGGCGTAGAACCGGTAGTCCGAGCCGGTCAGCGGGCGGAAGCCG  
AGATCGGCCCGAGCCGACCACGGTCACCGGCACCGGGGCGGCCGTGCCCTGCCGGAAGTCGAGGTGGGCG  
CCGACCCCGACGCCGAGGTGGTGAGGGTCCGCCGGTCGGCGACGACCTGCCCGGCCGCGCTCGGCCAG  
GTGCCCTCGTCGACGGTGAACCAGCGCACCGACGCGGTGCGCGGGATGCTCTGCACGTTGGCCGAGCCG  
CGCCGGTCGCGGCCGAAGACGCTGACCGTCCGTGCGTACTGCGGGTCGACGCTCCGCACGCCCCGCACC  
CCGGCGGCCCGCTGGTACCCTGCGGGTCTGTCACCGTGTCTGTCGGCGTCGAGCACGATGTCCGCCGTG  
GTCAACGGGGCCGCGCGGGTGAGCCGACAGCCCTCGTCCGAGGTGCTCGCGAACGTGGCCGTGGCGGCC  
AGGAAGCCGGTGGCCAGCACGACCGCGGCGACGATCGCGAGCAGCCGGCCGGGGTGGCACCAGGACCTGC  
GACCAGGCCAGCGTGAAGATCATGACGTACCGTCACCGACGCCATGACCGACATGATCGACTCCAGTG  
TGGGCGCTCGGAGCTCGTCCCAGAGCCGCCCGTCGGCCATGATCAAGACCCGATCGGCGTACGTGCGGG  
CGGCAGCGTCATGCGTCACCATGATGATCGTCTGGCTGGCTGCCGGGCGGCGTTCTGCAGCCCGGCGA  
GCAGCGCCCGGCCGGTGGCGATGTCCAGCGCGCCGGTCGGCTCGTCCGGCAACACCACCGACGGCTCGG  
TGAGCAGCGCGCTGCCACCGCGACCCGCTGCTGCTGGCCACCGGACAGCTCGGCGGGCCGATGGCCGA  
GCCGGTTCGCCGATCTGCAGCGACGCGGCGATGCGCTGCAGCCGGTCGCGGTCCACCGGACGGCCGGCCA  
GCCGCAGCGGCAGCACGATGTTCTGCTCGGCGGTGAGCGTGGGCAGCAGGTTGAACGCCTGGAAGATGA  
AGCCGACCCGGTCCCGGCGCAGGTGCGGTGAGCGCGCGGTGCTCCAGCCCGGCGAGCGCGGTGCCGGCCA  
GGCTGACCTCGCCCTCGGTGCGGTGTGTCCAGCCCGGCGAGCAGGTGCATCAGCGTGGTCTTGCCGGAGC  
CGGAGGGGCCCATGATCGCGGTGAACTCGCCGGCCGCGAAGGAGGTGACACCCCGTCGACGGCGACCA  
CCGCGGCGTCCCGGTTCCGTACCGTTTACGCAGGTTGCGGCAACTGACCATGGAGCTGCTCTTCTGCTCT  
GCATCATCACGGTTGCGACGTTAGGGAAGTGCGGGGCCGCCACATCCGTCCGGGGCATCGCCGCGACC  
GATACCAAGGTATCGACCTGGCGTGCGGATGGTGCCGCGAGCGCTCGCGGTCCCTACTCTGAGAGCGTG  
GGAAGCTGGGACAGGAGAAGCTGACGGTCGATACGGTGATCGCCGGCGCCGCGGTGATGGTGTGCGCTG  
CTGCTCGGGTTGGCCGGTCTGGACGAGTGGTACTGGTTCGGCCGCGCTCTGCGTCCCCCTGGTGATCCGG  
CGCTCGGCTCCGGTGGTCTTCTCGCGCTGGTTCGCGGTGCTCTCCGGCATCCACATGATCTACTCCGGC  
AGCTTCGCGTTCCCGGTGACCTGGTTCGATCTGGTTCGCCGTGCACGCCGTGCGCGGTACCGGCCGGCC  
CGGGTCCGGCACCTGGGCCTGCTGCTCGGCGTGGCCGGCAGCCTGGTGGTGACCGCCCCGGGCGCTGCAC  
GACGGGCTGCCGTGCTCCGCGACGCTGCCCCGCCGCGCTGATCGTCCGCGGACCCCTGGCCGCTGGTTCG  
ACCGGCCTGATGCAACGCCGGCAGCGGGCCGACGTGATCGAGGCCGATCACCGCCGCCGGCTCGCCGAG  
CAGGACAGCGCGATGCGTGCCCGGCTCGCGGCGATCGAGGAGCGCACCCGGATCAGTCAGGAGATGCAC  
GACATCATCGCCACTCGCTGGCCTCGGTGATCGCCCAGGCCGAGGGCGGCGGGTTCGCCGCGCGCGCC  
GACGCGGTGCTCGCCGGCCCGCTGTTTCGACCGCATCGCGCAGATCGGCCGGGAGGCCCTCAACGACGTG  
AAGCGGCTGCTGAACTCGATCGACGGCGACACGCCGGACGACTTCGCGCAGGGCCTGCCCGACCTGCCG  
GGCTGCTGGCCGGGGTCTCCGCCGCCGGCCTCGACGTGACGTTTCGAGGTGGCCGGCCCGGAGCAGCCG  
CTCGCGTCCGGCATGGACCTGGCGGTGTACCGGGTGTACAGGAGTCACTGACCAACGTGCTCAAACAC  
GCCACGCAGCGGCAGGCCCGGCTCAGCCTGGTGTGGACGCCGGCGTGGCTCGAGGTGACGCTGACCAGC  
CCGTGACCTTCGCCGGCGCGCTCCGCGAGGGTTCGCGGGCTGTCCGGCATCCGGCAGCGGTGCTCGTTG  
TTCAACGGCGACTGCGAGATCGTGCCGGGCGAGACCTTCAGCGTGATCACCCGCTGGCCGCTCGCCCGT  
CCGGAGGTGGCCGTCCCATGACCGAGCCGAGATCGACGTGGTGATCGCCGACGATCAGGACCTCGTCC  
GGACGGGCTTCGCCCTGGTTCGTCGACTCGGCCCGGACATGCGGGTGGTTCGCGACCGCGGCGGACGGCG  
CCGAGGTGGTGGCGTGGCCGCGGAGTTCGCCCCGACGTGGTGTGATGGACATCCGGATGCCCGGG  
TCGACGGCATCACCGCGGCCCGGGCGATCCTGGAGGGCAACGCTCAGCCGCCGAAGATCGTCGCGCTGA  
CCACCTACGACAACGATGAGTACGCCAGCCGATCCTGGCCGCCGGCGCCAGTGGGTACCTGTTGAAGG  
ACACCACCGCCGAGGGCCTGACCGCGGCGATCCGGACGGTGCACCGCGGCGGCTCGGTGCTGGCCCCGT  
CCACCACCCACCGCTGGTGACCGCGCACCGTCAGCATCCGGCACGCCCCGTCCGCGCTGCTGGATTCTT  
TCACCACCCGGGAGCGCGAGGTCTTCGACCTGATCGTGGCCGGCGCCAGCAACCGGGAGATCGCCGACC  
GCCTGAACCTGGCGGAGGTACGATCAAGACCCACGTCCGGCCGGGTGCTGGCCAAGATCGGCGTCCGCG  
ACCGGGTGAACGTCGTGATCTGGGCCTACCGGAACGGCGCCGGGCGGAGCTGAGCAGGGGCCAGGCCGC  
GCTCGCACTCGGCCGAATCCCGAAAAGCGGCTAGGGGAAGCGGATCTCGGCGGCCAGCGGCAGGTGGTC  
ACTGCCGGTCCGGGGGAGGGCCGTGACCCGCGTGACCGTCATCGATCGCGCCAGCACCTGGTTCGATCCG  
GGCGACCGGCGTCCGCGCCGGCCAGGTGAACGCGAAGTCCGCGGGCGAATCGATCATCACTTGCCGGAT  
CGGCCGCAGCCCGCGGTGCTCCACCGACGTGTTTCAGGTCCCCGATCACGACCAGCCGCGGAACCGGGTC  
GGCGGCGAGCAGCGCGCCGAGCTTCGGGGCGCTCTCGTCCCGCCGTGCCGAGGTGAGTCCCGCCGCGGT  
GACCCGCACCGACGGCAGATGCGCGACGTACACCGCGGTGTCCCGCCCGGCGTCCGGGCCACCGCCG  
CAGCCCCCGGTTCAGTCTTCGCCCAGGTGATGCGGCCGGATGTCGATCGCCTCCGCGCCGGTCAGCGG  
ATAGCGCGACACAGCGCCACGGTCCCTGCACGGTGTGGAACGGAAGCTCCGCGTCGAGCACACCCCG  
GTAGGCGGCCACCGCCTCCGGCAGCACCTCTCCAGCCCGACCGGTCCGGATGTGCCGCGAGCAGCGC

CCGGGCGGTCCCCGCCGGGTCCGGGTTCTCGTCGCTCACGTTGTGCTGCACGACGATCAGATCCGGCGT  
GCCGGTGTCCCGGTTCGACCACGTAGCCGCCGAAATGGATCAGCCACGCGCCAGCGGCAGCAGCACCGC  
GGCGAGGCCGGTTCAGCGATCGCCGCATCAGCGCCAGCAGGAGCAGGACCGGCACGGCCAGCCCGAACCA  
CGGCAAAAACGCCTCGATCAGGCTGCCCCGCTTGCCGACCGCGTTGGGGACCAGACGGTGACCGAGAAT  
GACCGCCCCGGCCAGAACAGCGGTCCACAAGATCGGCATCGGGAGATGTTATCTTATTTCGCATGCGAA  
ATAATGAGACGGTACGGATCCCGGTAGCGACCGGCGGCGCCGTCACGGCCACGCTGTTGCGCCCCGAGT  
CCGCGCGTGCCGTGCTGGTCTGTCACCCGCGGACCGCGACTCCGCGAGGGCTTCTACGCGTCTGTTGCGCA  
CCTACCTGGCGGAGAACGGGATCGCCACGGTCACCTACGACTACCGTGGCACCGGCCGTTCCGGCTCCC  
CGCGCGACACCGTGACCTGGGCATGCGCGACTGGATCGGCGCCGACGCCCCGGCCGTCGCGGCGTGGG  
CCGCGGACCGCTTCCCGGGCCTGCCCCGGCTCGCCGCCGGTTCACAGCCTCGGCGGGCACGTGATCGCGC  
TCGGCGCGGCGCGGCCGGATCTGGCCGCTCGGTGATCGTCGCTCGCACATCGCCGCGTTCGCGACCA  
TCCCGAGCCGGCTGGAGCGGTTCCGCGTGCGAATCATGCTTCACATCCTGGGTCCGGCCCTCGGGCGGC  
TGCTGGGCTACGTGCCGGCCAGAACTCTGGGCTCGGCGAGGACCTGCCGGCCGCCGCGATGTTGGAGT  
GGGGCGGCTGGGCGCGCAGGGACAACCTACTTCTCGACGACCCGTCGATGCGCGCCGCCGAGCGCGCCG  
CCACCCTGACCGGCCCGGTCCTGGCCGTCGGCACACGAGACGACCCGTGGTCCACGCCGCGCCAGATGG  
ACGCCCTCACCGTGACCTGACCAGCGCATCCGTGGAGCGGCGAACCTACTCACCGGCCGCCGCCGGGG  
TCCCGGTGATCGGGCACACGGCCTGTTCCGGCGCGCCGTGCGCGACACCGTCTGGCCGGAACCTGCTGG  
CCTGGTTGCACGCCCACTCGGAGAAGGCGTCAAGATGACCGCGTGCGCCTGCCTCGCCTGATCTTCTTG  
CTCTTCAACGCCGACCGCGCGGTCCGGCGCTGGATCGACGCCCCGTTCCGGTGACACCGGGATCGGCGCG  
TCCGGCGCGCGCGTGCTGTTCTACCTCGCCGGTCACGAGAACGCCCTCATCGGCGACGTGACCGCGGCC  
CTGGGTGCGTCACCATCCGGCATGAGCGGCTGGTGAACCGCTGGAGCGTGGCGGTTGCCTGACCCGC  
TCGCGAGACCCGGCCGACGCCCGCGCGGTGCGGCTCGCCCTGACCCCGCGCGGACACACAGGTGGTATC  
CACGCCCGCGGGCTGGTTCGACGACCTGAACGAGCAGCTCACGGCCGGTTTCGACGACGCCGAGATAGCG  
GTCGTCCAACGCTGGTTCGAGCAGCTGACCCGGGTGTCCGTGCAACGTGAGGAGCGACTCGGTTAGCGT  
CCGGTGATCAGCTAAACCTCCGGGACGGGCCGCCGAAGTGAGGTGTCGCAAGGCACACCGCTTCCGAT  
CCGCCCTTGAGGAGACCATGGCTGACTCCGTCTCTTCGACAGATCCCCGTCGTCCGCGCCATCGGC  
CGAGGCACACCTCGCTGAGCGCCTTCCACGACGCGCTGGTGACCATGGAGTGGGTTTCTACAACCTT  
GTCCGGCTCTCCAGCGTCATCCCGCCGGGCACCGCCGTGACCCCGAGCGGCAAGGCTCCGGTCCCGGTC  
GGGGCGTGGGGCGACAAGCTGTACTGCGTCTACGCCGAGCAGCACGCCAGCCAGCCGGGCGAGGAGGCG  
TGGGCGGCATCGGTGGGTGCAGCGCCGCGACGGCCAGGGCGGCTGTTCTGTCGAGCACGAGGGCAC  
AGCGAGTCGTTCTGTCGCGAGGCGATCAAGGCCAGCCTCCGCGACCTGGTCAAGGGCCACGAGGACGAC  
TTCGACGGCCCCGACTTCGTCTGTCACGGCGTGGTCTCCGACGGCGAGCCGGTCTGCGCGATGGTGCTC  
GCCCCCTACGAGACCGCCCCGTGGCGCGGCGTCCGTGCCACCGACCCGCCCGGCATGAACTGACCTGTC  
ACCACGAGAGCCTCGCCTCGCACGTGCGGGCGAGGCTTTCGGCATTCGCTTTTCACTTCGACAACCCCC  
TCCGTACCGAAGAAATCGAAGATCAAATTTCTTATTGCTATCTCCGCTGTGGTCCAGATCGTCGCTCC  
CGCGGGGACCCCTCCGGTCTGAGCAGGGCCGTAACCGGCCAGAACGCTCAGCCCGAAAGGGCGACCTC  
CGTGGCGGGGCACGGTCAGGGCAAAATTCGCTCGCCGGTTCGCCAGGATTCGCCGCCGGTTCGTGCGGG  
TGGGTGCGGCGGTCCGCCGAGGATGGGTTGCCGCTACGCTCGGCTAGCTTTCGACACCGGTAACG  
TGTAATTTCCCTTCGGTCCACCTACGCCGTTGTGATCTCCTTATGGTGGGCTCGCTTATCCTCGTCCT  
GTGCATCTGGGCCGATCGGCCGCGATGTTCTTCGGCTGTCTCTCCGGCCCGGACGCCGGGCACCCG  
TGCTGGCGTTTCGCCCCGATCCTGACCGCCGGTTTCGCGATTCGCTCGCCGCCGCTACGGTTCGCA  
GGACTGGTCCACCTTCGGGTCTGCTGGTTCGCCGGCCTGGCCGCTTCGCGGTCTGCTCTACCTGCT  
CCGTCCCGTCCGCTCTGATCCATACTCGGTCTGTTGTGACGACCCGGATGCCCGCCTCGAGATCGAAT  
ACTGCACACAGTGCCGCTGGCTGCTCCGCGCCGCGTGGCTGGCGCAGGAGCTGCTCACCACATTTCCCC  
GCGATCTCGGCGAGGTTGCCCTGGTCCCCGGGATCGGCGGCGTCTTCGAGGTCCGCCTCGACGGCGAGA  
TCCTCTGGAACCGCAAGCCGACGGCTTCCCCGACCTCGCCAGCTCAAGCGCCTGGTCCGCGACCGGG  
TCGCCCCGGGCGGACCTCGGCCACTCCGACCGCGCGCCACCGACTCCTAGCCACACGTCCTCCTC  
CACCGTCCGGCCCGCGGACCGCTCTACGCTACAGGATCCGGTTCGATCGTCTCGCGGGGAGCAGCGT  
GCGGGCCCGCCACCGCAGACGGCGACACCGCGCGGCCGCGACATAGACATCGCGCCGCTGGTGGTA  
CGCCCCGGTACCGGACAGCGCCACCATACCGGCCGCCATGTCACCCGGCAGCTCCGCCACCGCCAC  
CGTCGCGTCCCCGACGGAACCGTGATCGACCGGCCCGGCGCGGCGACGCCCGGCGATCAGCGCTGC  
CGTGTGACGGCCGGCGAGTCCGCCCCGGGCGAGGAGTCCGGGACGTCCCGTCCAGCTCGATCACGCC  
GTCCCCGGCGGCCACCACTCGGTGCACCGTGATCCCGGCACGACCCAGCAGCGCACGCCGGGCGACAC  
GTGCACCTCGGGCGGCTCAATGCCGTACCCCTCGGAAGTGACCTGCGCGACCGCCCGCAACCTGGCAGC  
GAAAATGCCACCGGGAGATCGGCGGCATGGCCGCCGCGAGGTTGAGCACGGGCACCGTCTCGCGAG  
TCGCGCGACGAACCCGATCGCCTCACGCAGGCAACTCTCGTAGGTGCCGAAGCGGTTACGCCGGTGCCC

GAGCGAGCAGTCCAGCCCGGCCAGCACCAGCCTCCGCGACCGGGTCACCGTGGCGACGGCGGGGAGCGC  
GGCCGAGCTGTTTACGCCGACCCCGTAGCCGCGCTGTGCGCTGCCCGCCGGACCCTCAGCAGGACGCG  
CTGCCCCGGCCGGGACCGCGCCGCGACCACTTCCGCCTCGGAGGCCGAGCCGATCACGACGGCGGCGCC  
GCAGGACAGTGCCGCCTCCAGATCCGCCACGGAATTCCTCGCTCCCGAACAGCGCCAGCGACTCCGGGCG  
GATCCCCGCGTCGAGCGCCGTCCGCAGTTCCGCGGCCGAGCGGCAGTAGCAGCCAGCCCGTCCCGCGC  
GATCCACCGCGCGGCACCCCTGAGCAGGCCGCCCTTGGCGCTGCAGCACACCGCGCCCGGCCGAAACGC  
CGCGACGTACTCCGCGCAGCGACTGTGCACATCGGTCTCGTCGATCACGTGCACCGGAGTGCCGTGGGC  
CGCGGCATCCGGGTACCCGGCACCCCGCCGACGGTGAGGTACCCGGGCTCGGTCCAGCGTGCGGTGAG  
CGGCCAGTTCGCGGGGTGAGGCGCGGGCGGAGCGACGCGCCAGTGAGGGCAGAATCTCGGACAACGT  
CATACCGCGAGCCTGCTCGCCGGCCCGTGGCCGAGAAGATCCGGATACGCCACGTTGACGATCGACGTG  
CCGTTGTTGACGTGTGCGCTACAGGTCCGGGCTGTGCGCGGGAGCCATGATCCGCAGCGCGTTCCGGTC  
GATCGCGATCCGCATCGGCGTCTGCCCCGAGGCTCGCCGTGACCTCGACGGCCACCGGCCGATCCGT  
CTCCAGCCAGAGTTCCCGGACCGCGATGAAAGGGCGTTCATGCAGGGTACGGCGATGCCCGGTGGCCGC  
GTTCCGGGGCGTCTCGCGCAGCAGCTCCCGCCGGCTCGCCCCGCCACCGGATAGGCGACCAGCAGCCG  
GTCGTGCGCGTGCAGCGTCCGCGGTGATCGGCCGGCGCGCGTGGAAACCGCCGTTCCGCGACGTACACCTG  
ATGGGTGTGGAACCGCAGCTCCCGCCCCCTCGGCCCGGATCACCGCGCGGACCGGCCGGTGCCGGGCGAG  
CAGCCCGAGCGCGGTTCATCGGATACGCCAGCCGCCCCACGGCCCGTTTCAACCGCGGCGGCGCGCTGAT  
CATACCTCGCCGGAGAGCCCGATCCCGACGTGTTGGTGAACGGCACGTCCCGGGCGACGCCAGGTG  
CACGTGATGACTTTGCCGTCCACCAGGGTCCGCGATGGCGGCTTCGAGGTCCGGCTCGATCCGCACGGT  
CCGGGCGAAGTTGTTGGTGGTGCCGAGCGGCAGCACCCCGAGCGCCACGTCCCGGTGTGCCAGCATCCG  
CCCCCGGGTGTGATCGTCCCGTCCCCGCCCCCGCGATCAGCAGGTCCGGCTCTTTCTGAGCGCTTC  
GGAGATCAGCCCGTCCAGCCCGCGGACTGCTCCAGCGCATACGTCCCGAGCAGCTCGAACC CGGCTC  
GACAAGCCGCGCCGCGCCTCTTCGTAGAGGAGCGCCCCCGCGGGACCGCGTGTGACAACAAGAAC  
CGCCCGCGCCCTCGCCGATGTGAGCACTGTGCTCCCGCTTACTACGCACCCGCCGACCTATCCGCC  
CCAACCACCAACGCCACGCCTCGTTCCCGCCCGCCGGTCCCGAGCCGTGATCACCAGCCGCAAGG  
CCTCAGCTGGCGTTGACGCTGATTTCCGGCCCGGTTTTCGCGCCGTAGGCGCCAGCCGCGAGGTGTCCGC  
TGGCGTTTCAAGGCTGGATTCGGCTCGGTTTTCGAGCCGTGGGTGCCAGCTGGGGGCTGCGGCCGGAAGG  
GGTGGATGTGGGGATGCCCCGGGGCCGTAGGCCTCGGGGCACCGAAAAAGCGAACAGCTCAGCTGTTG  
GCGATGAACGCCGGATGCGACAGCAGGAACGTGTGATCTGGTCCCTTCTTGATCGACTTCAGCAGGTCC  
ATGCTGTGCTGCTCAGGAGCTCGACGCTGCCGGAGCCCGGCACGTTCCTCCGAGTTTCACTTGCAGGCG  
TTGGTCTTGATGGTGAGGAGCTTGTCCGGCTTGAGGCTGCGCATCGCGAGCGCCAGTCGGCCAGGTGCG  
ATGCCACCGTCTGTCGACCGTTCATCGCCTTGCCGAACGCGCTGAGCAGCTTCGGCAGCTTGGTTCGGCGAG  
TCGAGGCGCTCCTTCAGCGCCTGGTTGATGATCGCCTTGAAGAACTGCTGCTGGTGCCGCTGGCGACCG  
TAGTCGAGCGAGTTGTGCGCGAGCAGGTACGCTGGCGGACGAAGTCGAGCGCCTGGCCGGGGGTGAAG  
CAGTGGTCCGCCCTTGGTGTACGTGTTCCGGCTCACGCCCGAGATCTTGCTCTTCAGCGTGCCGTCCGGG  
TTGATCACGAACGGCTTGGCGGTCTTGCCGTTCTGGTCCCTTGCCCAGGTGGATCGACTTGGTTCGTGGTG  
TCCACGTACATGCAGACCTTGCCGAGCAGTTCACCACGTGCGGGAAGCCCTGGAAGTCGATGATCGCC  
CCGGCGTCCGGCGTGATGCCGGTCACTTCTTGACGGTCATGGTGAGCAACTCGAAGCCGTGCTGCAGC  
GCCTCGTTGCCCTTGAGCCCGCGGGTGCCGAACGCGAACGCGGCGTTGATCTTCGTCTTGCCGCCGGCC  
CACTTCTGCTTCCCGTTGTGCTACGCCGGGATGTAGACGTAGCTGTCCCGGGGAGCGAGATCATGTAA  
CCGCTGCTGTGGTCTTGTGATGTGCAGCAGGATGATCGAGTCCGACCGAGGGGCTCGCCGTGGTTC  
TGCGTGGGCGCTGGTTCGATGCCGACGAGCAGGAGATTCTTCGCCCCGTGAGGTTTCGCTTCTTCTTC  
TCTTCGGCGGGTTTCGCGGAGCCGAGCAGCGACTCTGCCCGACCGAGGAGGTGGCCGCGGCCACCGTC  
GCGTTCAGCCCGATCGCGCCACCGCCACCGCCGACCAGCAGCAGTGAGCCGAAGACCACGGTCCAGAAA  
GCCCAGCGGGGTGTGCGGCGTGGCGGCTTGCTGGTCTACGCAAGGGGTCTCCTCTAATCGGGGCGGTG  
GTTCCGCCCCAGGACAATCTAGGACGGTCGTTCCACCCATGAAGACGGGCGCCGCGAGCGAAAAGATTG  
CCTGCGGAACATTGAATTTTCTGAGCGTCAGGTAGCTACAAGCTGAACGTGCGCGCTCGTCCCCGTGG  
CGTCGCGGAACGTTGCTTCCCTGTTACTCCCCAACGCGACGTCTGGTGTAAACCATCGCACCACCAAGG  
CGACCTAGGTTCTGCACGGCCCCGGGCCCCGAGGTTCCGACAAAGTTGCCCAAAGCGAGTCGCCATGC  
CCTCAGAGCCGGATGTGAGTGTGTCATTCGACGTGTAACCGGCCGGAGTTGGCCGTCCGCGCGGTGCG  
GCAGCGCCCTCGGGCAGACGCACCGGAACCTCGAGGTGATCGTCTGTCGTCGACGGTCCCGACGAGGCGA  
CCGTGACGGCGCTGGGCGAGGTCCGTGACCCGCGACTCAGCGTGATCGTCTGCTGCCGGAGCGTGGCAAGG  
CCCCGAACGCGCGGAACACTGGCGCCCCGGGCGCCCGCGGCGCTGGACGGCGATGCTCGACGACGACG  
ACGAGTGGCTGCCACCAAGATCGAACGGCAGCTGGAGACGGCCGCGGCGGCGACCGTGGAGCGCCCGG  
TGGTGGCCTGCCGGATGATCAGCCGGACGCCGCGGGCCGACACGATCATGCCGCGCCGGCTGCCGAGC  
CGGGTGAGCCGATCAGTGAATACCTGCTGGTCCGCCGCGGTCTGTTCTACGGCGACGGCTTCGTGCGA

CGTCCTGCATCATGGCGCCGACCGAGCTGTGGCGGAAGGTGCCGTTACCGTCGGCCTGCGCCGCGCGC  
AGGAGCTGGACTGGACGCTGCGGGCGATGCGCGAGCCGGGCACCGCGCTGATCTACGCCGAGGAGCCGC  
TGGTCTCTGGCACCAGGACGAGAACCGTGACCGGATCAGCCTGCAGAACCGTGGCGCGAGCAACTGG  
AGTGGCTGCGCGGCAACCGGGAGCTGTTACGCCCGCTGCCCTACGCCCGCTTACGCTGAGTGTGCTGA  
GCTCGATGGCCGCGCCGACCCGGGACACCGGGCTCTTCCGTGAGCTGCTCGCCGAGGCCCGCACGCACG  
GCGACCCGGGCACCGTCGACTACCTTACGCACATGCAGATCTGGGCCCTCCCGCCGAGCGTCCGGCACC  
GCCTGCGTGACGTCGTGGTGGGCCGCGGCAAGACGAGCAGCAATGCCGGCTGAGCGCCGGGTGCGGATC  
TGGCGCAGTTCGATGCTGCCC GGCTCCGAGACGTTCTGTCGCAATCAGGCCGACGCGCTGACCCGCTGG  
ACCCCGGCCTATGTGCGCGCGGTCCGGCACGAGTCGGTGCTGTCCCGCCCCGACGACGTGATCGCCTTC  
CCGGGCGGCAAGGGGTTCTGCGACTGCGGCTGACCGGGGCGTCTCCCCAGCTGCAGAAGACGATTTCC  
GCCGTACGGCCGAATCTGGTTTCATGCCCATTTCTGGTGGTGACGGATGGCTGGTGAGCCACTCCGCCAG  
CAGCTGGGGGTGCCGCTGGCCGTACCGTGCACGGGCACGACGTGACCCGTCAGCCGTCCAGCCCGGGC  
GCGAAGGGCGTGCGCTACCGCCGAACCTGCAGACCGTTTTTCACGCGGGCGTCGCTGGTCATCGCGGTC  
TCGGAGGTGATCCGCGGCCAGGCGATCAGGTGGGGCGCCGACCCGGCGAAGGTGAAGGTGCACTACACC  
GGCATCGCGGTTCCCCCGGAGCAGCCGGAGGAGGTGCCGAAGCGGTGGGACGTGGTGTTCATCGGCCGT  
TTCGTGCGCAAGAAGGGCGTCGACGACCTGCTCACCGCGCTCGCCGCGGTGAGTCGCGCCCCGCGCGC  
CTGCTCATCGGCGACGGCGAGCTGATGACGGCGATGCGTGCCCGTGGCGAGCAGCTGGGCGTGGACGTC  
ACGTTTCGCCGGCAGCCGGACCCCGAGCAGGTGCGGCGCCACCTGCTGGAGTCGCGGGCTGCTGGCCTGC  
CCGTGCAAGACCGCGCCGGACGGGGACACCGAGGGCCTGCCGACCACGATCCTGGAGGCGGCCGCGCTC  
GGCCTGCCGGTGGTCGAACCCGGCACAGCGGCATCCCGGAGGCCGTGATCGACGGTGAGACCGGCCTG  
CTCAGCCCGGAGGCGGACCCGGCGGCGCTGGCGGTGTCGCTGACCCGGCTGCTCGGTGACGAGGATCTC  
CAGCGCCGGCTCGGTGCGCGGGCGCGGCGGCACGTCACGGCACACTTCGATCTCGTGAGCAGACCAGG  
CGGCTGGAGGACCTGTATGACGAGGTGCTGGCGGGCGCTAGGGTCTAGGCCTGCCTGTCCGGCTTCTTG  
GGGAAATCATGATCTGCACGCACTGCGGCTCACCGGCGGCGCCGACCGTCGGCCCCGTGCCCGGGTGC  
GGCGTCCGGTCTCCGCCCGGGCGGGGTGTTCCCGGACCCGCTGGCCGCGCCGGCCGAGCGGTGCGTC  
CCCAGCTCGGCGTACAGCGTGCCGGTTCGATCCGTACACCGGCCCGACACCGGTGACCCGTTCCGCCGC  
GACTCGTTCCACACCCCGCCGCCGGCGTCCCCGATGCCGCCGCGCGGTGACCGAGCCGATCACGCCGACG  
CCGCCCGCGCCGGCTACGCCCGCCCCCGCAGTACAGCCCGCCGCGCTATGCGGGGGCGCCCGGTGAG  
CCCTATCCGGGTGAGCCTTACCCAGGTGAGCCCTACCCGGGACAGCCCTACCCCGGGCAGCAGCCGTAT  
CCCGGCCAGCAGCCCTACCCCGGATACGGCCAGTGAACAGCCAGGCCCTGACCATCGTCGCCTACGTC  
CTCGGCGGCTTCGGTCTCCTGACCATGACCCCGCTCATCGGCGTCTCGGCGCTGGTCTGGCGAATTC  
GCGAAGCGCCGTACAGAGCGCAATGCCTCGATCGCGGTCAAGATCGTCGCCGGCCTGGTGATCGCCGCT  
TTGGTGCTGAGCATCGTGACCGCATCGTTTAAAGGCCTGCCCTGCCGATCACTCAGGGCGGCGACTTCT  
GGGCGGCGGGGGTGCTCCCCCTCGCCGGAAGGATGTCCGGTTCCCGGGATGGTCAGTACTGTGGGCTCA  
TGGGCGAACATTTTCGATCTTGTCGTGCTGGGCGCCGGTCCGGGTGGATATGTGCGGGCGATCCGCGGCG  
CTCAACTGGGCCTGACCACCGCGATCGTCGAGGACAAGTACTGGGGCGGCGTCTGCCTCAACGTGGGCT  
GCATCCCGTCGAAAGCGTGCTGCGCAACGCGGAGCTGGCGCACATCTTCCACCACAGGCGCAGACCT  
TCGGCATCGAGGGGAAGGTACCTTCGACTTCGCCGTGCCCCACCAGCGCAGCCGACGCTCGCCGACG  
GCCGCGTCAAGGGCGTGAATTCCTGATGAAGAAGACGGGATCACCAGATCCAGGGGCGTGCGGAGT  
TCACCGACGCGCACACGCTGCGGGTGGCGACCGGACGGTCACGTTGACAACTGCATCCTGGCGACCG  
GCGCGAGCACCCGGATGATTCGCCGACGAGCGTCTCGAAGCGGGTCTGACGTACGAGGAGCAGATCC  
TCGACCCCGACCTGCCGGACAGCATCGTGATCGTCGGCGCCGGCGGATCGGCGTCTGAGTTCTGCGTATG  
TGCTGCGCAACTACGGCGTCGACGTGACCATCGTGAGTTCTTCGACCGGATGCTGCCGCTGGAGGACG  
AGGAGGTCTCCAAGGAGCTGCTCCGGCAGTACCGCAAGCTCGGCGTCGACGTGCGGGTGGCACCCGGG  
TGGAGGGCATCGAGGAGGGCGCCGACTCGGTCCGCGTCACCGTCTCCAAGAACGGGAAGACCGAGGTCC  
TCGAGGCCGACAAGGTGATGCAGGCGATCGGCTTCAAGCCCAACGTGGAGGGCTACGGGCTGGAGACCA  
CCGGCGTCACGGTGTCCGACCGCGGCGCGGTGAGATCGACGACTTCTGCCGTACGAACGTGCCCGGCA  
TCTACGCCATCGGCGACGTACCCGGAAGCTGATGCTGGCGCACGCCGCCGAGGCGATGGGCATCGTGG  
CGGCCGAGACGATCGCCGGCGCCGAGACGATGGCCCTCGACTACCGGATGATCCCGCGCGGACTTTCT  
GCCAGCCGACGGTCCGACGTTCTGGGTGGACCGAGGCGCAGGCCCGCGAGCAGGGCTTCGACGTCAAGG  
TGGCCAAGTTCCCGTTACCGCGAAGCGCAAGGCGCACGGCCTCGGCGACGCGACCGGGTTCGTGAAGA  
TCCTCAGTGACGCGAAGTACGGGGAGCTGCTCGGCGCGCACCTGATCGGGCCGACGTGACCGAGCTGC  
TACCCGAGCTGACCTGGCCAGCAGTGGGACCTGACCGTCCACGAGGTGGGGCGAACGTGCACGCGC  
ATCCGACGCTCGCCGAGGCGGTCAAGGAGGCGATCCACGGCCTGGCCGGCCACATGATCAACTTCTGAG  
TCCGCCGCTCGGCCGACGGGTCCGCCGCTCGGCCGACGGGTCCGCCGCTCGGCCGACGGGTCCGCTGTC  
CGGCCGACGAGTCCGCCCGCCGGCTGACGCCCGGGGAGTAGCCGGAGCCGCCCGTACCGCGCCTGCGG

TAGCCCGAATACCTACGCGGGGATGACGACTCCGCCCCGCCGGTCGGGAACGCTGAGCCTTGTGACCCT  
 GACCGTCGAGCCTCCGATCGCCCCGGCGCCGCCGCCGCCCGGGGCGATCCCGGCGGCGCAGGCTGGG  
 CTATCTGGCGTTTCGTGCTGGTTCGCGGTGGTGGCGGTGGTACGCTGCGCGACCGGCTGCCGGATCCGGG  
 GGAGTTCCTCGACGCGCTGCGAGCGGCCGACTGGCGGTGGGCGGCGCTCGCGGTGGGGGCCGGGGTGCT  
 GTCCAGATCGCGTACGCCGAGCAGCAGCGCCGGCTTCTCGCCGCGTTCGGAGTGCGCGTGCCGGCCCCG  
 GCGGGCGATCGCGATGACGTACGTCCGGTCGGCGCTGAGCATGGCGCTGCCGGCCGGGTGCGCGGCCTC  
 CGCGGCGTACGCCCTTCAGGTCTATCGTCGCCACGGCGCCACCGCGGCGATCTCCGCGACGGCGACCCT  
 GATCTCGACGGTCGTGACCGTGATGTGCTGGGCTGCTCTACGCCGCGACCTGGTCTGCTGACCGCCAC  
 CGTCGTGGCCGGCCTGGCCGTTCTCCTGCTGTGGATCTACCGGACCGTGCGGGGGCCCGGTCCCCGCGCG  
 TGCCGGCGTGCCGCGCCGTCTGAGGGTCCGCCCGATCGCCCGCCTGCTCCAGCGGCCCGCGGTGGCCCA  
 GCGCTCCGGGGTGACGGTCCGTCCCGGCCCGGACGTGGCTCACGGTGACCCTGGCCGGCGTGATCAA  
 CTGGCTGCTGGACATGGCCTGCCTGCTCCTCGCGGCCGACGCGCTGCACGCCGGGCTCGGCTGGAGCCG  
 GCTCGCGCTGATCTACCTGGCCGTCCAGGTGGTCCGGCAGATCCCGCTCACCCCCGGCGGCATCGGCCT  
 GATCGAGACCAGCATGCTCGCCGGCCTGATCGCCGCGGGCGCCCCGAGGTACCGCCGCGGGGATCGT  
 CCTGATCTACCGGCTGATCTCGTTCTGGCTGATCCTGCCAGCGGGCTGGCCGCCACCTGACCCTGCG  
 CCGGGGGACCGTGCCGCCGGTGACTCCGGGCTGACCGCCGGGGCTCAGGCGCGGTGCAACTGCTGGAGG  
 ACGGCGCCGAGGCCGAGCGTCAGGGCCGGGGCCGGGCAGTCGTGAGTCCAGCCCGTACCGCCGGCCGGGG  
 TTCTCCCGATCGACAGGTCTGCCAGAGCCCGGTACCGGATCGCCAGCCCCGGCAGGTCTGCCCCG  
 GCCGCCCGGAGCAGCTCGGCCAGGTGTTCTGCTGGCGCAGCCGCTGCTCCTCGAAGATCTGGCTCCAC  
 TGCATCAGCGAGATCGACGCCTCCAGTGTGTCCGGCACCGTCTCGTCCAGCATCGGCAGCGGTGCGGCC  
 GCGGCCACGCCGCGAGCTGCGCCCCGGTCACCGTGCCGGGATCGTCTGCGGGCTGCGCGCCAGCGCG  
 TACGACAGCAGCGCCGTGGTCTCCACCAGCAGGTGCTGCACCGGGACGCGCGCCGGCGGGTCCAGCCAC  
 CAGTGGTCCGCCTCGAGCAGGGCCGCGGCGTGCCGGGTGCGCGCCCGCCACCAGCCGTCTGCGCGCCG  
 GCCAGGGCGTGGTGTAGGACCACAGGAAGGTACCCCTTCCGTGCCTACCAGCGGCACTTGACACGAGT  
 CAATCACGCGAACGCGTGCGGACGACCGCGATCGCTTCGACGGAGATCAGCAGCCCGTGCGGGAGAACC  
 ACGCCGGCGGGCGCGGACCGGGCCGGCGGCGGGTTCGGGCATGTGCCGCGCCAGCACCTGGTTGATCACC  
 GGGAAGTCTCGACGTGCGTGTAGAAGACGGTCAGCTTACCAGCCCGTCCAGCCCGTCCAGCCCGGAG  
 CCGGCCGCGCCAGCACCGCCCCGAGGTTGCACAGCGCCTGCTCGGTCTGCGCCTCGATGCCGTCCACC  
 GGCTTCCCGGTGCGGGGGTTCGATCCCGGGCATCCCGGAGCAGAACACGAACCCGCGGGCGACGATCGCC  
 TGGCTGTAGGGCCCCAGCGCCGTGGGGGGCCGCTCACTCGTCAACCGCGATGCGATCCATCGGGCGAGTC  
 TCCCGCACCGGGCATCGATGGTCTTTTCGGCTTTATGACACGGGTGGGAGCTTGACAGCCGTACATCGTT  
 TATCGATTATCAATATATCGAAAAACGATAAGGGGGTACGTCTGTGAAGACCGACTGGTTGACGGAGTTC  
 CACCAGGTGCTGCTGGTTCGCCGCGCTGGGCGCCGGGGCCGCCCGCTCCTCGGGACCGTCGGGCTCGTG  
 GTGCAGGACGAGGTGCTGCTCCCGGGCGGCACGTCCGGCTCCGGCGAGCGTGGTCAGCGACCCGAGCGTC  
 ACCAGCTCCTGCTCGCCGTGGCGGCCCGCGTCCCGACGTACCTGCTGGCGACCGCCATGCTGGTGCTG  
 CTGTACCGGTTGGTTCGGCGCCGCGCGTTCGCGGGGATC

**SEQ ID NO: 201 orf1 Secretion system protein 334aa**

VLRIQRRRLADELGPTLYTTGQSEDLARVRDALHDLLAREETPLSGADR  
 ARITREVTDEVLGHPGPIESLLRDPVTEVMVNGPYSVYVERFGRLEKVAA  
 EFDDEQHLRRIIDRICSRVGRVDEASPTVDARLPDGSRVNAVVPPIALD  
 GSTLTIRKFAAVPLTVGDLLQYGTLTRQAAEFLDACVRGRRTILVSGGTG  
 SGKTTILNLVSSFPADERIVTIEDAAELQLVQDHVVRLESRPPNSEGRG  
 TITTRDLVRNALRMRPDRIVGVEVRDAAALDMLQAMNTGHDGSLTTLHAN  
 TPRDALSRLETMVLMDLPRAIRAIQISSAVD

**SEQ ID NO: 202 orf2 Response regulator 358aa**

MSELESYLPYPETLLVVIGAVVDLDEALMFTAYQRLQRPVIGVVLIRHA  
 IEAGVVLRLQSGIRDVLAEEDEEGIRSATVRSVDLSKALSTTMRGTGDN  
 APYASVTVTFAGKGGCRSVVAVNLAVALAGARRVLLMDLDLQFGDVAI  
 MMKLSPERNIAGGLQAGRLDEPGLRSIVTTYRQSLDALLAPASPAEGEQ  
 VRREFVVELLDVARPLYDFIVVDTPSVVTDQVLAALDMSDWFIPIVTPDL  
 PALKSVRLTAEMFDLLDYPKDKRLLVFNRAGAQVGLSPSEVEVAAGMPFA  
 VQVPASRDVTVSVNHGEPIAVTDPLHPVSRRAIRELGDRIAGVGTPQRRR  
 GFLGRLFP

**SEQ ID NO:203    orf3 Hypothetical protein 266aa**

MRRRILILLALVLAGISGA AVL SYARSADRRALSGREGTWILVAGQHIP  
 SGTSGAEIRARGLTERLLVPTVTVPAGALT TWDSALDPLRLGGDLQPRQL  
 LMRTLFPASPAPSHTGRIPVPRRSLAVSVALNVAPQVAGNITPGDQVAV  
 YYTYQAKVLM AEDHETIPMTELLLPKARVITIGEAEPGQAPVTPSPSPTP  
 GPSASAGDSTA AVKEIQRYVVT LAVGDT DALRLVHAAQSGSLYLALLGPN  
 ATASTGPAIDTSWVVR

**SEQ ID NO:204    orf4 Hypothetical protein 432aa**

MRRLIHALFP PRGEHGVIT ALVAVLAGAGVLLGMAALVIDIGALYAEREQ  
 LQSGADAASWKVAQACAGTAGRDLTSATCTVAAQRDNAQRYADRNAKDLV  
 SDVQFCITT VSAAGVTTADAGCPSSWNTPVTC PAPPASAGPYRYVEVRTS  
 TRNSDNTSVV PPLFGRGLAGSAYHGAKMGACGRVAVGAPAVTDVLALGVS  
 RCDFLRLTGDYTRFFAPPPPTGPNPQTGVHPLLGLSDPTAGYIPISDGIL  
 ATSCPADAGETVAGYTWLGQPDGPPALPGLLPVSAPDASCELTGIPATDG  
 PADSWVGGFTIGPGNASAATA CLDRLNVLITSGQPVLVPIFDRQVAVIGT  
 LPSYYRIAGFAPLVLTGYESPVSGPVSPGS AVPSLVPGAQRALCAAQSCL  
 YGYFTRALVTDHVPTRFATSRYYGAMVIGRTG

**SEQ ID NO:205    orf5 Response regulator ATP-binding 136aa**

MWKATERQGPVLLVEDDEDLRELA AQMLEMRGFVVLVAKDPVSAIMTCR  
 VHSGAIDVLLTDLGLPGVSGGELARSASEVRPGMKIVYVSGVP EEIAIKK  
 GLIRAGSPFVAKPYTADRLAGMLRTVLAQGAESSAR

**SEQ ID NO: 206    orf6 ABC sugar transporter 363aa**

MFNYVSFVRPKTF AATFAAAALLLGSGACAKSEDSGDTVAAGPAPSAAQV  
 VQSASAGSATC ALDQYGASKLDLKTASVGFSQSEKEANPFRIAETKSIKD  
 EAAKLGITNLKTSNANSQFNKQIADVEQMIDAGVQLLVIAPLNSDGWDSV  
 FAKATAKHIPIITIDRKINATA CKDYLTFIGSDFAEQGKRAADALAKSLG  
 NKGEVAILLGAPGNVTTLR TSGFKDEIAKVAPDIKITFEQTGNFSREDG  
 QKVAEQLLQSKPNINGIYG ENDEMALGAILKAGGKAGDVKIVSIDGT  
 KGA VQGIVDGWVS AVIESNPRFGPLAFDTATKFFGGEPVGQDIVIQDRAY  
 DESNAKTDIGSAY

**SEQ ID NO:207    orf7 ABC sugar transporter ATP-binding 501aa**

MLLEVSGVSKTFPGVRALDGV SFTLNPGEVHALVGENGAGKSTLIKVL TG  
 VYQPD SGELRYRGE PARFATPLDAQRAGISTIIYQEVNLVPLMSVAHNLF  
 GREPRNRFGLLDEARMVAEATEILAGYGVRTDVRRLGLTLALGAQOMVAL  
 ARAVMVDARVVVMDEPTSSLEPREVETLFGVIRELHTAGIGIVVSHRLD  
 ELYRVCDAVTILRDGKL VHTGRMADLDRRTLVSIMLGREFGADFTSFSES  
 PQSTPEGEFVL RVSGLT SRPRLDDISFDVRPGEVVGLGGLLGAGRSETIK  
 AIGGAYPIDSGVIEVGGVRLGRPSTVR A VRAGVATQPEDRKAEGIVPGLS  
 IRDNIALAILPRMARFGLVSDKRIDSIVATYMSRLRIKASGPDQAVGDL S  
 GGNQQKVLLARLLATGPKVLL LDEPTRGIDVGAKEVQALIDELAKEGLG  
 VVLVSSDAEELVEGADRVVLRD GAVVGTLTGDRVTTEALMATIAEAADE  
 H

**SEQ ID NO:208    orf8 ABC transport permease 319aa**

MSTETLTRPRMTFNPAAWAARYGVYAAIVLLIVVNIAFTPYFLTSLNLRIQ  
 LIQAAPVVIVALGMALVIGTEGIDLSVGSVMALAAAFIPLYLGYGVTAAI  
 LVSLLAGVAVGLINGVLVAKAGLQPIVATLALFVGGRGLAVVISGGQLKD  
 VRNADLLYLGSDDLGV PVLVWIAALLVLVVA FVVRRTVFGRRL LAVGGN  
 RPAAELAGLPVKRVLIGVYVFC AVLASIAGLLSVARIQSSDASAVGLLIE  
 LSAITAVVVG GTPLTGGRVRVLGT VAGALLMQLVVATMIKHDLPSTTEM  
 VQAVIILVAVYVARERRTR

**SEQ ID NO:209      *orf9 ABC transporter permease protein 320aa***

MSIPVPAFRNGGFVQRQ GALAVLTVVAISLAAPFGFRSADNAGTILVAA  
 APPMLIALGMTFVIITGGIDLSVGS LYVLGGVVA AWASQWGVVAALAAPL  
 LLCGAIGVLNGILISRTGMAPFIVTLAALLGARGLMRSISDEGSTTYLVR  
 SDVFHELGTGSSLGVGLPVWLA AVLVGAGILVLNRTRFGHAVHAIGGSED  
 AAALMGLPVRRIKVWVYLLSGLLAGLAGAINAAKLGSGVTVLGSGMELDA  
 IAAVVIGGTLLTGSGS IAGTVAGVLLLGV IQNLINQVGNVNSNWQQVIS  
 GGFLAAVVVAQTTLVRARRS

**SEQ ID NO:210      *orf10 metalloproteinase 486aa***

MRTSAGTRVLTVGATVVLALAAAAAPAQAGSPSPSGSAGLGDRLYPLLGN  
 GGYDVLDDYDLRLRYPEKDPKQVVS GVDVTITAVAGQSLSRFDLDFGGASIG  
 RVSV DQGPARAARDGDELTVIPRRPLPRGRLFRVTVANFTAAPAALVATP  
 DGTVLAAQPGSAHLLFP GNDHPRDKATFTITLTV PAGWTGTANGTLVSTT  
 EHDGHVSSVYRESAPMATELVQTAVGDFVVERRPAAGGTPIRDVVPRLA  
 GTLLPAIAGEREQ LAWMEKQAGPYPFEDYGS LVIDDDLGYALETQTLSTLY  
 GAALFTGPETTRGPSMTH ELAQWFGDSVSPFSWSDVWLN EGHA TWYEML  
 WSEETGGFPQYTGLADREAFFKAVYAAGDIFRARYGPVAAPLDAATTWDFV  
 FNPVYAGGALVLYALRQKIG AATFQ RVERAWLT TYRGRSASTGDFITLA  
 SRVARQDLRPFLTSWLLGVTT PMPNHPDWVVT PGS

**SEQ ID NO:211      *orf11 putative StrR-like regulator*      286aa**

MDSPLNGEDPRHTEVL AGFGAELPPIVVHRATMRVIDGAHRLS AARLRG  
 DDRIRAVLFDGTEQEAYVLSKANVTHGLPLSAAERTRAAERIITMHPDW  
 SDRMIAASSGLGARTVGG LRRRRAASGES PAGLRSRAGRDSRVRPAGSTA  
 GRLKAVDYLQDRPDASLREIARHAGVSPSTARDVRDRLHRGEDPIPATQR  
 AAARPGNDSPPLRSLVQGLASDPSLRFS ESGRDLLRWLIAHAVQDGEWKG  
 LVD TIPAHSQAQALAKIARHCSREWREFADILEKDAA

**SEQ ID NO: 212      *orf12 lanA*      64aa**

MSAITVETTWKNTDLREDLTAHPAGLGFGELS FEDLREDRTIYAASSGWV  
 CTLTIECGTLVCAC

**SEQ ID NO:213      *orf13 lanM*      1046aa**

MSSFAIAASPASAYLHERSAGPGGDPVAEHERVESWRESAFLDDPVLDIR  
 LRELGLSRAEFGRLLTDGAYDAGSTALDWAGELAAVLATGTGAVTGLARS  
 TKLWAQGFDRLPFAGLIERFLAYYEPRVPRTAGTVRVSLLES LANRLTV  
 ATRTLLELNVARVHGRLTGATPGERYDHYDRVLLTDPDYLRSLFGEYPV  
 LGRAMVECGRRWASAMAELFQRLDADRPALHAAGLLPAGAGEVTALRPDL  
 GDPHNSGRAVAILTFRSGAQLVYKPRPVGP ERAYAETA AALNRHGLSLPL  
 TAVDVLDRGAYGWCELV RHEPCADRADLDRFYRRTGAVLATTLLLGAVDV  
 HMENVIAAGSSCMPIDLETLLQPGVPSGDATDAYTRALD LLLNQSVLAIGI  
 LPARAFGGREKRSVDVSAIGGGEAQTA PRPVP MVVEPFTDVARIEAVEAT  
 MLGAQNRPVLVGA EVRPEEHT EAVVAGFTEAYDLIVRHREDFADLLAGFG  
 DVEVRYLPRPTRYSMFLTESYHPDYLRDARDRDLDDKLWTAAGARPDL  
 IPIESEKRQLLAGDIPC FRALAGDRAIR TASAPVAPDFFDAPGIEVLG  
 RLRQFGPVHRAAQLRI IRESMGTMPAPGPIAGTPAPSSERRGGLDPREAA  
 TLGDRLVRELADEAILGADDAGWIGV SIEGLDQETFSYKPMATGLYDGA  
 GMALTYAYAARTLGDERYLDLTRRTVKLVSGYLRYLAEHRIVETVGAYSG  
 MAGLLYTLDHVAHATGDASLLGEIEAALPWLRECATREECPDLIAGLAGC  
 AVVALSLYRRHGIAGYREVAEICGRRLAGTAVDVEGAAGWAATRTGVILG  
 GFSGHSAGIAWALHELAAEFGRDLRELADRAVEFDRRLYVPAAGAWRDL  
 RP EMAGTDGYPALWCHGAAGIGLSRLLIHRIRPDERLAEEARA AVALVRR  
 HGFHGHNSLCHGDFGALALLGLADRAWPGSGG HDERAGAVVRDIGETGLR  
 CGLNGIRMPGLMLGAAGAGLSLLRLAAPADVP AVTWLEPPRGTHV

**SEQ ID NO: 214 orf14 lanT 575aa**

MSETAGLLRRSLLDHRGKLAAGVAGVGCQLGQPFLLIRRVLTAVQSA  
 QPYRQLALAVLAVMVVGAALGAVQQFLLQRTGEAMVFTVRRTLVAHLLRL  
 PVAAYDERQSGDLVSRVGADTAQVRSVITSGVVDLAGGVLLVGGSIAGMI  
 IIDPVLLGVSLAPVLCGAAGVRLVGRRLRPLSSAVQESIGALTASTTRAL  
 GAIRTIRVAGATERETALIVAEADRARAAGVRLALVAAQAGPIVRLALQG  
 AFVAVIGFGGYRVANGAVSVGDLVAFSTLLFTLALPLAQLAEAAATRIQTG  
 LGALTRIEEILALPDEDSALGVRARTPATVRHDPVLLFDFHVSFRYPTGG  
 EILRDVSFRVPAGSTTALVGPSGAGKSTILALIARLYEVHGGRIILLHGRD  
 IRDYPLAELRAALGYVEQEAPVLAGTVRDNLTLAAPDVAEHAIHVHTASV  
 NLDDLLIARDPAGLDAPVGDGGVLFSGGERQRLAVARTLLAPGELLFLFDEP  
 TAHLDAARNEQALQHGLTAHAAGRTLTVVAHRLATVAHADQILVIDDGRSV  
 AAGRHEELLVRDPTYREFATRQLLT

**SEQ ID NO:215 orf15 lanO (Lucifrase type monooxygenase) 347aa**

MLSVLDQVPVFRGDDPAEAVREAVGLARAAESLGYHRFWIAEHHGSAANA  
 CAAPEIVAAAVAGATERIRVGTGGVLLPYYSPLKVAEAFRVLAAALYPGRI  
 DLGFGRGRGGPAVMAELLNPYAIATEEAYAEQVGRLLAFLGDARTVSRVS  
 VTPAVQDPPLPWLLGSGVGSARLAGMLGVPFCFAQFIATEECPEAIAAYQ  
 ESFRSSPWLDEPQAMLALRVLAAGTAEDAEELATGFWMSCCTGWRAQVRP  
 DDDYRGGVPNLADAQRYTLTEEDLAMRASRPYLQISGTAETVGEIIRRLR  
 KVDVDAEVMLTTNCPGAAAPAPVLAAGRRARADRAQVTRVRPARRW

**SEQ ID NO:216 orf16 lanR (Response regulator) 217aa**

MADVLVVAAEALVSIKIMILETMGGFSVVAADRDSALAAVSEHRPAVVL  
 LDAHATLPESVPLLRRLDLESGPALAVLATLAASSTVLESRLGGACGFL  
 LKDSQPEQLVAAVRALASGVTVLAPAEASSIMLGAAACRGTPAAENAVDEVK  
 QLSDREQAVLGLIGQGLTNAEIAGRFLFISDSTVKEYVSVILRKLGVANRV  
 QAAVLAYAAGLTTDELA

**SEQ ID NO:217 orf17 ABC-transporter associated permease 814aa**

MIFTLAWSQVRSHPGRLLAIVAAVVLATGFLAATATFASTSDEGLRLTAA  
 APLTTADIVLDADDTVHDPQWYQAAAGVRGVRSVDPQYARTVSVFGGDRR  
 GSANVQSIPTASVRWFVDEGTWPSAAGQVADRRTLTDLGVGVGAHLD  
 FRQGTAAFPVPTVVGSAADLGFRLTGSDYRFYAAASFFAGDTPPAALLTV  
 TDRDLAETVDAVGRALPPGATATDASAAADAAAGRFAGGNTQLVVLMLA  
 FAAVALLASILVIANTFQVIVSQVRVQVALLRLVGGHRRQVSRVVLAEAA  
 IAGSIGAVIGAVAGVGLGYLGAGLLDINGGGGLAVNPVLALCVLTGVGAT  
 VVAAWAPARRATRVPPVRALQEVDPALPAQVRGGRRLVAGLILIGLAVGV  
 LGLAAIGTSLPLALVGGVLLAAGLLTALPLGLIALLLPPAARGLERFGVAA  
 SLAGSNLRQNARRTASATMAVVVGAALITGLAVASASGRATVEADLEARY  
 PVAVSVHTDGAIDDRTVRALSGITGLTTATVATSAATFPAAGKPTPARI  
 AALPTDVAGRLAPELSASTGDPVLLVPASYLTARGLTDGAPLTVTAGGRD  
 LRFTARGSRLADTTGQLLGVTGDLTAAGVTVPTTVWGTAPGGFDRET  
 LAADVNAVAARDAGVEVGGGVTEGGDIMNVLSILLGLSLGMLAVTVVIAL  
 LGIANLLGLSVIERTREMAALLRALGTRRSRLRAMVAEAVTITLVGTVAG  
 IVIGVPVGLVGVIAAVGRQAEPVIMLAWPQLGLVLVAAAVTGVLASLAPA  
 RRATRIAPAEGLVR

**SEQ ID NO:218 orf18 ABC-transporter ATP-binding protein 240aa**

MVSCRNLRKRYGTGDAAVVAVDGVSTSFAGEFTAIMGPSGSGKTTLMHL  
 LAGLDTPTEGEVSLAGTALAGLDDRALTDLRRDRVGFIQAFNLLPTLTA  
 EQNIVLPLRLAGRPVDRDLQRIAASLQIGDRLGHRPAELSGGQQQRVAV

ARALLTEPSVVFADPTGALDIATGRALLAGLQNAARQASQTIIMVTHDA  
AAATYADRVLMADGRLWDELRAPTLESIMSVMASVTVTS

**SEQ ID NO:219 orf19 Histidine kinase 362aa**

VIAGAAVMVCLLLGLAGLDEWYWSAALCVPLVIRRSAPVVFLALVAVLSG  
IHMIYSGSFAPGDLVDLVAVHAVAGYGPVRHLGLLLVAGSLVVTAR  
ALHDGLPSSATLPAALIVAATLAAWSTGLMQRRQADVIEADHRRRLAEQ  
DSAMRRLAAIEERTRISQEMHDIIAHSLASVIAQAEGRVAARADAVVA  
GPLFDRIAQIGREALNDVKRLNSIDGDTPDFAQGLPDLPGLLAGVSAA  
GLDVTFEVAGPEQPLASGMDLAVYRVIQESLTNVLKHATQRQARLSLVWT  
PAWLEVSVTSPLTFAGALREGRGLSGIRQRCSLFNGDCEIVAGQTFSVIT  
RWPLARPEVAVP

**SEQ ID NO:220 orf20 Response regulator 218aa**

MTEPQIDVVIADDQDLVRTGFALVVD SAPDMRVVATAADGAEVVRLAAEF  
RPDVVLMDIRMPRVDGITAARAILEGNAQPPKIVALTTYDNDEYASRILA  
AGASGYLLKDTTAEGLTAAIRTVHRGGSVLAPSTTHRLVTAHRQHPARPS  
ALLDSFTTREREVFDLIVAGASNAEADRLNLA EVTIKTHVGRVLAKIGV  
RDRVNVVIWAYRNGAGPS

**SEQ ID NO:221 orf21 Putative membrane protein 301aa**

MPILWTAVLAGAVILGHRLV PNAVGNAGSLIEAFLPWFG LAVPVLLLLAL  
MRRSLTG LAAVLLPLGAWLIHFGGYVVDRTGTPDLIVVQHNVSDENPDP  
AGTARALLAAHPDLVGLLEEVLP EAVAAYRGVLD AELPFHTVQGTVALWSR  
YPLTGAE AIDIRPHDLGEDWNRGLRAVARTPGGDTAVYVAHLPSVRVTAA  
GLTSARRDESARKLGALLAADPVPRLVVIGDLNTSVDDRGLRPIRQVMID  
SPADFAFTWPARTPVARIDQVLARSMTVTRLTALPRTGSDHLPLAAEIRF  
P

**SEQ ID NO:222 orf22 alpha-beta hydrolase 290aa**

MRNNETVRIPVATGGAVTATL FAPESARAVLVVHPATATPQGFYASFATY  
LAENGIATVTYDYRG TGSGSPRDHRDLGMRDWIGADAPAVA AWAADRFP  
GLPRLAAGHSLGGHVIALGAAGPD LAASVIVASHIAALRTIPSRLEFRV  
RIMLHILGPALGRLGYPARSLGLGEDLPAAAMLEWGGWARRDNYFFDD  
PSMRAAERAATLTGPVLAVGTTDDPWSTPRQMDALT VHLT SASVERRTYS  
PAAAGVPVIGHGLFRRVAVRDTVWPELLAWLHAHSEKASR

**SEQ ID NO:223 orf23 Transcriptional regulator 145aa**

VRLPRLIFLLFNADRAVR RWIDARSGDTGIGASGAGVLFYLAGHENALIG  
DVTAAALGASPSGMSGLVNR LERGGCLTRSQDPADARAVRLALT PRGHQVV  
IHARGLVDDLNEQLTAGFD DAEIAVVQRWLEHVTRVSVQREERLG

**SEQ ID NO:224 orf24 Pyruvoyl-dependent arginine decarboxylase 175aa**

MADSVVFDQIPVVRAIGRTTSLSAFH DALVTMECGFYNLVRLSSVIPPG  
TAVDPSGKAPVPVGAWGDKLYCVYAEQHASQPGEEAWAGIGWVQRRDGQG  
GLFVEHEGTSESFVREAIKASLRDLVKGHEDDFDGPDFVVGHVSDGEPV  
CAMVLAPYETAPWRGVRATDPPGMN

**SEQ ID NO:225 orf25 Putative diamino pimelate decarboxylase 406aa**

MTLSEILPSLGASLRPRLD PANWPLTARWTEPGDLTVGGVPVTRIAAAHG  
TPVHVIDETDVHSRCAEYVAAF GPGAVCCSAKGGLLRGAARWIARDGLGC  
YCRSAAELRTALDAGIRPESLALFGSGKSVADLEAALSCGA AVVIGSASE  
AEVVAARSRPGQRVLLRVRPGSAQRGYGVRLNSSAALAAVATVTRSRRLV  
LAGLDCSLGHRLNRF GTYESCLREAIGFVARLR TTVPVNLGGGHAADLP

VGIFAARLRAVAQVTSEGYGIEPPEVHVSPGRALLGRAGITVHRVVAAGD  
GVIELDGDVDPDCLPGADCAGLHTAALIGRASAPGRSITVRCGDATVAVA  
ELPGDMAAGDLVALSGTGAYHQRRDVYVGRPAVVAVCGGRARTLLPRETI  
DRILYA

**SEQ ID NO:226      *orf26 Kinase*      309aa**

VRSKREHSADIRRGRAVLVVNTRSRRGRLLYEEARRRLVEAGFELLGTY  
ALEQSGGLDGLISEALRKEPDLLIAGGGDGTISTAGRMLAHRDVALGVLP  
LGTNNFARTVRIEPDLEAAIATLVDGKVIDVDLGVAGDVPFTNHVGIGL  
SGEVMISAPPRLKRAVGRLAYPMTALGLLARHRPVRAVIRAEGRRLRFHT  
HQVYVANGGFHAGRPITADAHADDRLLVAYPVGGASRRELLRETARNAAT  
GHRRTLHERPFIARELWLETDRPVAVEVDGEPRGTTMPRIAIDPNALRI  
MAPADSPDL

**SEQ ID NO:227      *orf27 Transcriptional regulator*      367aa**

VVFGSLLLVGSGGGAIGLNATVAAATSSVGQESLLGSAKPAAEKKANLD  
GAKNLLLVGIDQRPTQTNGEPLRSDSIILLHINKDHSSGYMISLPRDSYV  
YIPAYDNGKQKWAGGKTKINAAFAFGTRGLKGNEALQHGFEELLTMTVKEL  
TGITPDAGAIIDFQGFDRDVVNLGKVCMYVDTTTTKSIHLGKDQNGKTAKP  
FVINPDGTLKSKISGVTPTNTYTKGDHCFTPGQALDFVRQRDLLADNSLDY  
GRQRHQQQFFKAIINQALKDGLDSPTKLPKLLSAFGKAMTVDDGGIDLAD  
WALAMRSLKPKLLTIKTNAGKLNSENVPGSGSVELLSDDSMDLLKSIKK  
DQIDTFLLSHPAFIANS

**SEQ ID NO:228      *orf28 Glycosyl transferase*      317aa**

MPSEPDSVSVIPTCNRELA VRARSALGQTHRNLVIVVVDGPDEATVT  
ALGEVGDPRLSVIVLPERGKAPNARNTGARAARGRWAMLDDEWLPTK  
IERQLETAATAATVERPVVACRMISRTPRADTIMPRRLPEPGEPISEYLLV  
RRGLFYGDGFVQTSICIMAPTELWRKVFFTVGLRRAQELDWTLRAMREPQT  
ALIYAEPLVLWHQDENRDRISLQNPWREQLWLRGNRELFTPRAYAAFT  
LSVLSSMAAPTRDTGLFRELLAEARTHGDPGTVDYLTMHQI WALPPSVRH  
RLRDVVVGRGKTSSNAG

**SEQ ID NO:229      *orf29 Glycosyl transferase*      369aa**

MPAERRVAIWRSSMLPGSETFVRNQADALTRWTFAYVGAVRHESVLSRPD  
DVIAFFPGGKGLRLRLTGASPLQKTISAVRPNLVHAHFGGDGWLVSLSA  
QQLGVPLAVTVHGHVDTRQPSPPGAKGVYRRNLQTVFTRASLVIAVSEV  
IRGQAIRWGADPAKVKVHYTGIAVPPEQPEEVPKRWDVVFGRFVAKKGV  
DDLTLAALAVESRPRALLIGDGELMTAMRARAELGVDVTFAGSRTPEQV  
RRHLLSRLLACPSKTAPDGDTEGLPTTILEAAALGLPVVATRHSGIPEA  
VIDGETGLLSPEADPAALAVSLTRLLGDEDLQRRLGARARRHVTAHFDLV  
EQTRRELDLYDEVVAGARV

**SEQ ID NO:230      *orf30 Dihydrolipoamide dehydrogenase*      459aa**

MGEHFDLVVLGAGPGGYVAAIRGAQLGLTTAIVEDKYWGGVCLNVGCIPS  
KALLRNAELAHIFHHQAQTFGIEGKVTFDFAVAHQRSRSVADGRVKGVHF  
LMKKNGITEIQGRGEFTDAHTLRVGDRVTFTDNCILATGASTRMIPGTSV  
SKRVVTYEEQILDPLPDSIVIVGAGAIGVEFAYVLRNYGVDVTIVEFLD  
RMLPLEDEEVSKELLRQYRKLGVDRVGTRVEGIEEGADSVRVTVSKNGK  
TEVLEADKVMQAIIGFKPNVEGYGLETTGVTVS DRGAVEIDDFCRTNVPGI  
YAIGDVTAKMLAHAAEAMGIVAAETIAGAETMALDYRMIPRATFCQPOV  
ASFGWTEAQAREQGFDVKVAKFPFTANGKAHGLGDATGFVKILSDAKYGE  
LLGAHLIGPDVTELLPELTLAQQWDLTVHEVGRNVHAHPTLAEAVKEAIIH  
GLAGHMINF

**SEQ ID NO:231 orf31 Putative membrane protein 348aa**

MTTPRRSGTSLVTLTVEPPIAPAPPAAPGRSRRRLGYLAFVLVAVVA  
 VVTLRDRLPDPGEFLDALRAADWRWAALAVGAGVLSQIAYAEQQRRLLAA  
 FGVRVPARRAIAMTYVRSALSMALPAGSAASAAYAFQVYRRHGATAAISA  
 TATLISTVVTVMSLGLLYAATWSLTATVVAGLAVLLLWIYRTVRGPVPAR  
 AGVPRRLRVAPIARLLQRPVAQAALRGARSVPARTWLTVTLAGVINWLLD  
 MACLLLAADALHAGLGWSRLALIYLAVQVVRQIPLTPGGIGLIETSMLAG  
 LIAAGAPQVTAAGIVLIYRLISFWLILPSGLAAHLTLRRGTVPPVTPG

**SEQ ID NO:232-SEQ ID NO:299**

(Null)

**SEQ ID NO:300 O/SBDIG-1**

TGGGTSTGCACSCTSACSATCGARTGCGGNACSGTSATCTGCGCSTGC

**SEQ ID NO:301 O/ACT08F**

TCCAGCACGCGCGGGG

**SEQ ID NO:302 O/ACT09R**

GTTTCGACCAGCCGCC

**SEQ ID NO:303 O/AGvar01bF**

TTCTAGACGTTGTTCTCCCATTTTCAC

**SEQ ID NO:304 O/AGvar02bR**

AAGATCTTCGAAGGTGAGCTCGCCGAA

**SEQ ID NO:305 O/AGvar03F**

GATCTTCGCGAGGACCGCACCATCTACGCCGCCAGCAGCGGCTGGGTGTGTACACTGACGATCGAGTGC  
 GGCACCGTGATCTGCGCCTGCTGAC

**SEQ ID NO:306 O/AGvar04R**

CTAGGTCAGCAGGCGCAGATCACGGTGCCGCACTCGATCGTCAGTGACACACCCAGCCGCTGCTGGCG  
 CGGTAGATGGTGCGGTCCTCGCGAA

**SEQ ID NO:307 O/AGvar05F**

GCCTGCTGACCTAGGTCGACGATCGT

**SEQ ID NO:308 O/AGvar06r**

TGAATTCGGCTGCTCCCCGCGCGAAAT

**SEQ ID NO:309 O/SB50F**

ATTCGCCCCGGAAGTCCACCGAAAGGAAGACACACCATGATTCCGGGGATCCGTCGACC

**SEQ ID NO:310 O/SB51R**

GGGCGATGCCCCCCCCGGGCCGAAACGATCGTCGATCATGTAGGCTGGAGCTGCTTC

**SEQ ID NO:311 O/SB52F**

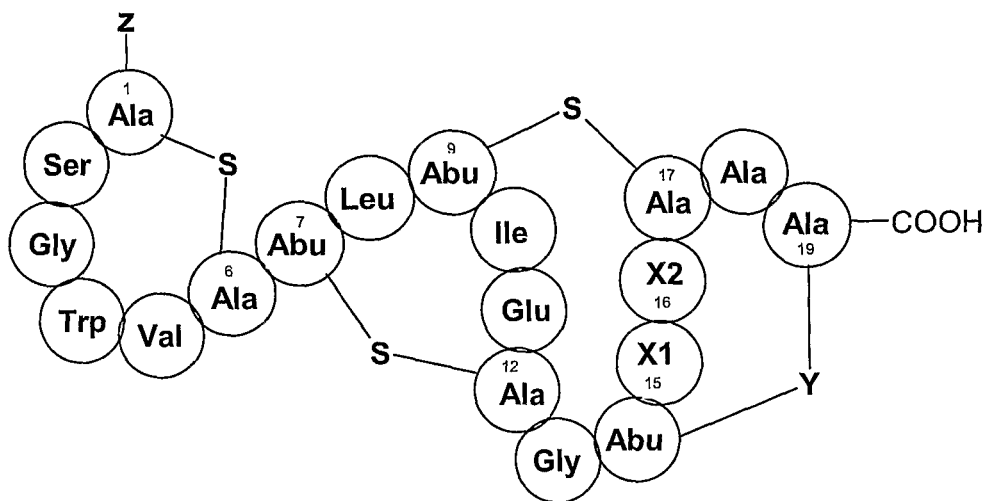
AAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGATTCCGGGGATCCGTCGACC

**SEQ ID NO:312 O/SB53R**

GCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCATGTAGGCTGGAGCTGCTTC

Claims

1. A compound of the formula:



wherein:

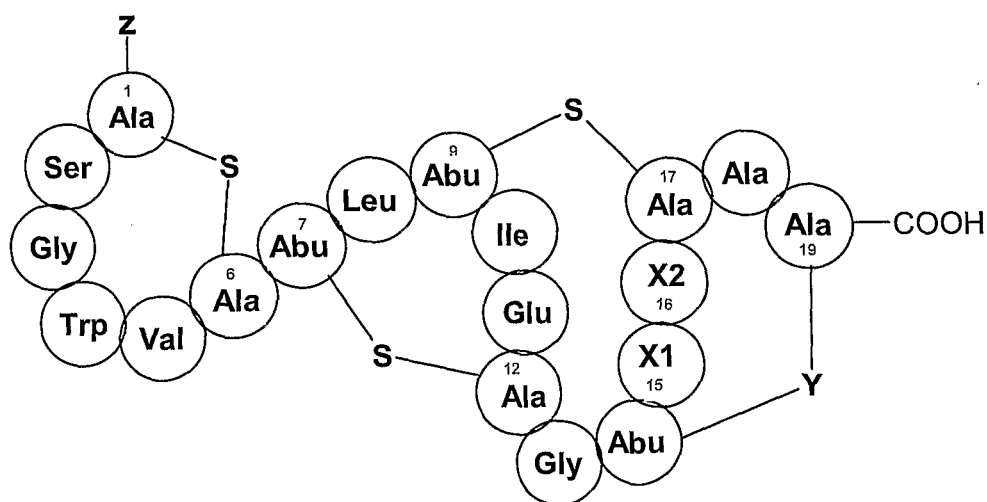
-X1-X2- represent -Leu-Val-; -Val-Val- or -Leu-Ile-;

Y is -S- or -S(O)-; and

Z is either  $\text{NH}_2$  or Ala-,

or a pharmaceutically acceptable salt thereof.

2. A compound of the formula:



wherein:

-X1-X2- represent -Leu-Val-; -Val-Val- or -Leu-Ile-;

Y is -S- or -S(O)-; and

Z is either NH<sub>2</sub> or Ala-,  
or a biologically active variant thereof, or a derivative of said compound or variant,  
or a pharmaceutically acceptable salt of said compound, variant or derivative.

3. A compound according to claim 2 wherein the variant is a compound in which Z is an amino acid.

4. A compound according to claim 3 wherein the amino acid is selected from the group Ala, Ile-, Lys-, Phe-, Val-, Glu-, Asp-, His-, Leu-, Arg-, Ser- and Trp and said amino acids are in the D- or L- configuration.

5. A compound according to claim 3 wherein the variant is a compound in which Z is an amino acid selected from the group Ile-, Lys-, Phe-, Val-, Glu-, Asp-, His-, Leu-, Arg- and Ser-.

6. A compound according to any one of claims 2 to 5 wherein said derivative is a compound which comprises a C-terminal of the formula -COR, in which R represents the group -NR<sup>1</sup>R<sup>2</sup>, wherein R<sup>1</sup> and R<sup>2</sup> independently represent:

(i) hydrogen;

(ii) a group of formula -(CH<sub>2</sub>)<sub>n</sub>-NR<sup>3</sup>R<sup>4</sup>, in which n represents an integer from 2 to 8 and R<sup>3</sup> and R<sup>4</sup> independently represent hydrogen or (C<sub>1</sub>-C<sub>4</sub>) alkyl or R<sup>3</sup> and R<sup>4</sup> taken together represent a group -(CH<sub>2</sub>)<sub>3</sub>-, -(CH<sub>2</sub>)<sub>4</sub>-, (CH<sub>2</sub>)<sub>2</sub>-O-(CH<sub>2</sub>)<sub>2</sub>-, -(CH<sub>2</sub>)<sub>2</sub>-S-(CH<sub>2</sub>)<sub>2</sub>- or -(CH<sub>2</sub>)<sub>5</sub>-; or R<sup>1</sup> and R<sup>2</sup> taken together with the adjacent nitrogen atom represent a piperazine moiety which may be substituted in position 4 with a substituent selected from:

(a) (C<sub>1</sub>-C<sub>4</sub>)alkyl;

(b) (C<sub>5</sub>-C<sub>7</sub>)-cycloalkyl,

(c) pyridyl,

(d) -(CH<sub>2</sub>)<sub>p</sub>-NR<sup>5</sup>R<sup>6</sup> in which p represents an integer from 1 to 8 and R<sup>5</sup> and R<sup>6</sup> independently represent hydrogen or (C<sub>1</sub>-C<sub>4</sub>) alkyl;

(e) piperidinyl;

(f) substituted piperidinyl, wherein the substituted piperidinyl bears a N-substituent which is (C<sub>1-4</sub>)alkyl;

(g) benzyl; and

(h) substituted benzyl, wherein the phenyl moiety bears 1 or 2 substituents selected from chloro, bromo, nitro, (C<sub>1</sub>-C<sub>4</sub>)alkyl and (C<sub>1</sub>-C<sub>4</sub>)alkoxy.

7. A compound according to any one of claims 2 to 6 wherein said derivative is a compound which comprises a C-terminal of the formula -COR, in which R represents the

group  $-NR^1R^2$ , wherein  $R^1$  and  $R^2$  independently represent hydrogen, a group of formula  $-(CH_2)_n-NR^3R^4$ , in which  $n$  represents an integer from 2 to 8 and  $R^3$  and  $R^4$  independently represent hydrogen or  $(C_1-C_4)$  alkyl or  $R^3$  and  $R^4$  taken together represent a group  $-(CH_2)_3-$ ,  $-(CH_2)_4-$ ,  $(CH_2)_2-O-(CH_2)_2-$ ,  $-(CH_2)_2-S-(CH_2)_2-$  or  $-(CH_2)_5-$ , or  $R^1$  and  $R^2$  taken together with the adjacent nitrogen atom represent a piperazine moiety which may be substituted in position 4 with a substituent selected from  $(C_1-C_4)$ alkyl,  $(C_5-C_7)$ -cycloalkyl, pyridyl, benzyl, and substituted benzyl wherein the phenyl moiety bears 1 or 2 substituents selected from chloro, bromo, nitro,  $(C_1-C_4)$ alkyl and  $(C_1-C_4)$ alkoxy.

8. A compound according to any one of claims 2 to 7 in which the compound comprises a modification to a carboxy function of a side chain of an internal residue such that said function is modified from  $-COOH$  to a group  $-COOR^5$  in which  $R^5$  represents hydrogen,  $(C_1-C_4)$ alkyl or  $(C_1-C_4)$ alkoxy  $(C_2-C_4)$ alkyl.

9. A compound according to any one of claims 2 to 8 in which the compound comprises a modification to the N-terminal amino group such that this group  $-NH_2$  is instead a group  $-NHR^6$  wherein  $R^6$  represents  $C_{1-4}$ alkyl.

10. An isolated nucleic acid encoding a polypeptide selected from any one of the group of SEQ ID NOs:1-4, SEQ ID NOs:11-14, SEQ ID NOs:212, 22, 23 and 119.

11. An isolated nucleic acid encoding a polypeptide selected from any one of the group consisting of SEQ ID NOs:1-4, SEQ ID NOs:11-14, SEQ ID NOs: 212, 22, 23 and 119 or a biologically active variant thereof.

12. An isolated polypeptide comprising a sequence selected from any one of SEQ ID NOs:101-132, or a variant having at least 90% sequence identity thereto.

13. An isolated polypeptide comprising a sequence selected from any one of SEQ ID NOs:201-231, or a variant having at least 90% sequence identity thereto.

14. An expression construct comprising the nucleic acid of claim 10 or 11.

15. A set of expression constructs comprising a first nucleic acid according to claim 10 or 11 and a second nucleic acid encoding a *LanM* polypeptide comprising SEQ ID NO:120 or SEQ ID NO:213 or a variant having at least 90% identity to SEQ ID NO:120 or SEQ ID NO:213.

16. A set of expression constructs comprising a first nucleic acid according to claim 10 or 11, or a first and second nucleic acid according to claim 7, and a third nucleic acid encoding a *LanR* polypeptide comprising SEQ ID NO:122 or SEQ ID NO:216 or a variant having at least 90% identity to SEQ ID NO:122 or SEQ ID NO:216.
17. A set of expression constructs according to claim 15 or 16 and further comprising a nucleic acid encoding a *LanT* polypeptide of SEQ ID NO:123 or a variant thereof, or SEQ ID NO:214 or a variant thereof.
18. A set of expression constructs according to claim 15, 16 or 17 further comprising a nucleic acid encoding a *lanO* polypeptide of SEQ ID NO:122 or a variant thereof, or SEQ ID NO:215 or a variant thereof.
19. A recombinant vector or vectors comprising the expression constructs of any one of claims 14 to 17.
20. A host cell transformed with the vector or vectors of claim 19.
21. The host cell of claim 20 which is an actinomycete.
22. The actinomycete of claim 21 which is a non-producer of a lantibiotic polypeptide.
23. The actinomycete of claim 21 or 22 which is selected from the group consisting of *Streptomyces lividans*, *S. coelicolor* and *S. cinnamoneus*.
24. A method of making a compound according to any one of claims 1 to 9 comprising expressing a nucleic acid encoding a polypeptide comprising a sequence selected from the group of SEQ ID NOs:1, 2, 3, 11, 12, 13, 212, 22 and 23, or a variant of any of said sequences in a lantibiotic-producing host cell.
25. The method of claim 24 wherein said host cell has a mutation in its endogenous *LanA* gene such that the gene is not expressed or the gene product is inactive.
26. The method of claim 24 or 25 wherein the host cell is *A. garbadinensis* or *A. liguriae*.

27. The method of any one of claims 24 to 26 wherein the host cell is *A. liguriae* NCIMB 41362.
28. The method of claim 24 wherein the host cell is an actinomycete which does not produce a lantibiotic and wherein the host cell is transformed with SEQ ID NO:100 or SEQ ID NO:200, or a fragment or variant of said sequences, wherein the fragment or variant confers on the cell the ability to produce a lantibiotic.
29. The method of claim 28 wherein the host cell is transformed with nucleic acid encoding SEQ ID NOs:116-127, or a variant thereof wherein said nucleic acid encodes polypeptides having at least 90% identity to each of SEQ ID NOs:116-127.
30. The method of claim 28 wherein the host cell is transformed with nucleic acid encoding SEQ ID NOs:206-220, or a variant thereof wherein said nucleic acid encodes polypeptides having at least 90% identity to each of SEQ ID NOs:206-220.
31. *A. liguriae* NCIMB 41362.
32. A derivative of *A. liguriae* NCIMB 41362 comprising a mutation in its *LigA* gene such that the gene is not expressed or the gene product is inactive.
33. A recombinant DNA cassette which comprises a nucleotide sequence encoding the lantibiotic LigA pre-pro-peptide of SEQ ID NO:212, wherein said sequence comprises  
a first restriction site at or adjacent the N-terminal encoding region of the encoding sequence;  
optionally a second restriction site downstream of the first restriction site and within the encoding sequence; and  
a third restriction site at or adjacent the C-terminal encoding region of the encoding sequence,  
wherein at least one of said restriction sites does not occur within the ligA coding sequence of SEQ ID NO:200.
34. A recombinant DNA cassette which comprises a nucleotide sequence encoding the lantibiotic LanA pre-pro-peptide of SEQ ID NO:119, wherein said sequence comprises  
a first restriction site at or adjacent the N-terminal encoding region of the encoding sequence;

optionally a second restriction site downstream of the first restriction site and within the encoding sequence; and

a third restriction site at or adjacent the C-terminal encoding region of the encoding sequence,

wherein at least one of said restriction sites does not occur within the lanA coding sequence of SEQ ID NO:100.

35. A method of making a library of lantibiotic-encoding gene sequences which comprises: providing an expression cassette according to claim 33 or 34

introducing into the cassette, between the first and second restriction sites, the first and third restriction sites, or the second and third restriction sites, a multiplicity of sequences each of which corresponds to the corresponding lantibiotic-encoding sequence apart from having from 1 to 15 nucleotide changes; and

recovering an expression library encoding a multiplicity of lantibiotic-variant sequences.

36. A method which comprises:

providing an expression library produced according to claim 35;

introducing one or more members of said library into host cells capable of expression of a lantibiotic; and

screening said host cells for lantibiotic activity.

37. The method of claim 36 which further comprises recovering the lantibiotic.

38. The method of claim 36 or 37 which further comprises isolating a host cell comprising a member of the library.

39. An isolated nucleic acid comprising SEQ ID NO:100 or SEQ ID NO:200, or a fragment or variant of said sequences, wherein the fragment or variant confers on an actinomycete host cell the ability to produce a lantibiotic.

40. The isolated nucleic acid of claim 39 wherein the nucleic acid encodes SEQ ID NOs:116-127, or a variant thereof wherein said nucleic acid encodes polypeptides having at least 90% identity to each of SEQ ID NOs:116-127.

41. The isolated nucleic acid of claim 39 wherein the host cell is transformed with nucleic acid encoding SEQ ID NOs:206-220, or a variant thereof wherein said nucleic acid encodes polypeptides having at least 90% identity to each of SEQ ID NOs:206-220.
42. A pharmaceutical composition comprising a compound according to any one of claims 1 to 9 together with a pharmaceutically acceptable carrier or diluent.
43. A method treatment or prophylaxis of a bacterial infection in a subject, which method comprises administering to said subject a compound according to any one of claims 1 to 9, or a composition according to claim 42.