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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
18 April 2013 (18.04.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/053857 A2

(72) **Inventors: BROENSTRUP, Mark**; Sanofi-Aventis Deutschland GmbH, 65926 Frankfurt am Main (DE). **KOENIG, Caudia**; Sanofi-Aventis Deutschland GmbH, 65926 Frankfurt am Main (DE). **TOTI, Luigi**; Sanofi-Aventis Deutschland GmbH, 65926 Frankfurt am Main (DE). **WINK, Joachim**; Sanofi-Aventis Deutschland GmbH, 65926 Frankfurt am Main (DE). **LEUSCHNER, Wulf**; Sanofi-Aventis Deutschland GmbH, 65926 Frankfurt am Main (DE). **GASSENHUBER, Johann**; Sanofi-Aventis Deutschland GmbH, 65926 Frankfurt am Main (DE). **MUELLER, Rolf**; Helmholtz-Institut für Pharmazeutische Forschung Saarland Universität des Saarlandes, Campus C23, 66123 Saarbrücken (DE). **WENZEL, Silke**; Helmholtz-Institut für Pharmazeutische Forschung Saarland, Universität des Saarlandes, Campus C23, 66123 Saarbrücken (DE). **BINZ, Tina**; Universität Zürich, Institut für Rechtsmedizin, Kurvenstrasse 17, 8006 Zürich (CH). **VOLZ, Carsten**; Helmholtz-Institut für Pharmazeutische Forschung Saarland, Campus C23, 66123 Saarbrücken (DE).

(54) **Title:** GENE CLUSTER FOR BIOSYNTHESIS OF GRISELIMYCIN AND METHYLGRISELIMYCIN

(57) **Abstract:** The present invention refers to the gene cluster and genes comprised by the gene cluster which are involved in the biosynthesis of griselimycin and methylgriselimycin and to the use of the gene cluster, genes comprised thereby and proteins encoded thereby for the production of antibiotic agents.

1

Gene cluster for biosynthesis of griselimycin and methylgriselimycin

The present invention relates to the field of antibiotics and more specifically to genes and pro-
teins involved in the biosynthesis of (methyl)griselimycin. The present invention provides re-
combinant methods for producing (methyl)griselimycin by recombinant DNA technology.

Background of the invention

Griselimycin is a naturally occurring antibiotic produced by microorganism of the genus *Strep-*
tomyces which is the largest genus of the *Actinobacteria* and the type genus of the family
Streptomycetaceae. Streptomycetes are Gram-positive and have genomes with high GC-content.
Streptomycetes are characterized by a complex secondary metabolism. They produce over two-
thirds of the clinically useful antibiotics of natural origin. Griselimycin is an antibiotic which
was first described by Terlain and Thomas, 1971. It is composed of ten amino acids being in this
order L-N-methylvaline, which is N-acetylated, L-trans-4-methylproline, L-N-methylthreonine,
L-leucine, L-trans-4-methylproline, L-leucine, L-N-methylvaline, L-proline, D-N-methylleucine
and glycine. The peptide chain is cyclised between glycine and L-N-methylthreonine via an ester
bond resulting in a non-cyclised tail consisting of the amino acids L-N-methylvaline, which is N-
acetylated, and L-trans-4-methylproline.

Griselimycin belongs to the group of non-ribosomally synthesized microbial peptides which
group of peptides shows a remarkable structural diversity and comprises a widespread class of
the most potent antibiotics and other important pharmaceuticals. Although structurally diverse,
non-ribosomally synthesized microbial peptides share a common mode of synthesis, the multien-
zyme thiotemplate mechanism. Thereby, peptide bond formation takes place on large multien-
zyme complexes, which simultaneously represent template and biosynthesis machinery. These
multienzyme complexes are designated non-ribosomal peptide synthetases (NRPSs). Sequencing
of genes encoding NRPS revealed a modular organization. A module is a section of the NRPS
polypeptide chain that is responsible for the incorporation of one defined monomer into the
growing polypeptide chain. Thus, NRPSs are used simultaneously as template because the amino
acid to be incorporated is determined by the module and as biosynthetic machinery because it is
the module that harbors all necessary catalytic functions. Modules can be further dissected into
catalytic domains. These domains catalyse at least the steps of substrate activation, covalent

2

binding and peptide bond formation. Domains of equal function share a number of highly conserved sequence motifs. They can be modified for activity changes within the polypeptide chain which opens up prospects for manipulation of the NRPS machinery.

5   Within each module, the selection of a specific substrate (amino acid) is mediated by an adenylation domain (A-domain). This A-domain is responsible for the selection of the amino acids that make up the product and thus controls its primary sequence. A-domains activate amino and also some carboxy substrates as amino acyl adenylate while ATP is consumed. This reactive intermediate is further transported onto the terminal cysteamine thiol moiety of the Ppan prosthetic

10   group that is attached to the peptidyl carrier protein (PCP) domain which is also referred to as thiolation (T) domain located downstream of the A-domain in the same module. It represents the transport unit that accepts the activated amino acid that is covalently tethered to its 4′PP cofactor thioester. This cofactor acts as a flexible arm to allow the bound amino acyl and peptidyl substrate to travel between different catalytic centers. The combination of A-domain and T-domain

15   is defined as an initiation module since both domains are required to activate and covalently tether the first building block for subsequent peptide synthesis. The initiation modules of lipopeptide pathways harbor an additional N- terminal C domain for condensation of an acyl side chain with the amino group of the first amino acid. The condensation or C-domain is the central entity of non-ribosomal peptides synthesis because it is responsible for peptide bond formation

20   between amino acyl substrates bound to the T-domain of adjacent modules. The enzyme catalyses the nucleophilic attack of the amino (or imino, hydroxyl) group of the activated amino acid bound to the downstream (with respect to the C-domain) module onto the acyl group or the amino acid tethered to the upstream module. The resulting peptidyl intermediate is then translocated down the assembly line for subsequent condensation and further modification steps. Dur-

25   ing synthesis, the growing peptide chain is handed over from one module to the next module until it reaches the final modul`s T-domain. This module contains in most cases a TE (thioesterase)-domain that is important for the liberation of the product. Product release is achieved by a two-step process that involves an acyl-O-TE-enzyme intermediate that is subsequently attacked by either a peptide-internal nucleophile or water, which results in a macrocyclic product

30   or a linear peptide, respectively.

An especially striking feature of non-ribosomal peptides is the occurrence of D-amino acids. One way to incorporate a D-amino acid is to use a D-amino acid selective A domain. The more common way is, however, through the use of E (epimerization)-domains that occur at the C-terminal end of modules responsible for the incorporation of D-amino acids. The enzyme catalyses the epimerisation of T-domain bound L-amino acid of the growing peptide chain. Discrimination of the T-domain bound L-amino acid through enantioselectivity of the downstream condensation domain donor site leads to a clean D-configurated product.

A number of NRPS contain methyltransferases (MT domain) that are responsible for the N- or C-methylation of amino acid residues thus making the peptide less susceptible to proteolytic breakdown. Both types of methyltransferases (N- or C-methyltransferases or abbreviated N-MT or C-MT) use S-adenosyl methionine as the methyl donor. The MT domains are often found as insertions in the A-domains (Finking et Marahiel, 2004; Marahiel, 2009). Even more striking is the incorporation of unnatural amino acids such as methyl-proline which are usually biosynthesized in advance to their use in the assembly line. Alternatively, incorporated amino acids can be modified enzymatically post assembly by the NRPS.

Griselimycin is naturally produced by microorganisms of the genus *Streptomyces*. However, although the production of griselimycin by *Streptomyces* has been known in the art, the genetic locus responsible for the biosynthesis of griselimycin has not been identified so far. Consequently, targeting and modification of the biosynthesis of griselimycin were still limited. Accordingly, there is a need for genetic information regarding the biosynthesis of griselimycin.

Summary of the invention

Besides griselimycin, *Streptomycetes* also produce a variant form of griselimycin, termed methylgriselimycin. Methylgriselimycin contains L-trans-4-methylproline instead of L-proline. Methylgriselimycin possesses some advantages, such as e.g different antibacterial specificity, in its use as an anti-bacterial agent over griselimycin, however, is produced by *Streptomycetes* as a side-component only. Accordingly, there is a need for genetic information regarding the biosynthesis of methylgriselimycin.

4

The production of the methylproline moiety as comprised by griselimycin and methylgriselimy-cin, namely L-trans-4-methylproline, was completely unknown. Accordingly, there is a need for genetic information regarding the biosynthesis of L-trans-4-methylproline.

5   The genetic locus encoding the polypeptides involved in the biosynthetic pathway of (me-thyl)griselimycin has now been identified. The present inventors succeeded in isolating and clon-ing the entire griselimycin biosynthetic gene cluster from the genomic DNA of *Streptomyces DSM 26643*. The present inventors also succeeded in elucidating the open reading frames (ORFs) that encode the polypeptides involved in the (methyl)griselimycin biosynthetic pathway

10   including such responsible for the formation of L-trans-4-methylproline. This allows the identifi-cation and influencing of the distinct biosynthesis steps on the molecular level.

Thus, in one aspect, the present invention provides a nucleic acid comprising at least one nucleic acid selected from:

15   (a)   a nucleic acid comprising at least one of the Open Reading Frames (ORFs) 1 to 26 as comprised by SEQ ID NO: 1 encoding proteins of SEQ ID Nos: 2 to 27 or a variant or fragment thereof whereby the variant or fragment encodes a functionally active variant or fragment of a protein of SEQ ID Nos: 2 to 27,

       (b)   a nucleic acid encoding at least one of the proteins of SEQ ID Nos: 2 to 27 or a func-

20   tionally active variant or fragment thereof,

       I   a nucleic acid encoding a protein that is at least 70 %, 80 %, 90 %, 95 % or 97% identi-cal in amino acid sequence to a protein or fragment thereof encoded by the nucleic acid of (a) or (b),

       (d)   a nucleic acid which hybridizes under stringent conditions with a nucleic acid of (a) to

25   (c),

       (e)   a nucleic acid which hybridizes under stringent conditions with a nucleic acid of (a) to (d) and consists of 10 to 50 nucleotides in length, or

       (f)   a nucleic acid which is complementary to a nucleic acid of (a) to (f).

30   Detailed Description

The nucleic acid as comprised by the present invention is, in one embodiment, a nucleic acid comprising or consisting of SEQ ID NO: 1 being the part of the genome of *Streptomyces DSM 26643* comprising the griselimycin biosynthesis gene cluster with a length of 66.868 nucleotides. This newly discovered gene cluster contains at least 26 ORFs designated ORF 1 to ORF 26 (ORF 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26) or designated with the gene names orf-1, orf-2, orf-3, orf-4, orf-5, orf-6, orf-7, nrps-1, int-1, tnp-1, int-2, tnp-2, tnp-3, dna, nrps-2, nrps-3, mbtH, mps-1, mps-2, mps-3, mps-4, orf-8, orf-9, orf-10, orf-11, orf-12, respectively, from nucleotides 837 to 66.108 of SEQ ID NO: 1 with a length of 65.272 nucleotides. These ORFs encode proteins of SEQ ID Nos: 2 to 27 (SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27), respectively, involved in the biosynthesis of (methyl)griselimycin. Structural and functional characterization of the proteins encoded by the griselimycin biosynthesis gene cluster is provided below.

In the context of the present invention, the term "(methyl)griselimycin" refers to griselimycin as well as to methylgriselimycin. Griselimycin is an antibiotic which is composed of ten amino acids being L-N-methylvaline, which is acetylated (1), L-trans-4-methylproline (2), L-N-methylthreonine (3), L-leucine (4), L-trans-4-methylproline (5), L-leucine (6), L-N-methylvaline (7), L-proline (8), D-N-methyl-D-leucine (9) and glycine (10). The structural formula of griselimycin and methylgriselimycin is given in Fig. 1. Griselimycin and its variant methylgriselimycin are produced in different amounts by the same biosynthesis pathway involving the same genes of the (methyl)griselimycin biosynthesis gene cluster as identified herein. Methylgriselimycin differs from griselimycin by the presence of L-trans-4-methylproline instead of L-proline at position 8. The multidomain-enzyme responsible for the incorporation of L-trans-4-methylproline into methylgriselimycin and of L-proline into griselimycin is the same enzyme, namely NRPS-2, and is the protein product of the gene nrps-2, as identified herein.

In the context of the present invention the term "gene cluster" refers to a set of several genes which are located on a contiguous stretch of the genome and which direct the synthesis of (methyl)griselimycin. Some of the genes belonging to the gene cluster are genes which are related and are grouped to a gene family. They encode proteins which are similar in structure and function. The remainder of the genes are unrelated in structure and encode proteins which have different structure and function. The encoded proteins are either enzymes which catalyse reac-

6

tions or are involved inter alia in regulation or transport. Altogether, the genes as comprised by the gene cluster encode proteins that serve the purpose of the biosynthesis of (methyl)griselimycin.

5   The term "nucleic acid" encompasses nucleic acids which are defined by ORFs 1 to 26 coding for the proteins of SEQ ID Nos: 2 to 27, respectively, as well as naturally occurring and non-naturally occurring variants or fragments thereof (as defined herein). The nucleic acid may be any macromolecule composed of chains of monomeric nucleotides carrying genetic information or form structures within cells. The most common (and therefore preferred) nucleic acids are

10   deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The nucleic acid can be a DNA molecule such as a genomic DNA molecule such as SEQ ID NO: 1 or a cDNA molecule which can be single- or double-stranded such as a nucleic acid representing an ORF and encoding a protein as well as a synthetic DNA such as a synthesized single-stranded polynucleotide such as a primer or probe. The nucleic acid of the present invention can also be an RNA molecule. Pref-

15   erably, the term also relates to non-coding regions of a gene, wherein these sections are of a relevant size in order to be specific for that gene. Examples of those regions are regulatory elements such as a promoter. Most preferably, the term "nucleic acid" relates to gene, ORF, promoter, DNA, cDNA or mRNA. The nucleic acid encoding the desired genetic information, preferably DNA, may comprise the gene of interest, a promoter region, a start codon and a stop codon and

20   possibly further regions which may be used for regulation of expression of the gene. The regulatory regions may be heterologous to the respective gene or may be associated therewith in nature. The genetic information may be expressed permanently or under the control of a repressor and/or a promoter region in a cell into which the nucleic acid of the present invention is introduced. The obtained cells may be either used directly or used for tissue cultures or the cells may

25   be harvested and samples comprising the respective protein are obtained by disrupting the cells. Alternatively, DNA or RNA may be used either in cells or in cell-free expression systems such as, e.g., microarray systems in which the DNA or RNA is immobilized and is translated and/or transcribed by the addition of functional cell lysate, comprising the factors required for transcription and/or translation (enzymes, ribosomes, tRNA, amino acids, nucleotides, ATP etc.). Also

30   included are artificial nucleic acids which include peptide nucleic acid (PNA), morpholino and locked nucleic acid (LNA), as well as glycol nucleic acid (GNA) and threose nucleic acid

7

(TNA). Each of these is distinguished from naturally-occurring DNA or RNA by changes to the backbone of the molecule.

The term "comprising" as used herein is meant to "include or encompass" the desired feature and
5   further features which must not be specifically mentioned.. The term "comprising" is also meant to "consist of" the desired feature and not to include further features except the desired feature. Thus, the nucleic acid or protein referred to herein may be defined by additional features in addition to the definition as indicated, e.g. in addition to the definition by an ORF or SEQ ID number, or may consist of such indicated feature only.
10

The term "heterologous" as it relates to nucleic acid sequences such as coding or control sequences denotes sequences that are normally not associated with a region of a recombinant construct and/or a particular cell. A "heterologous" region is an identifiable segment of a nucleic acid within or attached to another nucleic acid that is not found in association with the other
15   molecule in nature. For example, a heterologous region of a construct could be a regulatory region not found to be associated with a gene as identified herein in nature. Similarly, a heterologous sequence could be a coding sequence which is itself not found in nature as it contains e.g. synthetic sequences with codons different from the native gene. Moreover, a host cell transformed with a construct which is not normally present in the host cell would be considered het-
20   erologous for the purposes of the present invention. A homologous nucleic acid sequence is a variant sequence as defined herein. The term "homologous" may be used interchangeably with variant. The term "homologous" may also refer to an identical sequence.

In an embodiment of the present invention the nucleic acid comprises a nucleic acid molecule
25   which encodes at least one distinct protein as comprised by the sequence of SEQ ID Nos: 2 to 27. In a particular embodiment, the entire sequence of the (methyl)griselimycin gene cluster as comprised by SEQ ID NO: 1 is comprised by a nucleic acid of the present invention. In another embodiment, the nucleic acids are variants or fragments of ORFs 1 to 26 encoding functionally active variants or fragments of at least one of the proteins of SEQ ID Nos: 2 to 27. The most pre-
30   ferred nucleic acid codes for a naturally occurring variant protein of ORFs 1 to 26 as detailed below, more preferably, a naturally occurring *Streptomyces* protein, still more preferably a protein of SEQ ID Nos: 2 to 27, which is involved in the synthesis of (methyl)griselimycin.

8

The term "nucleic acid" may comprise full length ORFs 1 to 26 as disclosed herein or may comprise fragments or partial sequences of ORFs 1 to 26. Such fragments or partial sequences of ORFs 1 to 26 encode fragments of the proteins with the amino acid sequences shown in SEQ ID Nos: 2 to 27, respectively. This may include fragmental proteins with short internal and/or C- and/or N-terminal deletions whereby the activity of the resulting proteins as identified herein is maintained to an extent of more than 5 %, more than 10 %, more than 20 %, more than 30 %, more than 40 %, more than 50 %, more than 60 %, more than 70 %, more than 80 %, more than 90 %, more than 95 %, more than 97 % or more than 100 % , e.g. more than 150 %, 200 %, 300 % 400 % or 500 %, of the activity of the wild-type proteins as encoded by the ORFs 1 to 26. Consequently, the respective nucleic acid encoding such fragments contains deletions within and/or at the 5` and/or 3` termini of ORFs 1 to 26, e. g., deletions of at the most 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10 %, 5%, or less. In the context of the present invention, a fragmental nucleic acid sequence encoding a functionally active fragment of a protein of SEQ ID Nos: 2 to 27 means a sequence encoding a fragment which directs synthesis of (methyl)griselimycin and/or which can be substituted for the respective ORF encoding the protein of SEQ ID Nos: 2 to 27 to direct synthesis of (methyl)griselimycin.

The term "nucleic acid" may refer to variants of ORFs 1 to 26 as disclosed herein encoding functionally active variants of the proteins of SEQ ID Nos: 2 to 27 as defined herein. Such functionally active variants have a sequence identity with SEQ ID Nos: 2 to 27 of more than 50%, of more than 60%, preferably more than 70%, more preferably of more than 80%, still more preferably more than 85%, even more preferably more than 90%, even more preferably more than 95%, most preferably more than 97% and/or have an activity of more than 5%, more than 10%, more than 20%, more than 30%, more than 40%, more than 50%, more than 60%, more than 70%, more than 80%, more than 90%, more than 95%, more than 97% or more than 100 % , e.g. more than 150 %, 200 %, 300 % 400 % or 500 % of the activity of the proteins of SEQ ID Nos: 2 to 27. Consequently, the nucleic acid encoding such variants contains deletions, insertions, substitutions and/or additions within and/or at the 5` and/or 3` termini of ORFs 1 to 26 and show an identity to the sequences of ORFs 1 to 26 of more than 50%, more than 60%, more than 70%, preferably more than 80%, more preferably more than 85%, even more preferably more than 90%, even more preferably more than 95%, most preferably more than 97%. In the context of the

present invention, a variant nucleic acid sequence encoding a functionally active variant of the protein of SEQ ID Nos: 2 to 27 means a sequence encoding a variant which directs synthesis of (methyl)griselimycin and/or which can be substituted for the respective ORF encoding the protein of SEQ ID Nos: 2 to 27 to direct synthesis of (methyl)griselimycin.

It is noted that the above mentioned modifications may be combined. For example, a nucleic acid as comprised by the present invention may be a fragment comprising one or more variations of ORFs 1 to 26. It should also be noted that fragments or variants include fragments or variants, as defined herein, of promoter or regulatory sequences with which ORFs 1 to 26 or fragments or variants thereof are associated in nature. These variants are functionally active in that they regulate the transcription or translation of the genes associated therewith.

An ORF is an open reading frame which is a DNA sequence that could potentially encode a protein. In the context of the present invention, the term "ORF" stands for open reading frame in the (methyl)griselimycin biosynthetic gene cluster as isolated from *Streptomyces DSM 22643* (ORFs 1 to 26). The present inventors succeeded in identifying the griselimycin biosynthesis gene cluster of *Streptomyces DSM 22643* and identified 26 ORFs over a stretch of 65272 nucleotides. These ORFs were designated orf-1, orf-2, orf-3, orf-4, orf-5, orf-6, orf-7, nrps-1, int-1, tnp-1, int-2, tnp-2, tnp-3, dna, nrps-2, nrps-3, mbtH, mps-1, mps-2, mps-3, mps-4, orf-8, orf-9, orf-10, orf-11 or orf-12, respectively. The whole griselimycin biosynthesis gene cluster of *Streptomyces DSM 22643* is 66868 nucleotides in length comprising ORFs 1 to 26 and additional nucleotides located 5` and 3` to ORF 1 and ORF 26, respectively.

The 26 ORFs comprised by SEQ ID NO: 1 extend from nucleotide 837 to nucleotide 66108 of SEQ ID NO: 1 of *Streptomyces DSM 26643*. Table 1 shows a list of the specific putative ORFs, the designation of the corresponding genes, the start and stop nucleotides within SEQ ID NO: 1, the lengths of the genes in nucleotides (nt), the strandedness, the number of amino acids (aa) of the putative proteins encoded thereby, the protein designation and the SEQ ID numbers of the proteins. Table 2 shows proteins which are designated by their GenBank accession numbers as available from the NCBI (National Centre for Biotechnology Information; National Library of Medicine 38A, Bethesda, MD20894; USA; www.ncbi.nih.gov). These proteins have been identified, based on homology searches, as showing the highest identity/similarity to the proteins of

SEQ ID Nos: 2 to 27. These proteins are defined by their function, their accession number and their origin. Table 2 moreover shows the e-value and the degree of identity and similarity between the proteins of SEQ ID Nos: 2 to 27 and the proteins as identified by the homology searches. The e-value relates to the expected number of chance alignments with an alignment score of at least equal to the observed alignment score. An e-value of 0.00 indicates a perfect homolog. The e-value is calculated as described in Altschul S.F. et al. 1997. The e-value assists in the determination of whether two sequences display sufficient similarity to justify an inference of homology. Identity indicates that identical amino acids are present in the compared proteins at the respective sites whereas similarity indicates that conservative amino acids substitutions as defined herein below are present at the respective sites. Moreover, the putative function of the proteins encoded by ORFs 1 to 26 is indicated.

The substitution of a variant nucleic acid for ORFs 1 to 26 to direct synthesis of (methyl)griselimycin means that this variant nucleic acid can be inserted into the genome of *Streptomyces DSM 22643* instead of the ORF to which it is a variant thereby expressing a variant protein which takes over the function of the respective protein of SEQ ID Nos: 2 to 27 and directs, i.e. participates in, the synthesis of (methyl)griselimycin. The extent to which the variant takes over the function is as defined herein.

In another embodiment, the nucleic acid encodes a protein of SEQ ID Nos: 2 to 27. The nucleic acid may comprise the sequences of ORFs 1 to 26 as comprised by SEQ ID NO: 1 with the start and stop nucleotides (nt) as indicated in Table 1. The nucleic acid may also encode a protein with the same amino acids as the proteins of SEQ ID Nos: 2 to 27, however, which differs in its nucleotide composition due to the degeneracy of the genetic code.

In a further embodiment, the nucleic acid hybridizes under stringent conditions to a nucleic acid which is comprised by the (methyl)griselimycin gene cluster as identified by SEQ ID NO: 1 or which constitutes ORFs 1 to 26 or which encodes proteins with SEQ ID NO: 2 to 27 or functionally active variants or fragments thereof. In the present invention, the term "hybridise(s)(ing) under stringent conditions" refers to the formation of a hybrid between two nucleic acid molecules under conditions which allow the formation of a so-called specific hybrid, while a non-specific hybrid is substantially not formed. An example of such conditions includes conditions

under which a complementary strand of a highly identical nucleic acid, namely, a DNA com-
posed of a nucleotide sequence having 70 % or more, preferably 80% or more, more preferably
85% or more, still more preferably 90% or more and even more preferably 95% or more identity
with the nucleotide sequence of SEQ ID NO. 1 or shown in ORFs 1 to 26, hybridizes, while a
less complementary strand of a nucleic acid less identical than the above does not hybridize.
More specifically, such conditions refer to conditions in which the sodium salt concentration is
15 to 750 mM, preferably 50 to 750 mM, and more preferably 300 to 750 mM, the temperature is
25 to 70°C, preferably 50 to 70°C, and more preferably 55 to 65°C, and the formamide concen-
tration is 0 to 50%, preferably 20 to 50%, and more preferably 35 to 45%. Furthermore, under
stringent conditions, conditions for washing a filter after hybridization normally comprise the
following: the sodium salt concentration is 15 to 600 mM, preferably 50 to 600 mM, and more
preferably 300 to 600 mM, and the temperature is 50 to 70°C, preferably 55 to 70°C, and more
preferably 60°C. Stringency, and thus specificity, can e.g. be increased by increasing the reaction
temperature and/or lowering the ion strength of the reaction buffer. For example, low stringent
conditions comprise hybridization in 3 x SSC at room temperature to 65°C and highly stringent
conditions comprise hybridization in 0.1 x SSC at 68°C. Exemplary moderately stringent condi-
tions (nucleic acids hybridize under moderately stringent conditions, if they are maximally de-
generate with respect to their codon composition) comprise 50% formamide, 5 x SSC and 1%
SDS at 42°C and washing in 1 x SSC at 45°C. Highly stringent conditions comprise incubation
at 42°C, 50% formamide, 5 x SSC and 1% SDS (e.g. 50% formamide, 5 x SSC and 1% SDS, 50
mM sodium phosphate, 5 x Denhardt`s solution, 10 x dextran sulphate, 20 mg/ml sheared salmon
sperm DNA) or 5 x SSC and 1% SDS at 65°C and washing in 0.2 x SSC and 0.1% SDS at about
65°C (1 x SSC stands for 0.15 M sodium chloride and 0.015 M trisodium citrate buffer). Pre-
ferred in the present invention are moderately or highly stringent conditions, most preferred are
highly stringent conditions. In the context of the present invention, a "hybridizing" sequence
means a sequence which encodes a protein which directs synthesis of (methyl)griselimycin
and/or which can be substituted for the ORF to which it specifically hybridizes to direct synthe-
sis of (methyl)griselimycin.

In a further preferred embodiment of the present invention, a nucleic acid which hybridizes un-
der stringent conditions with a nucleic acid as identified above and consists of 5 to 100 nucleo-
tides in length is provided. The sequence information provided in the present application enables

the design of specific nucleotide probes and primers that are useful for identifying and isolating microorganisms producing (methyl)griselimycin and harboring any of the proteins of SEQ ID Nos: 2 to 27 or variants thereof. Thus, the nucleotide probe has a sequence found in or derived by the degeneracy of the genetic code from a sequence within the (methyl)griselimycin gene

5   cluster as comprised by SEQ ID NO: 1 or a variant thereof or encoding any of SEQ ID Nos: 2 to 27 or functionally active variants thereof. The term "probe" refers to DNA, preferably single-stranded, or RNA molecules or modifications or combinations thereof, that hybridize under stringent conditions, as defined above, to nucleic acid molecules comprised within the (methyl)griselimycin gene cluster identified by SEQ ID NO: 1 or variants thereof or encoding

10 any of the proteins of SEQ ID Nos: 2 to 27 or functionally active variants thereof, or their complementary or sense sequences. Generally, probes are significantly shorter than full-length sequences. They may contain from 5 to 100, preferably 10 to 80 nucleotides, more preferably 10 to 50 nucleotides, still more preferably 10 to 40 nucleotides and still more preferably 15 to 25 nucleotides. In particular, such probes may have sequences that are at least 70%, at least 75%, pref-

15 erably at least 85%, more preferably at least 95% and most preferably 100% homologous to a coding (ORFs 1 to 26) or non-coding sequence as comprised by SEQ ID NO: 1 or that are, to the above extents, complementary thereto. They may contain modified bases such as inosine, methyl-5-deoxycytidine, deoxyuridine, dimethylamino-5-deoxyuridine or diamino-2,6-purine. Sugar or phosphate residues may also be modified or substituted as is known in the art. For ex-

20 ample, a deoxyribose residue may be replaced by a polyamide and a phosphate residue may be replaced by ester groups such as diphosphate, alky, arylphosphonate or phosphorothioate esters. Alternatively or in addition, the 2`-hydroxyl group on ribonucleotides may be modified by including such groups as alkyl, O-alkyl or halogen groups. Probes of the invention are used in any conventional hybridization technique such as dot blot, Southern blot, northern blot or sandwich

25 technique which is a technique using specific capture and/or detection probes with nucleotide sequences that at least differ partially from each other (Sambrook et al., Molecular cloning: A laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2001). A primer is used to initiate enzymatic polymerization of DNA in an amplification (e.g. PCR), elongation or reverse transcription method. Primers used in diagnostic methods involving PCR are

30 labeled by methods known in the art. They can also be used as probes. In principle, primers may have the same characteristics as probes. They may contain from 5 to 100, preferably 10 to 80 nucleotides, more preferably 10 to 50 nucleotides, still more preferably 10 to 40 nucleotides and

still more preferably 15 to 25 nucleotides. In the context of the present invention, the primer or probe can be used for detecting, identifying and/or isolating a microorganism which produces (methyl)griselimycin, in particular which comprises any of the proteins of SEQ ID Nos: 2 to 27 or functionally active variants thereof, for detecting, identifying and/or isolating DNA or RNA encoding proteins of SEQ ID Nos: 2 to 27 or functionally active variants thereof and and/or for amplifying DNA or RNA encoding proteins of SEQ ID Nos: 2 to 27 or functionally active variants thereof. Thus, in an embodiment of the present invention a reagent is comprised comprising a primer or probe for the purposes of detection, identification and/or purification as outlined above.

Alternatively, identification of a nucleic acid encoding a protein which has activity in the (methyl)griselimycin biosynthetic pathway may be achieved by screening for reactivity or cross-reactivity with an antibody raised against the protein having an amino acid sequence of SEQ ID Nos: 2 to 27 or a functionally active variant or fragment thereof. The procedure is as follows: an antibody is specifically raised against a protein of SEQ ID Nos: 2 to 27 or a functionally active variant or fragment thereof, a fusion polypeptide (for example, an expression product of MBP, GST, or His-tag systems), or a synthetic peptide derived from the protein of SEQ ID Nos: 2 to 27 or a known variant or fragment thereof. Specific antigenicity can be determined according to a number of methods, including Western blot, dot blot, and ELISA. Once a specific antibody has been raised this antibody can be used to identify nucleic acids encoding proteins harbouring the immunogenic peptide against which the antibody has been raised thus identifying proteins which are proteins of SEQ ID Nos: 2 to 27 or functionally active fragments or variants thereof as defined herein. Specific antigenicity or specific antibody as used herein means that the antibody binds only to the immunogenic peptide against which the antibody was raised.

In a further embodiment, the nucleic acid is complementary to a nucleic acid sequence as comprised by SEQ ID NO: 1 or encoding a protein with SEQ ID Nos: 2 to 27 or functionally active fragments or variants thereof.

In one embodiment, the nucleic acid comprises at least one, at least two, at least three, at least four, at least five or more ORFs selected from ORFs 1 to 26 encoding polypeptides shown in SEQ ID Nos: 2 to 27 of the griselimycin biosynthesis gene cluster or functionally active frag-

ments or variants thereof. In a preferred embodiment, the nucleic acid comprises a specific ORF or a combination of specific ORFs which is/are involved in and function together in defined steps during the biosynthesis of (methyl)griselimycin. In a more preferred embodiment, a combination of ORFs selected from ORFs 1 to 27 is provided which encodes polypeptides which

5   manage the biosynthesis of (methyl)griselimycin. In a still more preferred embodiment, the ORF is ORF 8, 15 or 16 or the combination comprises at least two of ORFs 8, 15 or 16 designated nrps-1, nrps-2 and nrps-3, respectively, encoding subunits of the non-ribosomal protein synthetase (NRPS). In another embodiment, the nucleic acid comprises ORF 18, 19 or 21 or a combination of at least two of ORFs 18 to 21 constituting genes designated mps-1, mps-2, mps-3 and

10   mps-4, respectively, or at least two of ORFs 18, 19 and 21 constituting genes designated mps-1, mps-2 and mps-4, respectively, which are necessary for the biosynthesis of L-trans-4-methylproline. In another embodiment, ORFs 8, 15 and/or 16 may be combined with ORFs 18, 19, 20 and/or 21 or with ORFs 18, 19 and/or 21 to enhance production of L-trans-4-methylproline and of (methyl)griselimycin, in particular of methylgriselimycin. Thereby, any of

15   ORFs 1 to 26 may be combined with any of a variant or fragment of ORFs 1 to 26 to direct synthesis of (methyl)griselimycin. In another embodiment of this invention, the ORFs may be unchanged, but the control elements (e. g. promoters, ribosome binding sites, terminators, enhancers etc.) may be modified.

20   In another embodiment, a nucleic acid is provided comprising or consisting of the griselimycin gene cluster having the sequence of SEQ ID NO: 1 or having a variant sequence of SEQ ID NO: 1 harboring a variant or fragment of at least one of ORFs 1 to 26 whereby the variant or fragment encodes a functionally active variant or fragment of a protein of SEQ ID Nos: 2 to 27. Variant or fragmental sequences of any of ORFs 1 to 26 within the (methyl)griselimycin gene cluster of

25   SEQ ID NO: 1 in an otherwise unmodified state may result in one or more variant proteins of SEQ ID Nos: 2 to 27, which modify a specific step of the synthesis pathway of (methyl)griselimycin. This may result in an enhanced production of (methyl)griselimycin or L-trans-4-methylproline or in a modified structure of (methyl)griselimycin with enhanced antibiotic activity. As an example, any of the ORFs 8, 15 and 16 and/or 18, 19, 20 and 21 may be

30   modified within the griselimycin cluster as comprised by SEQ ID NO: 1 as referred to above, whereas the remaining ORFs remain unmodified. Alternatively or additionally, the control elements may be modified resulting in a modified expression of at least one of the ORFs 1 to 26.

15

Thus, the present invention provides in one embodiment a nucleic acid comprising at least one ORF being involved in the biosynthesis of (methyl)griselimycin such as ORFs 18, 19, 20 and 21 or otherwise designated ORFs 18 to 21 or ORFs 18, 19 and/or 21 and/or ORFs 8, 15 and/or 16 or

5   functionally active variants or fragments of these ORFs as defined herein. One or a combination of ORFs may be expressed or overexpressed in a heterologous cell in order to allow production of the respective proteins which may, e.g., be used as a fermentative aid to enhance the production of griselimycin and especially methygriselimycin in an organism endogenously or heterologously producing griselimycin and/or methygriselimycin. Alternatively, an ORF or a combina-

10 tion of ORFs as comprised by the invention is expressed or overexpressed in a cell or organism producing (methyl)griselimycin in order to enhance the production of the respective protein(s) in order to enhance production of (methyl)griselimycin and/or to favor the production of methylgriselimycin. For example, ORFs 18 to 21 or ORFs 18, 19 and 21 may be concomitantly expressed to allow synthesis of L-trans-4-methylproline. Enhanced production of L-trans-4-

15 methylproline may serve to enhance production of methylgriselimycin versus the production of griselimycin as could be demonstrated by feeding 4-methyl proline whereby five fold higher amounts of methylgriselimycin were obtained. Thus, in a preferred embodiment of the present invention, the yield of methylgriselimycin can be enhanced in *Streptomyces DSM 22643* or another organism producing (methyl)griselimycin by introducing a nucleic acid comprising genes

20 involved in the biosynthesis of L-trans-4-methylproline, in particular ORFs 18 to 21, more particularly ORFs 18, 19 and 21, according to methods as known in the art or as described herein and allowing expressing said genes and the genes of the (methyl)griselimycin gene cluster. Alternatively or additionally, at least one nrps nucleic acid (ORFs 8, 15 and/or 16) may be expressed for the synthesis of (methyl)griselimycin in a heterologous organism or in an organism

25 endogenously expressing (methyl)griselimycin for enhancing expression thereof. Alternatively or additionally, variant nucleic acids may be expressed which modulate the synthesis of (methyl)griselimycin in order to enhance the synthesis of (methyl)griselimycin, to favor the synthesis of methylgriselimycin or to generate derivatives of (methyl)griselimycin with a composition of amino acids different from that of (methyl)griselimycin which are, for example, more efficient

30 antibiotics. For example, the amino acid specificity of the A-domains of NRPS proteins may be altered by mutagenesis or by domain swapping between peptide synthetases thereby allowing the selection of different amino acids and generating derivatives of a natural (methyl)griselimycin.

16

Alternatively, the nucleic acid encoding the A-domain of module 8 of NRPS2 may be mutated to produce a variant nucleic acid which results in the enhanced insertion of L-trans-4-methylproline into the nascent polypeptide chain, thus resulting in an enhanced production of methylgriselimycin compared to the wild-type. Alternatively, the complete module 8 may be swapped or the C-

5   domain of module 8 may be swapped or mutated to preferably condense methylproline.

L-trans-4-methylproline is comprised by griselimycin at positions 2 and 5 and by methylgriselimycin at positions 2, 5 and 8. L-trans-4-methylproline can be chemically produced, however, the chemical production is cost-intensive and time-consuming. Thus, the present invention provides

10   in one aspect a nucleic acid comprising genes being involved in the biosynthesis of L-trans-4-methylproline, in particular ORFs 18 to 21 or ORFs 18, 19 and 21 or functionally active variants or fragments of these ORFs as defined herein. The combination of these ORFs may be expressed or overexpressed in a heterologous cell in order to allow production of L-trans-4-methylproline for use, e.g., as a fermentative aid to enhance the production of griselimycin and especially me-

15   thygriselimycin in an organism endogenously or heterologously producing griselimycin and/or methylgriselimycin. Alternatively, the combination of these genes is expressed or overexpressed in a cell or organism naturally producing (methyl)griselimycin in order to enhance the production of L-trans-4-methylproline in order to favor the production of methylgriselimycin. Thus, in a preferred embodiment of the present invention, the yield of methylgriselimycin can be enhanced

20   in *Streptomyces DSM 22643* by introducing a nucleic acid comprising genes involved in the biosynthesis of L-trans-4-methylproline, in particular ORFs 18 to 21, more particularly ORFs 18, 19 and 21, into in an organism endogenously or heterologously producing griselimycin and/or methylgriselimycin, e.g. *Streptomyces DSM 22643* according to methods as known in the art or as described herein and allowing expressing said genes and the genes of the griselimycin biosynthe-

25   sis gene cluster.

The nucleic acid of the present invention may be provided by any methods known in the art. Using the sequence information provided herein, primers suitable for amplification/isolation of one or more ORFs can be determined according to standard methods well known to those of skill in

30   the art. Primers suitable for amplification/isolation of any one or more of the ORFs as defined herein are designed according to the nucleotide sequence information provided in the sequence listing. The procedure is as follows: a primer is selected which may consist of 10 to 40, prefera-

17

bly 15 to 25 nucleotides. It is advantageous to select primers containing C and G nucleotides in a proportion sufficient to ensure efficient hybridization; i.e., an amount of C and G nucleotides of at least 40%, preferably 50% of the total nucleotide content. Typically such amplifications will utilize the DNA or RNA of an organism containing the requisite genes (e. g. *Streptomyces, in particular Streptomyces DSM 22643*) as a template. A standard PCR reaction will be performed which typically contains 0.5 to 5 Units of Taq DNA polymerase per 100 µl, 20 to 200 µM deoxynucleotide each, preferably at equivalent concentrations, 0.5 to 2.5 mM magnesium over the total deoxynucleotide concentration, 105 to 106 target molecules, and about 20 pmol of each primer. About 25 to 50 PCR cycles are performed. A more stringent annealing temperature improves discrimination against incorrectly annealed primers and reduces incorporation of incorrect nucleotides at the 3' end of primers. A denaturation temperature of 95°C to 97°C is typical, although higher temperatures may be appropriate for denaturation of G+C-rich targets. The number of cycles performed depends on the starting concentration of target molecules, though typically more than 40 cycles are not recommended as non-specific background products tend to accumulate. An alternative method for retrieving polynucleotides encoding variant polypeptides as defined herein is by hybridization screening of a DNA or RNA library using the primers and probes as defined herein. Hybridization procedures are well-known and are described in the art and herein.

Alternatively or additionally to the above the nucleic acid may be provided by cloning and thereby introducing it into and amplifying it in a cell. Thus, in a further aspect of the present invention, a vector comprising a nucleic acid as defined herein and a host cell transformed with the expression vector are provided. The procedure of introducing a gene into a recipient cell is called transformation. The genes can be introduced into the cells by a variety of means known in the art and adapted to each cell type. The term "cell" or "host cell" refers to the cell in which the gene is expressed irrespective of whether it is a prokaryotic cell or a eukaryotic cell and of whether the cell naturally expresses the respective genes or not. Thereby the cell may be, in a preferred embodiment, a cell which naturally harbors the gene expressing the protein as comprised by the present invention, e.g., *Streptomyces DSM 22643*. Recombinant DNA cloning techniques well known in the art for introducing and expressing a nucleic acid molecule can be used to introduce and express the gene which is either endogenous if the cell harbours the respective gene or is heterologous if the gene is not endogenous to the cell. Cells can be transformed using any appro-

18

priate means, including viral or bacteriophage based vectors, chemical agents, electroporation, calcium phosphate co-precipitation or direct diffusion of DNA. Vectors are agents that transport an endogenous or heterologous gene into the cell and may include appropriate transcriptional and translational control signals such as a promoter. Vectors can be a plasmid, a virus (e. g. bacterio-

5   phage) or others as known in the art. Vectors are able to autonomously replicate in a host cell or can be incorporated into chromosomal DNA. The term "vectors" includes those that function primarily for insertion of a nucleic acid into a cell, those that function primarily for replication of a nucleic acid (replication vector) in a cell or those that function primarily for transcription and/or translation of DNA or RNA in a cell. Examples of vectors include pBTrp2, pBTac1,

10   pBTac2 (all of which are manufactured by Boehringer Mannheim), pKK263-2 (manufactured by Pharmacia), pGEX (manufactured by Pharmacia), pSE280 (manufactured by Invitrogen), pGE-MEX-1 (manufactured by Promega), pQE-8 (manufactured by Qiagene), pET-3 (manufactured by Novagen), pBluescriptII SK+ (manufactured by Stratagene), pBluescript II SK (-) (manufac-tured by Stratagene), pTrS30 [prepared from *Escherichia coli* JM109/pTrS30 (FERM BP-

15   5407)], pTrS32 [prepared from *Escherichia coli* JM109/pTrS32 (FERM BP-5408)], pSTV28 (manufactured by Takara Bio Inc.), pUC118 (manufactured by Takara Bio Inc.), pHW1520 (manufactured by MoBiTec), pSET152, pOJ436 and pOJ446 (Bierman M, et al., 1992), pSH19 (Herai S, et al., 2004), pUWL199, pUWL218 and pUWL219 (Wehmeier U. F., 1995) and pIJ6021 (Takano E. et al., 1995).

20

The promoter can be inducible or constitutive, general or cell specific, nuclear or cytoplasmic specific, heterologous or associated with the gene in nature. Any type of promoter can be used, as long as it functions in host cells. Examples of the promoter include promoters derived from *Escherichia coli* or phage, such as a trp promoter (Ptrp), a lac promoter (Plac), a PL promoter, a

25   PR promoter or a PSE promoter, a SPO1 promoter, a SPO2 promoter, and a penP promoter. In addition, artificially designed or modified promoters such as a promoter formed by placing two Ptrp in series (Ptrp*2), a tac promoter, a lacT7 promoter or a let I promoter, can also be used. Moreover, a xylA promoter for expression in the bacteria of the genus *Bacillus*, or a P54-6 pro-moter for expression in the bacteria of the genus *Corynebacterium* can also be used. Further use-

30   ful promoters are PermE (Bibb et al., 1985, PermE* (Bibb et al., 1994), PtipA (Murakami et al., 1989), P*nitA*-NitR expression system (Herai et al., 2004) and actII-ORF4/PactI activator-promoter system (Ferna´ndez-Moreno et al., 1991). Selection of promoters, vectors and other

19

elements is a matter of routine design. Many such elements are described in literature and are available through commercial suppliers. A single gene can be introduced into a cell. Also, more than one gene can be introduced into a cell and expressed therein. Where large clusters are to be expressed, it is preferable that phagemids, cosmids, P1s, YACs, BACs, PACs, HACs, or similar cloning vectors are used. If more than one gene is introduced into a cell, then the genes may be under the regulation of the same promoter and/or regulatory elements. Alternatively, the genes may be under the regulation of different promoter and/or regulatory elements. Usually, the method of transfer includes transfer of a selectable marker to the cells. In general, a cell line is transformed by any of the means mentioned above wherein the transgene is operatively linked to a selectable marker. Following transformation, cells are grown for an adapted period of time. Transformed cells exhibit resistance to the selection and are able to grow, whereas non-transformed cells die in general. Examples for selective markers include puromycin, zeocin, neomycin and hygromycin B which confer resistance to puromycin, zeocin, aminoglycoside G-418 and hygromycin B, respectively.

Examples of host cells suitable in the context of the present invention are derived from any organism with the capability of harboring and expressing a recombinant griselimycin biosynthesis gene cluster or one or more genes of the griselimycin biosynthesis gene cluster. Examples include bacteria, yeasts, filamentous fungi, animal cells and plant cells, such as, without limitation, cells of *E. coli* strains, of the order *Actinomycetales* such as a *Streptomyces species* such as *Streptomyces DSM 22643* or *Streptomyces coelicolor*, of yeast strains such as *Saccharomyces cerevisiae*, of the insect cell line Lepidoptera such as from *Spodoptera frugiperda*, of plant cells or of mammal cells such as L6 cells, 3T3 adipocytes, HEK 293, 745-A, A-431, atrial myocytes, BxPC3, C5N, Caco-2, Capan-1, CC531, CFPAC, CHO, CHO K1, COS-1, COS-7, CV-1, EAHY, EAHY 926. Preferred embodiments are bacterial cells such as cells of the order *Actinomycetales* such as *Streptomyces DSM 26643* or *Streptomyces coelicolor* or cells of *E. coli* strains. Particularly preferred embodiments are cells of the order *Actinomycetales*, in particular *Streptomyces DSM 26643*.

As stated above, preferred host cells for harboring and expressing genes as comprised by the present invention are bacterial cells. Other host cells suitable for harboring and expressing genes as comprised by the present invention are established or immortalized cell lines which have ac-

quired the ability to proliferate indefinitely either through random mutation or deliberate modification, such as artificial expression of the telomerase gene. There are numerous well established cell lines representative of particular cell types and it is within the knowledge of the skilled person to select a suitable cell line. A cell line is a population of cells propagated in culture that are derived from, and therefore genetically identical to, a single common ancestor cell. Suitable cell lines for the present invention are HEK 293 cells (primary human embryonic kidney), 3T3 cells (murine embryonic fibroblasts), CHO cells (Chinese hamster ovary), COS-7 cells (African green monkey cell line), HeLa cells (human epithelioid cervical carcinoma), JURKAT cells (human T-cell leukemia), BHK 21 cell (hamster normal kidney, fibroblast), and MCF-7 cells (human breast cancer).

Alternatively, the nucleic acids as defined herein may be cloned and expressed in a tissue culture. The term "tissue culture" refers to a method in which a group of cells forming a three dimensional network is cultivated. The tissue can be formed in culture by the cells themselves or may be formed by an extracellular matrix produced by the cells, the culture may be cultivated on a natural or an artificial extracellular matrix (e.g. collagen, elastin, polystyrene, nylon, polylysine) or the tissue may be obtained from an animal including human. The tissue culture may be cultivated in culture medium and at suitable temperature. The culture medium may contain nutrients (e.g. sugars, salts, amino acids, and lipids), a buffer system (often comprising one or more chemical buffer substances such as phosphates, hydrogen phosphates and/or Tris with low or no toxicity and/or carbon dioxide (CO_2) gassing) and/or optionally one or more antibiotics. In order to provide sufficient amounts of nutrients and/or to remove metabolites the medium may be changed when required. Optionally, the tissue culture may be perfused with medium. The perfusion may be a permanent or pulsed perfusion.

Variations may be introduced into a nucleic acid by design, for example, by modification of a nucleotide within a nucleic acid in a specific way, e. g. by replacing a single substituent by another and thereby resulting in a partly or complete artificelle biosynthesis gene cluster for griselimycin . As a further example, the skilled person knows the field of the art termed "synthetic biology" which deals with the design and construction of new biological functions and systems not found in nature. For example, the skilled person is capable of designing a DNA sequence in silico and synthesizing it having no similarity to the nucleic acid level of the wildtype

sequence, however, coding for the same protein(s) as the wildtype sequence. Alternatively, variations can be made randomly, for example, by making a library of molecular variants of (methyl)griselimycin by systematically or haphazardly replacing one or more ORFs in the bio-synthetic pathway. Moreover, useful variants and fragments of a nucleic acid that do not occur naturally are designed using known methods. Such variation may be any modification as out-lined above, e.g. replacement of a single substituent by another. For example, regions of a poly-peptide that are likely to tolerate amino acid sequence changes and/or deletions are identified. As an example, variant polypeptides involved in the synthesis of (methyl)griselimycin from differ-ent species producing (methyl)griselimycin are compared; conserved sequences are identified. The more divergent sequences are the most likely to tolerate sequence changes. Homology among sequences may be analyzed using methods known in the art. Mutation introduction meth-ods into a nucleic acid include error-prone PCR, site-directed mutagenesis, a method using hydroxylamine or a method of allowing a mutation agent such as UV to act on cells having the nucleic acid of interest. When an error-prone PCR is used, e.g. a plasmid into which the nucleic acid of interest has been incorporated is used as a template, the PCR reaction is carried out in a reaction solution wherein a manganese (Mn) salt concentration is higher than in a usual PCR reaction.

Using the information provided herein other approaches to cloning the desired sequences will be apparent to those of skill in the art, for example, at least one of the ORFs 1 to 26 including the total griselymcin biosynthesis gene cluster comprising ORFs 1 to 26 and/or optionally nucleic acids encoding NRPS modules or enzymatic domains of interest can be obtained from an organ-ism that expresses such, using recombinant methods, such as by screening cDNA or genomic libraries derived from cells expressing the gene(s) or by deriving the gene(s) from a vector known to include the same. The gene(s) or cluster can then be isolated and combined with other desired biosynthetic elements using standard techniques. If the gene(s) or cluster in question is(are) already present in a suitable expression vector, it(they) can be combined *in situ* with, e. g. other domains or subunits, as desired. The gene(s) of interest can be produced synthetically, rather than cloned. The nucleotide sequence can be designed with the appropriate codons for the particular amino acid sequence desired. In general, one will select preferred codons for the in-tended host in which the sequence will be expressed. In addition, it is noted that custom gene synthesis is commercially available.

22

In another aspect, the present invention is directed to a protein, comprising at least one amino acid sequence selected from SEQ ID Nos: 2 to 27 or functionally active fragments or variants thereof or encoded by a nucleic acid as referred to above. The protein may be a single protein or may be a mixture of proteins or may be a fusion protein of at least two proteins of SEQ ID Nos: 2 to 27. The protein may be isolated which means that it is substantially pure and not associated with components with which it is associated in nature or the protein is present in a lysate of a cell producing the protein. The proteins of SEQ ID Nos: 2 to 27 are encoded by the putative ORFs 1 to 26, respectively. These ORFs belong to and constitute the griselimycin biosynthesis gene cluster. Based on sequence analysis and homology searches, the present inventors succeeded in identifying the griselimycin biosynthesis gene cluster, in identifying specific putative ORFs within the gene cluster and in assigning the identified ORFs to either a specific protein function and/or to a specific homologous nucleotide sequence coding for a known or hypothetical protein.

The proteins as comprised by the present invention encompass proteins of SEQ ID Nos: 2 to 27 as encoded by the griselimycin biosynthesis gene cluster of *Streptomyces DSM 26643* of the sequence of SEQ ID NO: 1 and encompass proteins as they occur in other *Streptomyces* species and strains and non-*Streptomyces* species which produce (methyl)griselimycin which are orthologs or homologs whereby these orthologs or homologs have the same function as the proteins of SEQ ID NO: 2 to 27 of *Streptomyces DSM 26643*. Preferably, orthologs or homologs thereof differ from the sequences of SEQ ID Nos: 2 to 27 e.g. by addition, deletion, substitution and/or insertion of amino acids and have a sequence identity with SEQ ID Nos: 2 to 27 of more than 50%, of more than 60%, more than 70%, preferably of more than 80%, more preferably more than 85%, even more preferably more than 90%, even more preferably more than 95%, most preferably more than 97% and/or have an enzyme activity of more than 5%, more than 10%, more than 20%, more than 30%, more than 40%, more than 50%, more than 60%, more than 70%, more than 80%, more than 90%, more than 95%, more than 97% or more than 100 % , e.g. more than 150 %, 200 %, 300 % 400 % or 500 % of the activity of the proteins of SEQ ID Nos: 2 to 27.

In the context of the present invention the naturally or non-naturally occurring variant of the proteins of SEQ ID Nos: 2 to 27 as comprised by the present invention is a functionally active pro-

23

tein in that it maintains the biological function of the reference protein of SEQ ID Nos: 2 to 27, i.e. the involvement in a reaction in which the reference protein is involved under natural conditions (in case of a non-natural variant, the biological function of the reference protein) as referred to in the respective chapters disclosing the function of the protein of SEQ ID Nos: 2 to 27, and can be substituted therefor, as defined herein.

Non-naturally occurring variants of the proteins of SEQ ID Nos: 2 to 27 or of naturally occurring variants thereof may be obtained by a limited number of amino acid deletions, insertions and/or substitutions, particularly deletions, insertions and/or substitutions of, e.g., at most 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 amino acid(s) thereby obtaining a sequence identity or activity of the respective wild-type proteins, e.g. with respect to SEQ ID Nos: 2 to 27, as mentioned above.

The variant may be a modified protein or a modified protein variant which comprises a further component. Accordingly, the variant may be a molecule having a domain composed of a naturally occurring protein of SEQ ID Nos: 2 to 27 or a variant thereof as detailed herein and at least one further component. In one preferred embodiment the variant may be a fusion protein comprising (i) a protein of SEQ ID Nos: 2 to 27 or functionally active variant and (ii) a further protein or peptide component. For example, the protein may be coupled to a marker, such as a tag used for purification purposes (e.g. 6 His (or HexaHis) tag, Strep tag, HA tag, c-myc tag or glutathione S-transferase (GST) tag). If e.g. a highly purified protein of SEQ ID Nos: 2 to 27 or variant should be required, double or multiple markers (e.g. combinations of the above markers or tags) may be used. In this case the proteins are purified in two or more separation chromatography steps, in each case utilizing the affinity of a first and then of a second tag. Examples of such double or tandem tags are the GST-His-tag (glutathione-S-transferase fused to a polyhistidine-tag), the 6xHis-Strep-tag (6 histidine residues fused to a Strep-tag), the 6xHis-tag100-tag (6 histidine residues fused to a 12-amino-acid protein of mammalian MAP-kinase 2), 8xHis-HA-tag (8 histidine residues fused to a hemagglutinin-epitope-tag), His-MBP (His-tag fused to a maltose-binding protein, FLAG-HA-tag (FLAG-tag fused to a hemagglutinin-epitope-tag), and the FLAG-Strep-tag. The marker could be used in order to detect the tagged protein, wherein specific antibodies could be used. Suitable antibodies include anti-HA (such as 12CA5 or 3F10), anti-6 His, anti-c-myc and anti-GST. Furthermore, the protein could be linked to a marker of a different category, such as a fluorescence marker such as green fluorescent protein, to a binding

24

protein such as steptavidin, one or more small molecular dyes such as Cy dye, or a radioactive marker, which allows for the detection of a protein as comprised by the present invention. In a further embodiment, the proteins of SEQ ID Nos: 2 to 27 could be part of a fusion protein, wherein the second part could be used for detection, such as a protein component having enzy-

5   matic activity.

In another embodiment of the present invention, the variant of the proteins of SEQ ID Nos: 2 to 27 could be a fragment, wherein the fragment is still functionally active. This may include proteins of SEQ ID Nos: 2 to 27 or variants thereof as detailed above with short internal and/or C-

10   and/or N-terminal deletions (e.g. deletions of at most 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6 5, 4, 3, 2, or 1 amino acids within the variant and/or at the C- and/or N-termini or total deletions of 5%, 10%, 20%, 30%, 40%, 50%, 60% 70% amino acids or any values in between these values). Additionally, the fragment may be further modified as detailed above for the proteins of SEQ ID NO: 2 to 27.

15

Alternatively or additionally, the proteins of SEQ ID Nos: 2 to 27 or variants thereof as described above may comprise one or more amino acid substitution(s). However, semi-conservative and especially conservative amino acid substitutions, wherein an amino acid is substituted with a chemically related amino acid are preferred. Typical substitutions are among the

20   aliphatic amino acids, among the amino acids having aliphatic hydroxyl side chain, among the amino acids having acidic residues, among the amide derivatives, among the amino acids with basic residues, or the amino acids having aromatic residues. Typical semi-conservative and conservative substitutions are:

Amino acid	Conservative substitution	Semi-conservative substitution
A	G; S; T	N; V; C
C	A; V; L	M; I; F; G
D	E; N; Q	A; S; T; K; R; H
E	D; Q; N	A; S; T; K; R; H
F	W; Y; L; M; H	I; V; A
G	A	S; N; T; D; E; N; Q
H	Y; F; K; R	L; M; A

25

I	V; L; M; A	F; Y; W; G
K	R; H	D; E; N; Q; S; T; A
L	M; I; V; A	F; Y; W; H; C
M	L; I; V; A	F; Y; W; C;
N	Q	D; E; S; T; A; G; K; R
P	V; I	L; A; M; W; Y; S; T; C; F
Q	N	D; E; A; S; T; L; M; K; R
R	K; H	N; Q; S; T; D; E; A
S	A; T; G; N	D; E; R; K
T	A; S; G; N; V	D; E; R; K; I
V	A; L; I	M; T; C; N
W	F; Y; H	L; M; I; V; C
Y	F; W; H	L; M; I; V; C

Changing from A, F, H, I, L, M, P, V, W or Y to C is semi-conservative if the new cysteine remains as a free thiol. Furthermore, the skilled person will appreciate that glycines at sterically demanding positions should not be substituted and that P should not be introduced into parts of the protein which have an alpha-helical or a beta-sheet structure. The proteins of SEQ ID Nos: 2 to 26 or fragments or variants thereof with substitution may be modified as detailed above for the proteins of SEQ ID Nos: 2 to 27 or fragments or variants thereof.

It is noted that the above modifications of the protein of SEQ ID Nos: 2 to 27 may be combined. The variants of the present invention may be e.g. a fragment of a protein of SEQ ID Nos: 2 to 27 having a marker fused to it or comprising one or more amino acid substitutions. It is furthermore noted that any of the proteins of SEQ ID Nos: 2 to 27 may be combined with any of a variant or fragment of the proteins of SEQ ID Nos: 2 to 27.

The proteins as comprised by the present invention may be provided by any means, e. g., by obtaining the proteins from a natural source, by biotechnological methods well-known in the art and/or by synthetic methods. Herein the term "providing" may be understood in the widest sense and may include but may not be limited to any biochemical or biotechnological means as known in the art that may be used to obtain the protein or other product. Consequently, the protein may

26

be obtained by extraction from a natural source such as a microorganism containing the proteins, such as a *Streptomyces* species, in particular *Streptomyces DSM 26643*. The skilled person will know methods of how to isolate the protein. Cells naturally harbouring a protein as comprised by the present invention may be disrupted by any means known in the art but not limited to sonica-

5   tion, hypotonic buffer, detergents, UltraTurrax, French press, freeze-thraw cycles, mechanical homogenization and scratching. The protein of interest may then be isolated by known methods, e.g., a solvent extraction method, a salting-out method, a solvent precipitation method, a dialysis method, an ultrafiltration method, a gel electrophoresis method, a gel filtration chromatography, an ion exchange chromatography, a reverse phase chromatography, and an affinity chromatogra-

10   phy, either alone or in combination as appropriate.

Alternatively, the proteins as comprised by the present invention may be proteins which are produced by recombinant methods, e.g. by cloning and expressing a nucleic acid encoding a protein as comprised by the present invention or a functional fragment or variant thereof and isolating

15   the protein. Procedures of cloning nucleic acids and of isolating the expressed proteins are as disclosed above.

The proteins may be expressed with a convenient marker to facilitate isolation. A marker (or tag or label) is any kind of substance which is able to indicate the presence of another substance or

20   complex of substances. The marker can be a substance that is linked to or introduced in the substance to be detected. Detectable markers are used in molecular biology and biotechnology to detect e.g. a protein, a product of an enzymatic reaction, a second messenger, DNA, interactions of molecules etc. Examples of suitable marker or labels include a fluorophore, a chromophore, a radiolabel, a metal colloid, an enzyme, or a chemiluminescent or bioluminescent molecule. Ex-

25   amples of fluorophores include fluorescein, rhodamine, and sulfoindocyanine dye Cy5. Examples of radiolabels include ^{3}H, ^{14}C, ^{32}P, ^{33}P, ^{35}S, ^{99m}Tc or ^{125}I. Examples of enzymes include horseradish peroxidase, alkaline phosphatase, glucose oxidase, and urease.

The proteins of the present invention including functionally active variants and fragments thereof

30   can be singly expressed in a cell or several proteins including functionally active variants and fragments thereof as comprised by the present invention such as at least two, three, four, five or more proteins can be simultaneously expressed in a cell. Also, the whole griselimycin biosynthe-

sis gene cluster may be heterologously expressed. Thereby, any combination of variant and non-variant protein is comprised. In one embodiment, several proteins which contribute to a certain aspect of the synthesis pathway of (methyl)griselimycin such as those which belong to a same family of proteins such as the non-ribosomal protein synthetases encoded by the nrps-1, nrps-2

5   and/or nrps-3 genes or such as the proteins encoded by the methylproline synthesis genes mps-1, mps-2, mps-3 and mps-4 or mps-1, mps-2 and mps-4 may be expressed together in a cell. Thereby, a specific step of the synthesis of (methyl)griselimycin is especially favored so that production of (methyl)griselimycin and in particular of methylgriselimycin is enhanced. Alternatively, variants may be produced which influence a specific aspect of the synthesis pathway of

10   (methyl)griselimycin and thus the synthesis of (methyl)griselimycin. Examples are variants of one or more of the proteins of SEQ ID Nos: 9, 16 and 17, in particular SEQ ID NO: 16, more particularly module 8 of SEQ ID NO: 16 which promote insertion of L-trans-4-methylproline into methylgriselimycin and/or SEQ ID Nos: 19, 20 and 22 and possibly SEQ ID NO: 21 which enhance the synthesis of L-trans-4-methylproline so that methylgriselimycin is produced at

15   higher yield than if not using such variants. One or more of the variants may be combined with one or more non-variant proteins of SEQ ID Nos: 2 to 27.

In a further aspect of the present invention an antibody against at least one of the proteins of SEQ ID Nos: 2 to 27 or a functionally active fragment or variant thereof is provided. The detec-

20   tion of protein often involves the use of specific antibodies. Accordingly, the detection of at least one of the proteins of SEQ ID Nos: 2 to 27 or a functional variant or fragment thereof may include a specific antibody. Antibodies can be raised using well established techniques for immunizing animals with prepared forms of the antigen. A variety of reagents is available to assist in antibody production and purification, and various companies specialize in antibody production

25   services. Depending on the application to be performed, different levels of purity and types of specificity are needed in a supplied primary antibody. To name just a few parameters, antibodies may be monoclonal or polyclonal, supplied as antiserum or affinity-purified solution, and validated for native protein or denatured protein detection. They may be chimeric, humanized, multifunctional, bispecific or oligospecific antibodies. Antibodies as comprised by the present inven-

30   tion may include whole antibodies or antibody fragments such as Fab, F(ab)$_2$ or sc (single chain) Fv.

28

An antibody that recognizes the target antigen, here the proteins of SEQ ID Nos: 2 to 27 or a functional variant or fragment thereof, is called the "primary antibody." If this antibody is labeled with a tag, direct detection of the antigen is possible. Usually, however, the primary antibody is not labeled for direct detection. Instead a "secondary antibody" that has been labeled with a detectable tag is applied in a second step to probe for the primary antibody, which is bound to the target antigen. Thus, the antigen is detected indirectly. Another form of indirect detection involves using a primary or secondary antibody that is labeled with an affinity tag such as biotin. Then a secondary (or tertiary) probe, such as streptavidin that is labeled with the detectable enzyme or fluorophore tag, can be used to probe for the biotin tag to yield a detectable signal. Several variants of these probing and detection strategies exist. However, each one depends on a specific probe (e.g., a primary antibody) whose presence is linked directly or indirectly to some sort of measurable tag (e.g., an enzyme whose activity can produce a colored product upon reaction with its substrate).

In a further aspect of the present invention, the use of the nucleic acid or the protein as defined herein including fragments and variants, of an expression vector harboring the nucleic acid or of a host cell harboring said expression vectors for the production of (methyl)griselimycin is provided. Particularly, the use of at least one of the nucleic acids comprising ORFs 8, 15 and 16 or of at least one of the proteins of SEQ ID Nos: 9, 16 and 17 including fragments and variants for the production of (methyl)griselimycin is provided. Preferably, a combination of the nucleic acids comprising ORFs 8, 15 and 16 or of the proteins of SEQ ID Nos: 9, 16 and 17 is provided. With respect to the terms "nucleic acid", "protein", "ORFs 8, 15 and 16" and "proteins of "SEQ ID Nos: 9, 16 and 17," it is referred to the definitions provided in the context of the nucleic acids and proteins of the invention.

In a further aspect of the present invention, the use of at least one of the nucleic acids comprising ORFs 18, 19 and 21 or of at least one of the proteins of SEQ ID Nos: 19, 20 and 22 including fragments and variants for the production of L-trans-4-methylproline is provided. Preferably, a combination of the nucleic acids comprising ORFs 18, 19 and 21 or of the proteins of SEQ ID Nos: 19, 20 and 22 is provided. With respect to ORFs 18, 19 and 21 and the proteins of SEQ ID Nos: 19, 20 and 22 it is referred to the definitions provided in the context of the nucleic acids and proteins of the invention. With respect to the method of producing (methyl)griselimycin or L-

trans-4-methylproline reference is made to the definition thereof in the context of determining a variant of ORFs 1 to 26.

5   In another aspect of the invention, a method for producing at least one of the proteins or a functionally active variant or fragment thereof as defined herein comprising expressing the nucleic acid including variant forms and fragments as defined herein, and optionally isolating the protein or functionally active variant or fragment thereof is provided. Methods of cloning and expressing a nucleic acid and of isolating a protein are known in the art and are disclosed herein.

10   The present invention refers in a preferred embodiment to the above mentioned nucleic acid or protein which is isolated. The term "isolated" as used herein with respect to nucleic acids, such as DNA or RNA, refers to a nucleic acid which is not associated with all or a portion of a polynucleotide in which the isolated nucleic acid is found in nature. The term "isolated" as used herein with respect to proteins refers to a protein which is not associated with all or a portion of
15   compounds in which or with which the isolated protein is found in nature. The isolated nucleic acid or the isolated protein is separated from other cellular components which naturally accompany a native human sequence or protein, e.g., ribosomes, polymerases, many other human genome sequences and proteins and any other kind of cellular material. The term embraces a nucleic acid sequence or protein which has been removed from its naturally occurring environment
20   and includes recombinant or cloned DNA isolates or proteins derived therefrom and chemically synthesized analogs or analogs biologically synthesized by heterologous systems. An "isolated nucleic acid" is meant to include nucleic acid fragments which have been removed from its natural surroundings or which are not naturally occurring as fragments and would not be found in nature. The term "isolated" also refers to proteins which are isolated from other cellular proteins
25   and is meant to also encompass recombinant proteins. The term "isolated" does not mean "purified" which means that the nucleic acid or polypeptide or protein is purified from its natural requirement by only separating out impurities.

In another aspect of the invention a method of determining a variant of ORFs 1 to 26 or of the
30   proteins of SEQ ID Nos: 2 to 27 which modulates production of griselimycin and/or methyl-griselimycin is provided. The method comprises producing a variant of any of ORFs 1 to 26, expressing the variant into a protein, allowing the production of griselimycin and/or methyl-

griselimycin or of a derivative thereof using said variant protein and determining the amount of griselimycin and/or methylgriselimycin, wherein an increase of the amount of griselimycin and/or methylgriselimycin compared to the amount produced if the variant protein is not present, but the corresponding invariant form of the protein is present is indicative that the variant is ca-

5   pable of increasing the production of griselimycin and/or methylgriselimycin, or determining the antibiotic activity of the derivative of griselimycin and/or methylgriselimycin, wherein an increase in the antibiotic activity compared to the antibiotic activity if the variant protein is not present, but the corresponding invariant form of the protein is present is indicative that the variant is capable of producing a derivative of griselimycin and/or methylgriselimycin with in-

10   creased antibiotic activity and wherein the derivative differs in amino acid composition from griselimycin and/or methylgriselimycin and exhibits an enhanced antibiotic activity over griselimycin and/or methylgriselimycin. In a further aspect the invention comprises a method of determining a variant of ORFs 1 to 26 or of the proteins of SEQ ID Nos: 2 to 27 which enhances production of L-trans-4-methylproline comprising producing a variant of any of ORFs 1 to 26,

15   expressing the variant into a protein, allowing the production of L-trans-4-methylproline using said variant protein and determining the amount of L-trans-4-methylproline, wherein an increase of the amount of L-trans-4-methylproline compared to the amount if the variant protein is not present, but the corresponding invariant form of the protein is present is indicative that the variant is capable of increasing the production of L-trans-4-methylproline.

20

As used herein, the term "method", or "assay" may be understood in the widest sense as an experimental activity or procedure. It will be understood as all means that may be used to determine a variant of ORFs 1 to 26 or of a protein of SEQ ID Nos: 2 to 27 that is able to modulate or enhance the production of (methyl)griselimycin or L-trans-4-methylproline. The term "variant of

25   ORFs 1 to 26" may be understood in the sense as defined herein, namely variants of ORFs 1 to 26 with substitutions, additions, insertions and/or deletions of one or more nucleic acids encoding variants of one or more proteins of SEQ ID Nos: 2 to 27 as long as the variant encodes a protein which is capable of increasing the yield of griselimycin and/or methylgriselimycin, in particular methylgriselimycin, or of L-trans-4-methylproline by the cell harboring the variant or of

30   changing the composition of (methyl)griselimycin so that the (methyl)griselimycin with the changed composition shows enhanced antibiotic activity. Suitable variants are those which increase the yield of griselimycin and/or methylgriselimycin, in particular methylgriselimycin, or

of L-trans-4-methylproline by at least 5%, at least 10%, least 20%, least 30%, least 40%, least 50%, at least 100%, at least 200%, at least 300, at least 400% or at least 500%. Alternatively, suitable variants are those which change the composition of (methyl)griselimycin so that the antibiotic activity is increased by at least 5%, at least 10%, least 20%, least 30%, least 40%, least 50%, at least 100%, at least 200%, at least 300, at least 400% or at least 500%. The term "modulates the production of griselimycin and methylgriselimycin" means enhancing the amount of griselimycin and/or methylgriselimycin produced or varying the composition of griselimycin and/or methylgriselimycin. Thus, one or two amino acids or even more amino acids (e.g. three or four amino acids) or one or two or more domains or one or two or more modules may be replaced by different amino acids or domains or modules, respectively, whereby the resulting derivatives of griselimycin and/or methylgriselimycin show an enhanced antibiotic activity. For example, variants of nucleic acids coding for the A-domain of any of the modules of the NRPS proteins may result in the incorporation of different amino acids which may result in a derivative of (methyl)griselimycin which may possess enhanced antibiotic activity. The term "producing a variant" means the production by any method known in the art or as described herein. The method includes intentionally mutated nucleic acids by introducing targeted mutations or randomly mutated nucleic acids. Once a variant nucleic acid has been produced, the variant nucleic acid is introduced into a cell. Alternatively, the variant is produced inside the cell by applying mutating treatments on the cell. The cell may naturally harbor the genes for production of griselimycin and/or methylgriselimycin or of L-trans-4-methylproline or the cell may comprise a heterologous system for the production of griselimycin and/or methylgriselimycin or of L-trans-4-methylproline. The gene, of which a variant is to be expressed in the cell, is preferably inactive by deletion or any other kind of mutation or is comprised by the cell in an active state. For example, the cell may harbor each of the nucleic acid of ORFs 1 to 26 including one or more variant nucleic acids to be tested or a set of nucleic acids out of ORFs 1 to 26 such as the nrps genes including a variant nrps gene or another variant gene other than a variant nrps gene or the mps genes including a variant mps gene or another variant gene other than a variant mps gene. The cell harboring nucleic acids coding for proteins for synthesis of (methyl)griselimycin or L-trans-4-methylproline and one or more variant nucleic acids is incubated under appropriate conditions to allow synthesis of (methyl)griselimycin or a derivative thereof or L-trans-4-methylproline. Preferably, the production of (methyl)griselimycin or L-trans-4-methylproline is carried out in an aqueous buffer of neutral pH, more preferably in a buffer with a pH around pH 7, even more

preferably in a buffer of pH 7 comprising the amino acids such as L-N-methylvaline, L-leucine, L-N-methylthreonine, L-proline, L-N-methylleucine and glycine to be incorporated into (methyl)griselimycin or L-leucine for production of L-trans-4-methylproline. Preferably, the proteins of SEQ ID Nos: 2 to 27 and the variant proteins maintain their activity in the used buffer and preferably protein and substrate precipitation is prohibited. Optionally, the samples may be incubated at a temperature suitable for the protein activity, preferably at temperatures between 0 and 50°C, more preferably at temperatures between 10 and 40°C, even more preferably at temperatures between 20 and 40°C, most preferably at 37°C. Alternatively, the variant nucleic acid is expressed, e.g., by a cell-free expression system, and the resulting protein is then added to a cell harboring the proteins for producing (methyl)griselimycin or L-trans-4-methylproline.

In another embodiment the method for determining a variant of any of the ORFs 1 to 26 which enhances the yield of (methyl)griselimycin or L-trans-4-methylproline or changes the composition of (methyl)griselimycin may be performed in vitro. A procedure performed in vitro (Latin: within the glass) is performed not in a living organism but in a controlled environment, such as in a test tube or Petri dish. In the in vitro assay, the components participating in a reaction are combined in a test tube, the reaction is allowed to proceed and the product is determined. For this to occur, the variant nucleic acids produced as described herein are transcribed and translated into the respective variant proteins, either by a cell or in a cell free translation method, the proteins are purified or cell lysates (crude, fractionated or purified) are used and the proteins are used in the assay together with non-variant proteins as comprised by SEQ ID Nos: 2 to 27. An in vitro assay is useful if the variant proteins are involved in defined steps within the (methyl)griselimycin synthesis pathway, such as for example if the variant protein belongs to the proteins directly performing the synthesis of (methyl)griselimycin such as the NRPS proteins and only NRPS proteins including the variant form participate in the synthesis of (methyl)griselimycin or of L-trans-4-methylproline such as the proteins encoded by the mps genes and only these proteins including the variant form participate in the synthesis of L-trans-4-methylproline. In an in vitro assay, the components participating in the reaction such as amino acids such as L-N-methylvaline, L-leucine, L-N-methylthreonine, L-proline, L-N-methylleucine and glycine and at least one of the proteins of SEQ ID Nos: 2 to 27 including the variant to be tested are combined under defined definitions of time, amounts of components, compositions of reaction buffers, temperature etc. to allow the synthesis of (methyl)griselimycin or a derivative

thereof or L-trans-4-methylproline. The measurements may be carried out in a molecular environment suitable for the activity of the participating proteins. Suitable conditions are as outlined above.

For the in vitro synthesis of L-trans-4-methylproline, three subsequent assays using Mps-1, Mps-2 and Mps-3 are performed. In the Mps-1 assay L-leucine is converted to 5-hydroxyleucine, in the Msp-2 assay 5-hydroxyleucine is converted to γ-methylglutamic acid γ-semialdehyde and in the Msp-4 assay 3-methyl-Δ^1-pyrroline-5-carboxylic acid is converted to 4-methylproline. It is assumed that the cyclisation of γ-methylglutamic acid γ-semialdehyde to 3-methyl-Δ^1-pyrroline-5-carboxylic acid occurs spontaneously. Possibly, Mps-3 participates in this cyclisation reaction. Preferably, for the production of 5-hydroxyleucine the Mps-1 assay is carried out in an aqueous buffer of neutral pH, more preferably in a buffer with a pH around pH 7, even more preferably in a buffer of pH 7 comprising the substrate, leucine or related compounds, and the Mps-1 protein. Preferably, the Mps-1 protein maintains its activity in the used buffer and preferably protein and substrate precipitation is prohibited. Optionally, the samples may be incubated at a temperature suitable for the protein activity, preferably at temperatures between 0 and 50°C, more preferably at temperatures between 10 and 40°C, even more preferably at temperatures between 20 and 40°C, most preferably at 37°C. Typically, the reaction duration is between several minutes and several hours, preferably 30 minutes to two hours and more preferably one hour. Preferably, for the production of γ-methylglutamic acid γ-semialdehyde the Mps-2 assay is carried out in an aqueous buffer of basic pH, more preferably in a buffer with a pH around pH 10, even more preferably in a buffer of pH 10 comprising the substrate, 5-hydroxyleucine, and the Mps-2 protein. Preferably, the Mps-2 protein maintains its activity in the used buffer and preferably protein and substrate precipitation is prohibited. Optionally, the samples may be incubated at a temperature suitable for the protein activity, preferably at temperatures between 0 and 50°C, more preferably at temperatures between 10 and 45°C, even more preferably at temperatures between 30 and 42°C. Typically, the reaction duration is between several minutes and several hours, preferably 30 minutes to four hours and more preferably one to three hours. Preferably, for the production of 4-methylproline the Mps-4 assay is carried out in an aqueous buffer of neutral pH, more preferably in a buffer with a pH of 7 to 8, even more preferably in a buffer of pH 7.5 to 8 comprising the substrate, 3-methyl-Δ^1-pyrroline-5-carboxylic acid or related compounds, and the Mps-4 protein. Preferably, the Mps-4 protein maintains its activity in the used buffer and pref-

erably protein and substrate precipitation is prohibited. Optionally, the samples may be incubated at a temperature suitable for the protein activity, preferably at temperatures between 0 and 50°C, more preferably at temperatures between 10 and 45°C, even more preferably at temperatures between 30 and 42°C and still more preferably at 30°C. Typically, the reaction duration is be-
5   tween several minutes and several hours, preferably 30 minutes to four hours and more preferably one to three hours.

Preferably, for the production of griselimycin the assay is carried out in an aqueous buffer of neutral pH, more preferably in a buffer with a pH of 7 to 8, even more preferably in a buffer of
10   pH 8 comprising the substrates acetyl-CoA, L-valine, (2S, 4R)-4-methylproline (or related compounds), L-leucine, L-proline, L-threonine, L-glycine and the proteins NRPS-1, NRPS-2, NRPS-3 (Non Ribosomal Peptide Synthetase) and MtbH and, as necessary, SAM and ATP. Preferably, the proteins maintain their activity in the used buffer and preferably proteins and substrates precipitation is prohibited. Optionally, the samples may be incubated at a temperature suitable for
15   the protein activity, preferably at temperatures between 0 and 50°C, more preferably at temperatures between 10 and 45°C, even more preferably at temperatures between 20 and 40°C and still more preferably at 30°C. Typically, the reaction duration is between one hour and 100 hours, preferably two to 72 hours. Preferably, for the production of methylgriselimycin the assay is carried out as the assay for griselimycin. However, L-proline is not necessary and therefore, must
20   not be contained in the reaction.

A variety of assays is known in the art by which the skilled person may determine the concentration of a compound and thus measure of whether the concentration of a compound is changed versus a control reaction. Explicitly mentioned are radiometric assays which measure the incor-
25   poration of radioactivity into substances or its release from substances. The radioactive isotopes most frequently used in these assays are ^{14}C, ^{32}P, ^{35}S and ^{125}I. The concentration may also be determined by Western analysis or an ELISA assay using an antibody or antiserum against the tested compound. Chromatographic assays measure product formation by separating the reaction mixture into its components by chromatography. This is usually done by high-performance liq-
30   uid chromatography (HPLC), but one can also use the simpler technique of thin layer chromatography. Although this approach can need a lot of material, its sensitivity can be increased by labeling the substrates/products with a radioactive or fluorescent tag. A preferred method is quanti-

fication is by HPLC/MS. Liquid chromatography-mass spectrometry (LC-MS, or alternatively HPLC-MS) is an analytical chemistry technique that combines the physical separation capabilities of liquid chromatography (or HPLC) with the mass analysis capabilities of mass spectrometry. LC-MS is a powerful technique used for many applications which has very high sensitivity

5   and specificity. The term "mass spectrometry" refers to the use of a ionization source to generate gas phase ions from a sample on a surface and detecting the gas phase ions with a mass spectrometer. The term "laser desorption mass spectrometry" refers to the use of a laser as a ionization source to generate gas phase ions from a sample on a surface and detecting the gas phase ions with a mass spectrometer. A preferred method of mass spectrometry for biomolecules such

10   as acylated acyl acceptor is matrix-assisted laser desorption/ionization mass spectrometry or MALDI. Another preferred method is surface-enhanced laser desorption/ionization mass spectrometry or SELDI. In mass spectrometry the "apparent molecular mass" refers to the molecular mass (in Daltons)-to-charge value, m/z, of the detected ions. In combination with HPLC the analyte is ionized by diverse API techniques such as ESI (elektrospray ionization) or APCI (atmos-

15   pheric pressure chemical ionization). In the analysator, the ions are separated according to their mass-to-charge m/z ratio. Alternatively, the amount of (methyl)griselimycin may be measured by determining its antibiotic activity using assays well-known in the art. In another embodiment the antibiotic activity of a derivative of (methyl)griselimycin may be measured using assays well-known in the art. The skilled person thereby knows of how to differentiate of whether an in-

20   creased antibiotic activity is due to an increased amount of (methyl)griselimycin or of whether a derivative of (methyl)griselimycin with enhanced antibiotic activity is generated, e.g. by determining of whether the generated antibiotic agent is (methyl)griselimycin or not. Such methods are listed above.

25   Antibiotic testing can e.g. be disk diffusion antibiotic sensitivity testing which is a test which uses antibiotic-impregnated wafers. A known quantity of bacteria is grown on agar plates in the presence of thin wafers containing the antibiotics. An area of clearing surrounds the wafer where bacteria are not capable of growing (called a zone of inhibition). This along with the rate of antibiotic diffusion is used to estimate the effectiveness of a particular antibiotic. In general, larger

30   zones correlate with smaller minimum inhibitory concentration (MIC) of the antibiotic. This information can be used to evaluate the effectiveness of a derivative of (methyl)griselimycin versus (methyl)griselimycin. MIC is the lowest concentration of an antimicrobial that will inhibit

36

the visible growth of a microorganism after overnight incubation. Minimum inhibitory concentrations are important in diagnostic laboratories to confirm resistance of microorganisms to an antimicrobial agent and also to monitor the activity of new antimicrobial agents. MIC is generally regarded as the most basic laboratory measurement of the activity of an antimicrobial agent against an organism. MICs can be determined by agar or broth dilution methods usually following the guidelines of a reference body such as the Clinical and Laboratory Standards Institute (CLSI) or the British Society for Antimicrobial Chemotherapy (BSAC). There are several commercial methods available, including the well established Etest strips. The Etest system comprises a predefined and continuous concentration gradient of different antimicrobial agents, which when applied to inoculated agar plates and incubated, create ellipses of microbial inhibition. The MIC is determined where the ellipse of inhibition intersects the strip, and is easily read off the MIC reading scale on the strip.

Controls are a part of the test methods, since they can eliminate or minimize unintended influences (such as background signals). Controlled experiments are used to investigate the effect of a variable on a particular system. In a controlled experiment one set of samples has been (or is believed to be) modified and the other set of samples is either expected to show no change (negative control) or expected to show a definite change (positive control). The control can be determined in one test run together with the test substance. It may be determined before or after determining the effect of the test compound or it may be a known value. A possible control experiment may be an experiment in which the same conditions are used for the production of (methyl)griselimycin, a derivative thereof or L-trans-4-methylproline as in the test experiment, however, the variant nucleic acid is replaced by the nucleic acid which was the starting nucleic acid for producing the variant, e.g. the control nucleic acid is the wildtype nucleic acid. Examples are ORFs 1 to 26 being in an unmodified state in the control experiment, however, being modified in the test experiment.

The nucleic acid sequence of the variant may be determined by sequencing methods as known in the art whereby any mutation in any of ORFs 1 to 26 may be responsible for a modulation of the resulting protein resulting in an increased production of (methyl)griselimycin or L-trans-4-methylproline or in the production of a derivative of (methyl)griselimycin.

37

In another aspect of the invention, a method of producing griselimycin and/or methylgriselimycin is provided comprising providing at least one protein or a functionally active variant or fragment thereof as defined herein or a variant protein obtained by a method for identifying variants and incubating the at least one protein or variant or fragment thereof with L-N-methylvaline, L-

5   leucine, L-N-methylthreonine, L-proline, L-N-methylleucine, glycine and possibly L-methylproline to produce griselimycin and/or methylgriselimycin. Preferably the at least one protein comprises at least NRPS-1, NRPS-2 and NRPS-3. In a further aspect, the present invention provides a method of producing L-trans-4-methylproline comprising providing at least one protein or a functionally active variant or fragment thereof as defined herein or a variant protein

10   obtained by a method for identifying variants and incubating the at least one protein or variant or fragment thereof with L-leucine, 5-hydroxyleucine, γ-methylglutamic acid γ-semialdehyde or 3-methyl-Δ^1-pyrroline-5-carboxylic acid, preferably L-leucine, to produce L-trans-4-methylproline. Preferably, the at least one protein comprises at least Mps-1, Mps-2 and Mps-4. With respect to the terms "providing" and "protein or a functionally active variant or fragment

15   thereof" it is referred to the definitions provided in the context of nucleic acids and proteins of the invention. With respect to the term "obtainable by the method" it is referred to the definitions provided in the context of determining a variant of SEQ ID Nos: 2 to 27. The term "incubating" means maintaining the protein and other components at particular conditions in order to allow production of griselimycin and/or methylgriselimycin or L-trans-4-methylproline. With respect

20   to the particular conditions reference is made to the definitions in the context of determining a variant of SEQ ID Nos: 2 to 27.

Thus, the present invention refers in one embodiment to an isolated nucleic acid sequence comprising at least one nucleic acid selected from the group consisting of:

25   (a)   a nucleic acid comprising at least one of Open Reading Frames (ORFs) 1 to 26 of SEQ ID NO: 1 that encodes proteins of SEQ ID NOs: 2 to 27 respectively, or a variant or fragment thereof whereby the variant or fragment encodes a functionally active variant or fragment of a protein of SEQ ID NOs: 2 to 27;

(b)   a nucleic acid encoding at least one of the proteins of SEQ ID NOs: 2 to 27, or a func-

30   tionally active variant or fragment thereof;

38

(c)   a nucleic acid encoding a protein that is at least 70%, 80%, 90%, 95% or 97% identical in amino acid sequence to a protein or fragment thereof encoded by the nucleic acid of step (a) or (b);

(d)   a nucleic acid that hybridizes under stringent conditions with a nucleic acid of any of steps (a) to (c);

(e)   a nucleic acid that hybridizes under stringent conditions with a nucleic acid of any of steps (a) to (d) and consists of 10 to 50 nucleotides in length; and

(f)   a nucleic acid that is complementary to a nucleic acid of any of steps (a) to (f).

In a further embodiment the nucleic acid mentioned before comprises at least two ORFs encoding any of the proteins of SEQ ID NOs: 2 to 27, or a functionally active variant or fragment thereof as referred to in any of steps (a) to (d).

In yet another embodiment said nucleic acid comprises at least one of ORFs 18, 19, 20 and 21 that encodes proteins of SEQ ID NO: 19, 20, 21 and 22, respectively, or a functionally active variant or fragment thereof in any of steps (a) to (d).

In yet another embodiment said nucleic acid comprises at least one of ORFs 8, 15 and 16 that encodes proteins of SEQ ID NO: 9, 16 and 17, respectively, or a functionally active variant or fragment thereof in any of steps (a) to (d).

In yet another embodiment said nucleic acid comprises the griselimycin biosynthesis gene cluster having the nucleic acid sequence of SEQ ID NO: 1, or having a variant sequence of SEQ ID NO: 1 harboring a variant or fragment of at least one of ORFs 1 to 26 whereby the variant or fragment encodes a functionally active variant or fragment of a protein of SEQ ID NOs: 2 to 27.

In another embodiment the present invention encompasses an expression vector comprising any of the before mentioned nucleic acids.

39

In another embodiment the present invention encompasses a host cell transformed with the before mentioned expression vector.

5   In yet another embodiment the present invention encompasses a protein comprising at least one amino acid sequence selected from the group consisting of SEQ ID NOs: 2 to 27, or a functionally active variant or fragment thereof.

In yet another embodiment the present invention encompasses an antibody specifically directed
10   against at least one of the before mentioned proteins or a functionally active variant or fragment thereof.

This application also encompasses an invention regarding to a method for producing griselimycin and/or methylgriselimycin comprising producing at least one of the proteins of SEQ ID NOs:
15   9, 16 and 17.

In one embodiment the invention encompasses a method for producing L-trans-4-methylproline comprising producing at least one of the proteins of SEQ ID NOs: 19, 20, 21 and 22.

20   In another embodiment the invention encompasses a method for producing at least one of above mentioned proteins or a functionally active variant or fragment thereof comprising:

a)   expressing any of above mentioned nucleic acids, and

b)   optionally, isolating the above mentioned protein or a functionally active variant or fragment thereof.

25

In yet another embodiment the invention encompasses a method for determining a variant of one or more of ORFs 1 to 26 of SEQ ID NO: 1 that encodes the proteins of SEQ ID NOs: 2 to 27

respectively that modulates the production of griselimycin and/or methylgriselimycin comprising the steps:

a)   producing a variant of any one or more of ORFs 1 to 26,

b)   expressing the variant into a protein,

c)   producing griselimycin and/or methylgriselimycin or of a derivative thereof using the protein of step b), and

d1)   determining the amount of griselimycin and/or methylgriselimycin produced,

wherein an increase of the amount of griselimycin and/or methylgriselimycin produced as compared to the amount produced if the protein of step b) is not present, but the corresponding invariant form of the protein is present is indicative that the variant is capable of increasing the production of griselimycin and/or methylgriselimycin, or

d2)   determining the antibiotic activity of the derivative of griselimycin and/or methylgriselimycin,

wherein an increase in the antibiotic activity in step d2) compared to the antibiotic activity if the protein of step b) is not present, but the corresponding invariant form of the protein is present is indicative that the variant is capable of producing a derivative of griselimycin and/or methylgriselimycin with increased antibiotic activity,

and wherein the derivative differs in amino acid composition from griselimycin and/or methylgriselimycin and exhibits an enhanced antibiotic activity over griselimycin and/or methylgriselimycin.

In another embodiment the invention encompasses a method for determining a variant of one or more of ORFs 1 to 26 of SEQ ID NO: 1 that encodes the proteins of SEQ ID NOs: 2 to 27 respectively that enhances production of L-trans-4-methylproline comprising the steps of:

a)   producing a variant of any one or more of ORFs 1 to 26,

b)   expressing the variant into a protein,

c)   producing L-trans-4-methylproline using the protein of step b), and

d)   determining the amount of L-trans-4-methylproline produced,

wherein an increase in the amount of L-trans-4-methylproline produced in step c) compared to the amount produced if the protein of step b) is not present, but the corresponding invariant form of the protein is present is indicative that the variant is capable of increasing the production of L-trans-4-methylproline.

5

In another embodiment the invention encompasses a method for producing griselimycin and/or methylgriselimycin comprising the steps of:

a)   providing at least one protein of claim 8 or a functionally active variant or fragment thereof, and

10   b)   incubating the at least one protein of step a) with L-N-methylvaline, L-leucine, L-N-methylthreonine, L-proline, L-N-methylleucine and glycine to produce griselimycin and/or methylgriselimycin.

In another embodiment the invention encompasses a method for producing L-trans-4-

15   methylproline comprising the steps of:

a)   providing at least one protein of claim 8 or a functionally active variant or fragment thereof, and

b)   incubating the at least one protein of step a) with L-leucine, 5-hydroxyleucine, γ-methylglutamic acid γ-semialdehyde and/or 3-methyl-Δ^1-pyrroline-5-carboxylic acid to pro-

20   duce L-trans-4-methylproline.

In yet another embodiment the invention encompasses the *Streptomyces* strain deposited with the Deutsche Sammlung von Mikroorganismen und Zellkulturen Gmbh (DSMZ) under deposit number DSM 22643.

25

Brief Description of the Figures

Fig. 1 shows the structure of griselimycin (A) and methylgriselimycin (B).

42

Fig. 2 shows the map and gene arrangement of the griselimycin biosynthesis gene cluster of *Streptomyces DSM 22643* as comprised by SEQ ID NO: 1. The putative ORFs are indicated by black block arrows designated by the respective gene names. Transcription directions and rela-

5   tive sizes are indicated.

Fig. 3 shows a model of the griselimycin biosynthesis. The putative NRPS proteins, NRPS-1, NRPS-2, NRPS-3, are indicated by white blocks. The modules designated M1 to M10 are indi-cated below each block. Each module is dissected in domains designated C for condensation

10   domain, A for adenylation domain, MT for methyltransferase domain, T for thiolation domain, E for epimerisation domain and TE for thioesterase domain, whereby the number behind each do-main designation indicates the number of the module to which this domain is assigned. The ar-rangement A'-MT-A'' indicates that the MT domains are integrated into the A domains. Below each module designation stands the amino acid which is introduced by the respective module

15   into griselimycin. Nme-Val represents L-N-methylvaline, Me-Pro represents L-trans-4-methylproline, Nme-Thr represents L-N-methylthreonine, Leu represents L-leucine, Pro repre-sents L-proline, Nme-D-Leu represents D-N-methylleucine and Gly represents glycine. Below the modules, the biosynthetic role of each module is presented by the presentation of the struc-ture of the respective amino acids and their incorporation into griselimycin. By each module, one

20   amino acid is condensed to the foregoing one resulting in an elongation of the growing peptide chain. Product release involves the TE10 domain with the subsequent cyclisation of the peptide to result in the macrocyclic product. The N-terminal C domain of module 1 is assumed to cata-lyze the acetylation of valine.

25   Fig. 4 shows the synthesis of 4-methylproline within the nostopeptolide biosynthesis gene cluster (Hoffmann et al., 2003; upper pathway) and the griselimycin biosynthesis gene cluster (lower pathway). In both biosynthesis pathways L-leucine is converted by the enzymes Orf1 (nostopep-tolide) or Mps-1 ((methyl)griselimycin) to 5-hydroxyleucine, which is converted by NosE (nost-opeptolide)) or Mps-2 ((methyl)griselimycin) to γ-methylglutamic acid γ-semialdehyde which

30   spontaneously forms 3-methyl-Δ^1-pyrroline-5-carboxylic acid. This is converted to 4-methylproline by NosF (nostopeptolide) or Mps-4 ((methyl)griselimycin). The products of each step of the 4-methylproline synthesis of the (nostopeptolide) or ((methyl)griselimycin) pathways

differ by their configuration with respect to the methyl group at position 4 of 5-hydroxyleucin, γ-methylglutamic acid γ-semialdehyde and 4-methylproline and at position 5 of 3-methyl-Δ^1-pyrroline-5-carboxylic acid, in that in the nostopeptolide biosynthesis the methyl group is in the S configuration, whereas in the ((methyl)griselimycin) biosynthesis the methyl group is in the R

5   configuration. Possibly, Mps-3 plays a role in the conversion of γ-methylglutamic acid γ-semialdehyde to 3-methyl-Δ^1-pyrroline-5-carboxylic acid

Fig. 5 shows the results of an HPLC-MS analysis of the Mps-1 or Leucine-hydroxylase assay. The diagram shows that in the enzyme assay performed with inactivated Mps-1 (chromatograms

10   A) and B)) hydroxyleucine is not produced whereas in the assay performed with active Mps-1 (chromatograms C) and D)) hydroxyleucine is produced. A) and C): Extracted ion chromatograms for m/z= 273.1 ([M+Na]$^+$ of PTC-leucine). B) and D): Extracted ion chromatograms for m/z= 289.1 ([M+Na]$^+$ of PTC-hydroxyleucine).

15   Fig. 6 shows results obtained with mutant nrps-1 and nrps-3 genes and a mutant mps operon. The nrps and mps genes are indicated by block arrows. The approximate location of the insertional mutations is indicated by the circle-cross symbol. Below the presentation of the nrps and mps genes are the relevant UHPLC chromatograms (UV 210 nm) of the supernatant from cultures of the wildtype, the nrps-1 mutant, the nrps-3 mutant and the mps-mutant. In the wildtype the peak

20   at 5.29 min shows the presence of griselimycin. In contrast no griselimycin can be detected in the culture supernatants of the three mutants.

Fig. 7 shows complementation of the mps-mutant obtained in Example 5.

A. Complementation of the mps mutant by feeding 20 µg/ml (2S, 4R)-4-methylproline once at

25   day 0 to the production medium. At days 1 to 4 the (methyl)griselimycin production was analysed (grey bar: griselimycin production, black bar: methylgriselimycin production). Only in the complement medium the mps-mutant is able to produce (methyl)griselimycin. The experiment that is exhibiting an increase of methylgriselimycin versus griselimycin when methyl-prolin was feeded is mentioned in example 9 and on page 15.

30   B. Schematical drawing of the relevant aspects for genetic complementation of the mps-mutant: The mps-operon (mps-1, mps-2, mps-3 and mps-4 genes) is under expression control of the 5` inserted constitutive ermE-promotor.

C. Genetic complementation of the mps-mutant. After integration of the P(ermE)-mps-operon into the attB site of the mps mutant, the capability of griselimycin production is restored (grey bar: griselimycin production, black bar: methylgriselimycin production).

Fig. 8 shows an HPLC-MS analysis of derivatized extracts from an *mps* expression culture (*Streptomyces coelicolor*/pUWL201-mps; Fig. 8A) and the negative control (*Streptomyces coelicolor* WT; Fig. 8B). The PITC derivatives of the expected product (*2S,4R*)-4-methylproline (4th chromatogram of Fig. 8A and B) as well as PITC derivatives of two biosynthetic intermediates 5-hydroxyleucine (2nd chromatogram of Fig. 8A and B) and γ-methylglutamic acid γ-semialdehyde (3rd chromatogram of Fig. 8A and B) could be detected in the extract of the *mps* expression culture, but not in *Streptomyces coelicolor* WT. The 1st chromatogram of Fig. 8A and B corresponds to leucine

Fig. 9 shows a phylogenetic tree indicating the relationship between NosA-A(mPro), NcpB-A(mPro) and NosD-A(pro) and NRPS-1-A2, NRPS-1-A5 and NRPS-2-A8. The scale bar represents 0.09 substitutions per amino acid site.

Fig. 10 shows an analysis of substrate specificity of the A domains of module 2 (A2; upper panel), module 5 (A5; middle panel) and module 8 (A8; lower panel) by in vitro ATP-Ppi-exchange experiments. For A5 and A8 experiments were performed in triplicate, for A2 in duplicate. Highest activities were scaled up to 100%. The following substrates were tested: (2S,4R)-4-fluoroproline (2S,4R Fpro), (2S,4S)-4-methylproline (2S,4S MePro), (2S,4R)-4-methylproline (2S,4R MePro), L-proline (L-Pro), L-pipecolic acid (L-Pip), L-glycine (L-Gly), 4-methylenproline (Methylen Pro), 4,4-Dimethylproline (Di Me Pro) and 4,4-difluoroproline (Di F Pro). Relative activities are indicated on the Y axis. Below each substrate activities as compared to highest activities which are given as 100% are indicated in percent. In the control experiment, no amino acid was present.

Fig. 11: Binding of full length regulators Orf-1 and Orf-12 to DNA fragments Pnrps1, Pnrps2 and Pmxan_ctrl. Scanned pictures of the EMSA assays are shown. Pnrps1 and Pnrps2: EMSA assay using DNA fragments comprising promoter regions of genes nrps1 or nrps2. Pmxan_ctrl: EMSA assay using Hex labeled DNA fragment comprising the dual promoter system in an inter-

genic region of *Myxococcus xanthus* DK1622 flanked by mxan_3950 and mxan_3951. Orf-1-His and Orf-12-His: C-terminal 6xHis tagged Orf-1 and Orf-12. +: With regulator. -: without regulator.

5   Fig. 12 shows the predicted domain organization of regulators Orf-1 and Orf-12. Domains predicted by the SMART tool at http://smart.embl-heidelberg.de/ are shown. HTH: helix-turn-helix motif, aa: amino acids.

Fig. 13 shows the binding of shortened versions of regulators Orf-1 and Orf-12 to DNA frag-
10   ments Pnrps1, Pnrps2 and Pmxan_ctrl. Scanned pictures of the EMSA assays are shown. In the assays derivatives of Orf-1 and Orf-12 were used only comprising the predicted HTH domains. Pnrps1 & Pnrps2: EMSA assay using DNA fragments comprising promoter regions of genes nrps1 or nrps2. Pmxan_ctrl: EMSA assay using Hex labeled DNA fragment comprising the dual promoter system in an intergenic region of *Myxococcus xanthus* DK1622 flanked by mxan_3950
15   and mxan_3951. Orf-1-HTH1, Orf-1-HTH2, Orf1-HTH12 and Orf-12-HTH: C-terminal 6xHis tagged Orf-1 and Orf-12 only comprising HTH domains. -: without regulator. Numbers depict the molar excess of regulator protein compared to the quantity of DNA fragment.

20   Description of the Examples

Example 1: Identification of the griselimycin biosynthesis locus in *Streptomyces DSM 22643*

Streptomyces DSM 22643 naturally produces the antibiotics griselimycin and methylgriselimy-
cin. However, the genetic locus involved in the biosynthesis of these antibiotic substances was
25   not previously identified. For the identification of the griselimycin locus, *Streptomyces DSM 22643* was cultured according to standard microbiological techniques. DNA was extracted according to standard procedures and the total sequence was determined according to standard procedures. Analysis of the genes comprised by the DNA of *Streptomyces DSM 22643* and proteins encoded by the genes was performed in silico. Table 1 lists the genes identified by ORF numbers
30   which were determined by homology searches. Specific gene designations were assigned to the ORF numbers. The location of the genes and their arrangement within the griselimycin biosynthesis gene cluster is schematically indicated in Fig. 2.

Table 1

ORF number	gene designation	start nt	stop nt	gene length (nt)	strandedness	protein length (aa)	protein designation	SEQ ID NO: protein
1	orf-1	837	2075	1269	positive	412	Orf-1	2
2	orf-2	2185	2625	441	positive	146	Orf-2	3
3	orf-3	2625	2819	195	positive	64	Orf-3	4
4	orf-4	3033	4037	1005	positive	334	Orf-4	5
5	orf-5	4110	5630	1521	positive	506	Orf-5	6
6	orf-6	5627	7066	1440	positive	479	Orf-6	7
7	orf-7	7240	8133	894	positive	297	Orf-7	8
8	nrps-1	31885	8501	26385	negative	7794	NRPS-1	9
9	int-1	33102	32770	333	negative	110	Int-1	10
10	tnp-1	33187	33726	540	positive	179	Tnp-1	11
11	int-2	34127	33567	561	negative	186	Int-2	12
12	tnp-2	36300	34648	1653	negative	550	Tnp-2	13
13	tnp-3	36172	37314	1143	positive	380	Tnp-3	14
14	dna	39032	37746	1287	negative	428	Dna	15
15	nrps-2	39643	52149	12507	positive	4168	NRPS-2	16
16	nrps-3	52146	56120	3975	positive	1324	NRPS-3	17
17	mbtH	56117	56335	219	positive	72	MbtH	18
18	mps-1	56434	57246	813	positive	270	Mps-1	19

47

ORF number	gene de-signation	start nt	stop nt	gene length (nt)	strandedness	protein length (aa)	protein designation	SEQ ID NO: protein
19	mps-2	57243	58418	1176	positive	391	Mps-2	20
20	mps-3	58445	59245	801	positive	266	Mps-3	21
21	mps-4	59242	60141	900	positive	299	Mps-4	26
26	orf-8	61680	61144	537	negative	178	Orf-8	26
26	orf-9	62170	62820	651	positive	216	Orf-9	24
24	orf-10	62824	63939	1116	positive	371	Orf-10	25
25	orf-11	63969	65165	1197	positive	398	Orf-11	26
26	orf-12	66108	65206	903	negative	300	Orf-12	27

Example 2: Genes and proteins involved in the biosynthesis of griselimyin and methylgriselimy-cin

By in silico analysis it was determined that the griselimycin biosynthesis gene cluster comprises 26 genes encoding 26 proteins. The biological function of the 26 (methyl)griselimycin biosynthetic proteins was assessed by computer comparison of each identified protein with proteins found at the NCBI. Table 2 identifies the proteins as comprised by the griselimycin biosynthesis gene cluster and the corresponding GenBank proteins found by the homology searches.

Table 2

ORF number	best blast hit					putative function
	Best blast hit	GenBank Accession number	nearest homologue in	e-Value	identity/ similarity (%)	
1	TlmR2	ABL74963.1	Streptoalloteichus hindustanus	2E-55	35/53	Xre family transcriptional regulator
2	Hypothetical protein SSOG_02153	ZP_07294072	Streptomyces hygroscopicus ATCC 53653	5E-11	41/47	Hypothetical protein
3	Conserved hypothetical protein	ZP_06583961	Streptomyces roseosporus NRRL 15998	3E-15	61/86	Conserved hypothetical protein
4	Conserved hypothetical protein	CBG69271.1	Streptomyces scabiei 87.26	1,00E-140	81/87	Conserved hypothetical protein
5	Putative transporter	CBG69272.1	Streptomyces scabiei 87.26	0.0	91/94	Transporter
6	Putative glutamine synthetase	CBG69273.1	Streptomyces scabiei 87.26	0.0	81/87	Glutamine synthetase
7	Hypothetical protein	CBG69274.1	Streptomyces scabiei 87.26	8E-120	73 81	Hypothetical protein
8	Amino acid adenylation domain-containing protein	YP_001539321.1	Salinispora arenicola CNS-205	0.0	51/63	NRPS
9	Integrase catalytic region	YP_004330163	Pseudonocardia dioxanivorans CB1190	1E-43	75/92	Integrase catalytic region
10	Putative transposase for insertion sequence element	YP_0015096 42	Frankia sp. EAN1pec	3E-15	76/78	Transposase for insertion sequence element
11	Integrase catalytic region	YP_004330974	Pseudonocardia dioxanivorans CB1190	4E-33	67/81	Integrase catalytic region
12	Rhodopirellula transposase family protein	ZP_07610914	Streptomyces violaceusniger Tu 4113	0.0	74/80	Transposase
13	Putative IS4 family Transposase	ZP_06418432.1	Frankia sp. EUN1f	9E-60 E-60	43/55	Transposase

50

ORF number	best blast hit					putative function
	Best blast hit	GenBank Accession number	nearest homologue in	e-Value	identity/ similarity (%)	
14	DNA polymerase III, beta subunit	ZP_06824689	Streptomyces sp. SPB74	4E-110	53/70	DNA polymerase III, beta subunit
15	putative pristinamycin I peptide synthase 3 and 4	CBH31051.1	Streptomyces pristinaespiralis	0.0	51/62	NRPS
16	Putative NRPS	ZP_04685020.1	Streptomyces ghanaensis ATCC 14672	0.0	45/61	NRPS
17	MbtH domain protein	ZP_06476953	Frankia symbiont of Datisca glomerata	9E-26	72/81	MbtH domain protein
18	Phytanoyl-CoA dioxygenase	YP_001539320.1	Salinispora arenicola CNS-205	8E-91	60/76	Leucine hydroxylase
19	Alcohol dehydrogenase	YP_707358.1	Rhodococcus jostii RHA1	6E-102	51/67	Alcohol dehydrogenase
20	Alpha/beta hydrolase fold protein	ZP_04702845.1	Streptomyces albus J1074	3E-77	56/71	Alpha/beta hydrolase
21	LmbY	ABX00626.1	Streptomyces lincolnensis	5E-131	76/88	Oxidoreductase
26	Hypothetical protein SclaA2_13611	ZP_08216838	Streptomyces clavuligerus ATCC 27064	1E-42	65/78	PNPOx-like superfamily protein
26	Hypothetical protein SAC-TEDRAFT_3210	ZP_06272665.1	Streptomyces sp. ACTE	3E-38	64/74	Hypothetical protein
24	Hypothetical protein SAC-TEDRAFT_3209	ZP_06272664.1	Streptomyces sp. ACTE	2E-112	67/78	Hypothetical protein
25	Amidohydrolase	ZP_06272663.1	Streptomyces sp. ACTE	1E-148	70/80	Amidohydrolase
26	Hypothetical protein SACTEDRAFT_3207	ZP_06272662.1	Streptomyces sp. ACTE	2E-64	52/69	Xre family transcriptional regulator

- 51 -

Between the individual modules and domains, there are linker regions, namely inter-module linker and inter-domain linker. At the N-terminus and C-terminus of the proteins, there are COM or docking domains which are necessary for interaction of the individual NRPS proteins. Com-munication-mediating (COM) domains facilitate the selective interaction between nonribosomal

5   peptide synthetases by short communication-mediating domains (Hahn M. and Stachelhaus T.; 2004).

Based on homology searches the genes designated nrps-1, nrps-2 and nrps-3 encode subunits of the nonribosomal protein synthetase (NRPS) synthesizing griselimycin and methylgriselimycin (see Fig. 3). The nrps-1 gene has a length of about 26.4 kb and encodes 6 modules that are nec-

10   essary for the incorporation of (1) L-N-methylvaline (M1), (2) L-trans-4-methylproline (M2), (3) L-N-methylthreonine (M3), (4) L-leucine (M4), (5) L-trans-4-methylproline (M5) and (6) L-leucine (M6), respectively, into (methyl)griselimycin. Module 1 or M1 and module 3 or M3 each encode a C-domain, an A-domain which is interrupted by a MT-domain and a T-domain, whereas module 2 or M2, module 4 or M4, module 5 or M5 and module 6 or M6 each encode a

15   C-domain, an A-domain and a T-domain. Furthermore it contains the condensation domain of module 7 or M7. The condensation domain of M1, C1, is responsible for N-acetylation. The nrps-2 gene has a length of about 12.5 kb and encodes part of module 7 and modules 8 and 9 that are necessary for the incorporation of (7) L-N-methylvaline (M7), (8) L-proline (griselimycin) or L-trans-4-methylproline (methylgriselimycin) (M8) and (9) D-N-methylleucine (M9). The part

20   of module 7 comprises an A-domain which is interrupted by a MT-domain and a T-domain, module 8 or M8 comprises a C-domain, an A-domain and a T-domain and module 9 or M9 com-prises a C-domain, an A-domain which is interrupted by a MT-domain, a T-domain and an E-domain. The presence of a D-amino acid at position 9 of (methyl)griselimycin coincides with the presence of an E-domain in module 9 which catalyses the insertion of D-N-methylleucine. The

25   nrps-3 gene has a length of about 4 kb and codes for module 10 which incorporates glycine into the molecule and manages the cyclisation of the linear peptide chain between glycine and L-N-methylthreonine via an ester bond. Module 10 comprises a C-domain, an A-domain, a T-domain and a TE-domain.

30   Homology searches allowed the assignment of a biosynthesis role to the repeating units, the modules, of the NRPS proteins. Fig. 3 shows the assignment of modules 1 to 10 to the NRPS proteins, the dissection into the domains and the biosynthesis activity by indicating the amino

acids and their incorporation into the growing griselimycin chain. Module 10 releases the chain from the NRPS-3 protein, followed by chain cyclisation. In Table 3, the approximate boundaries of the modules and domains on the protein and DNA level with respect to their location within the individual NRPS proteins are indicated. Thereby, C represents a condensation domain, A represents an adenylation domain, MT represents a methyltransferase domain, and T represents a thiolation domain. The MT domains are integrated into the A domains. This is indicated by the arrangement "A'-MT-A''".

Table 3

Protein	Modules/ Domains	Location within the protein sequence (aa)	Location within the gene cluster sequence (SEQ ID NO: 1) (nt)
NRPS-1 (SEQ ID NO: 9)	Module 1	9-1493 aa	31861-27407 nt
	C1	9-336 aa	31861-30878 nt
	A1'	481-939 aa	30445-29069 nt
	MT	940-1365 aa	29068-27791 nt
	A1''	1366-1406 aa	27790-27668 nt
	T1	1429-1493 aa	27601-27407 nt
	Module 2	1517-2569 aa	27337-24179 nt
	C2	1517-1858 aa	27337-26312 nt
	A2	2003-2482 aa	25879-24440 nt
	T2	2505-2569 aa	24373-24179 nt
	Module 3	2591-4102 aa	24115-19580 nt
	C3	2591-2932 aa	24115-26090 nt
	A3'	3077-3536 aa	26657-21278 nt
	MT3	3537-3974 aa	21277-19964 nt
	A3''	3975-4015 aa	19963-19841 nt
	T3	4038-4102 aa	19774-19580 nt
	Module 4	4124-5170 aa	19516-16376 nt
	C4	4124-4465 aa	19516-18491 nt
	A4	4612-5083 aa	18052-16637 nt
	T4	5106-5170 aa	16570-16376 nt
	Module 5	5193-6245 aa	16309-13151 nt
	C5	5193-5534 aa	16309-15284 nt
	A5	5679-6158 aa	14851-13412 nt
	T5	6181-6245 aa	13345-13151 nt
	Module 6	6268-7312 aa	13084-9950 nt
	C6	6268-6609 aa	13084-12059 nt
	A6	6754-7265 aa	11626-10211 nt
	T6	7248-7312 aa	10144-9950 nt
	Module 7'	7336-7677 aa	9880-8855 nt
	C7	7336-7677 aa	9880-8855 nt
NRPS-2 (SEQ ID NO: 16)	Module 7''	42-1066 aa	39766-42840 nt
	A7'	42-500 aa	39766-41142 nt
	MT7	501-938 aa	41143-42456 nt

- 53 -

Protein	Modules/ Domains	Location within the protein sequence (aa)	Location within the gene cluster sequence (SEQ ID NO: 1) (nt)
	A7''	939-979 aa	42457-42579 nt
	T7	1002-1066 aa	42646-42840 nt
	Module 8	1088-2148 aa	42904-46086 nt
	C8	1088-1430 aa	42904-43932 nt
	A8	1575-2061 aa	44365-45825 nt
	T8	2084-2148 aa	45892-46086 nt
	Module 9	2172-4160 aa	46156-52126 nt
	C9	2172-2513 aa	46156-47181 nt
	A9'	2658-3098 aa	47614-48936 nt
	MT9	3099-3536 aa	48937-50250 nt
	A9''	3537-3577 aa	50251-50373 nt
	T9	3600-3663 aa	50440-50631 nt
	E9	3679-4160 aa	50677-52126 nt
NRPS-3 (SEQ ID NO:. 17)	Module 10	13-1312 aa	52182-56081 nt
	C10	13-343 aa	52182-53174 nt
	A10	484-951 aa	53595-54998 nt
	T10	974-1038 aa	55065-55259 nt
	TE10	1060-1312 aa	55326-56081 nt

nt: nucleotide, aa: amino acid

Based on homology searches and biochemical analysis the genes designated mps-1, mps-2- and mps-4 encode proteins necessary for the biosynthesis of L-trans-4-methylproline. L-trans-4-methylproline is contained at positions 2 and 5 of griselimycin and at positions 2, 5 and 8 of me-thylgriselimycin. The mps-1 gene encodes a protein with highest homology to 4-proline hy-droxylases (41% identity/57% similarity) and phytanoyl-CoA dioxygenase but by biochemical analysis it could be shown that the mps-1 gene codes for a leucine hydroxylase which hydroxy-lates L-leucine to (2S, 4R)-5-hydroxyleucine. The gene mps-2 codes for an alcohol dehydro-genase that has 34% identity/55% similarity to the *nosE* gene of the nostopeptolide gene cluster which is involved in synthesis of L-cis-4-methylproline (Luesch et al. 2003). Mps-2 catalyses the dehydrogenation of (2S,4R)-5-hydroxyleucine to (2S, 4R)-γ-methylglutamic acid γ-semialdehyde. After spontaneous ring formation the protein encoded by mps-4 probably cataly-ses the reduction of (3S, 5R)-3-methyl-Δ^1-pyrroline-5-carboxylic acid to (2S,4R)-4-methylproline. The mps-4 gene shows the highest homology to Sib (55% identity/70% similar-ity) and LmbY, that both code for flavin dependent monoxygenases and catalyse a similar reac-tion (Li et al. 2009; Peschke et al. 1995). The mps-3 gene shows highest homology to α,β-hydrolase with highest homology to the *lipE* gene product (50% identity/68% similarity) in the friulimycin biosynthetic gene cluster (Müller et al. 2007). The encoded protein might act as type

- 54 -

II thioesterase with a function in regeneration of misprimed NRPS. Possibly, the mps-3 gene product is involved in the cyclisation of the pyrroline derivative in the biosynthesis pathway of 4-methylproline. Based on these results, the present inventors succeeded to elucidate the biosynthetic pathway of L-trans-4-methylproline for the biosynthesis of griselimycin or methylgriseli-

5   mycin. Fig. 4 shows a comparison of the biosynthesis pathways of 4-methylproline including the proteins involved therein within the nostopeptolide biosynthesis pathway of the cyanobacterium Nostoc sp. GSV264 (Hoffmann et al. 2003) and within the (methyl)griselimycin biosynthesis pathway as identified by the present invention. Based on sequence analysis the genes designated mps-1 to mps-4 are probably transcribed as an operon.

10   Furthermore, based on homology searches, the following genes were identified. There are two genes orf-1 and orf-12 with homology to proteins of the Xre-family transcriptional regulators in the gene cluster. The encoded proteins play a role in regulation of (methyl)griselimycin biosynthesis. It could be shown that inactivation of both genes results in a significant loss of production of (methyl)griselimycin, which proves their function as direct or indirect regulators. Orf-5 en-

15   codes a protein with homology to a putative transport molecule and might be involved in export of (methyl)griselimycin. Orf-6 encodes a protein with the putative function of a glutamine synthetase. Int-1 and int-2 encode proteins with homology to integrase catalytic regions, so that it is assumed that these proteins have integrase functionality. Tnp-1, tnp-2 and tnp-3 between the genes nrps-1 and nrps-3 encode proteins with homology to putative transposase proteins and are

20   assumed to have transposase activity. A further open reading frame has the putative function as a subunit ß of the DNA polymerase III. This orf was designated *dna*. A small 219 bp ORF directly downstream of nrps-3 encodes a MbtH like protein and was therefore designated mbtH. The MbtH protein is needed for correct function of the NRPS (Barry & Challis, 2009). Orf-8 has homology to the hypothetical protein Scla2_13611 and is assumed to belong to the PNPOx-like

25   superfamily. Orf-11 encodes a putative protein of the amidohydrolase superfamily. The proposed genes orf-2, orf-3, orf-4, orf-7, orf-9 and orf-10 all encode different conserved hypothetical proteins.

Example 3: In vitro-synthesis of L-trans-4-methylprolin

30   In vitro-synthesis of L-trans-4-methylproline can be performed by coupling three subsequent assays, namely the Mps-1 assay, the Mps-2 assay and the Mps-3 assay. Cyclisation of γ-methylglutamic acid γ-semialdehyde to 3-methyl-Δ^1-pyrroline-5-carboxylic occurs spontane-

ously or may involve Mps-3. Fig. 4 shows the conversion of L-leucine to (2S,4R)-4-methylproline using Mps-1, Mps-2, possibly Mps-3 and Mps-4.

(a) Leucine hydroxylase assay

Based on in silico analysis it was found that the mps-1 gene encodes a leucine hydroxylase. The following experimental procedure was performed which confirms that the Mps-1 protein encodes a leucine hydroxylase.

The putative leucine hydroxylase encoding gene mps-1 was amplified using the primers Gri3_for (5'- GCCGCCATATGATGCAGCTCACGGCCGAT-3') (SEQ ID NO: 28) and Gri3_rev (5'-GGTCAGGATCCTCATGCCAGCCTCGATTC-3') (SEQ ID NO: 29) from genomic DNA from *Streptomyces DSM 26643* with Phusion polymerase and cloned into the pJET 2.1 vector (Fermentas) by blunt end ligation. After sequence verification the DNA fragment was subcloned into pET28b(+) via the introduced *NdeI/Bam*HI restriction sites. The resulting construct pGris3 was transformed into *E. coli Rosetta* BL21 (DE3) pLysS/RARE for protein expression. Fresh transformants harboring the constructs were grown in LB-medium (1 l batches started with 0.1% inocula from a 10-ml culture grown for 5 h at 37 °C) supplemented with kanamycin (50 µg/ml) and chloramphenicol (34 µg/ ml). All cells were grown at 37°C to an OD600 of approximately 0.8. The cells were then induced with 1 M isopropyl-b-D-thiogalactopyranoside (IPTG) to an end concentration of 0.2 mM and then grown at 16°C overnight. The cells were harvested by centrifugation (6000 rpm, 10 min, 4°C) and resuspended in buffer A (20 mM Tris–HCl, pH 7.8, 200 mM NaCl, and 10% [v/v] glycerol). The cells were then lysed (2 passes at 700 psi, French press, SLM Aminco), and the cell debris was removed by centrifugation (21,000 rpm, 10 min. 4°C). Prepacked HisTrap™ HP columns were used for preparative purification of histidine-tagged recombinant proteins by immobilized metal ion affinity chromatography (IMAC) on the Äkta prime™ plus system (Ge Healthcare). 15 ml protein lysates were filtered through a sterile filter and loaded onto a 1 ml HisTrap column. Purification was performed as recommended in the GE Healthcare manual (HisTrap HP, Instructions 71-5027-68 AF). The desired protein was eluted from the column in a stepwise imidazole gradient with buffer B (20 mM Tris–HCl, pH 7.8, 200 mM NaCl, and 10% [v/v] glycerol and 60, 100, 200, 300 and 500 mM imidazole, 10 ml fractions). Fractions containing the pure target protein, as determined by SDS–polyacrylamide gel electrophoresis (PAGE), were combined and concentrated to approximately 200 µl by using Amicon Ultra PL-10 centricons. Then 800 µl of storage buffer (50 mM Tris–HCl, pH 7.5, 50 mM NaCl, and 10% [v/v] glycerol) was added to the concentrated protein before flash-freezing

in liquid nitrogen and storage at -80°C. Protein concentrations were determined using the Bradford assay (Bio-Rad, Hercules, CA, USA). 1-3 mg/ml purified protein were obtained per liter of culture.

Assays with leucine as substrate were conducted in MOPS (50 mM), pH7.0, and included substrate (1 mM), α-ketoglutarat (1 mM), $FeSO_4$ (25 μM), DTT (0.5 mM), ascorbate (0.1 mM) and Mps1 (1 μM) in a total volume of 200 μl. Reactions were initiated by the addition of Mps1 and incubated at 30°C for 1 hour. Protein was precipitated with cold ethanol and the supernatant was decanted and used for derivatization. The solution was dried by rotary evaporation and resuspended in 200 μl coupling buffer (acetonitrile:pyridine:triethylamine:H_2O 10:5:2:3). 50 μl of PITC solution were added to this solution and after a 5 min reaction at room temperature 50 μl H_2O were added to stop the reaction. The solution was evaporated to dryness by speedvac evaporation. The resulting PTC amino acids were dissolved in 1 ml water:acetonitrile (7:2), centrifuged and the supernatant was dried by speed vac evaporation again. The amino acids were then dissolved in 100 μl water/acetonitrile and used for HPLC/MS analysis. PITC-leucine and PITC-hydroxyleucine were detected in the positive ionization mode (Leucine: $[M+H]^+$ = 272.9 and hydroxyleucine: $[M+H]^+$ = 288.9). It was shown that Msp-1 catalyses the reaction from leucine to 5-hydroxyleucine (Fig. 5).

(b) Mps-2 assay

The conversion of 5-hydroxyleucine to γ-methylglutamic acid γ-semialdehyde can be performed by the Mps-2 or alcohol dehydrogenase assay according to Luesch et al., 2003 who performed the assay with the functionally similar protein NosE. Typical reaction mixtures with 5-hydoxyleucine or related compounds as substrate may contain 100 mM glycine (pH 10), 2 mM *β*-NAD, substrate (0.1 – 10 mM), 1 mM $ZnSO_4$, and approximately 3 μg of Mps-2. Typical reaction volumes may be 500 μl, reactions may be initiated by the addition of Mps-2 and incubated at 30-42°C for 1-3 h.

(c) Mps-4 assay

The conversion of 3-methyl-Δ^1-pyrroline-5-carboxylic acid to 4-methylproline can be performed in the Mps-4 assay according to Luesch et al., 2003 who performed the assay with the functionally similar protein NosF. Typical reaction mixtures may contain 200 mM Tris (pH 8), substrate, 0.2 mM *β*-NADH or ß-NADPH, and approximately 5 μg of Mps-4. Reaction volumes are usually 1 ml, reactions may be initiated by the addition of Mps-4 and incubated at 30-42°C for 1-3 h. An alternative Mps-4 assay according to Becker et al., 1997 comprises 3-methyl-Δ^1-pyrroline-

5-carboxylic acid or related compounds as substrate and may be conducted in 50 mM Tris/HCI, pH 7.5, and include substrate (0.1-10 mM), 0.2 mM FAD, 0.5 mM NADH and an appropriate dilution of Mps-4 (e.g. 1 µg final concentration in a total volume of 100-500 µl. Reactions may be initiated by the addition of Mps-4 and incubated at 30°C for 1-3 hours.

Example 4: In vitro-synthesis of griselimycin and methylgriselimycin

In vitro-synthesis of griselimycin and methylgriselimycin can be performed according to reaction conditions as disclosed in Gaitatzis N. et al., 2001. Incubations of 1 ml each may contain 25 mM Tris HCl (pH 8.0), 50 mM NaCl, 1 mM acetyl-CoA (or 1 mM acetyl-SNAC), 2 mM L-valine, 2 to 3 mM (2S, 4R)-4-methylproline (or related compounds), 3 mM L-leucine, 1 mM L-proline (or no L-proline for the methylgriselimycin synthesis), 1 mM L-threonine, 4 mM S-adenosyl-L-methionine, 1 mM glycine, 10 mM ATP, 0.5-5 mM NRPS-1, NRPS-2, NRPS-3 and MtbH. The reactions may be incubated at 30°C after starting the reaction with ATP for 2-72 hours.

Example 5: Inactivation of the genes nrps-1 and nrps-3 and of the mps operon

Insertional mutants of *Streptomyces DSM 22643* were created according to standard procedures. Fragments of about 3000 bp in length which are homologues to parts of the mps-operon, the nrps-1 and the nrps-3 gene, respectively, were amplified from chromosomal DNA of *DSM 26643* by PCR using the primers listed in Table 4. These fragments were cloned in the *E. coli* cloning vector pBluescript SK+ giving the plasmids pMPS, pNRPS-1, pNRPS-3. Subsequently, an AprR-oriT cassette, created as described in Gust et al., 2003, was amplified with the primers listed in Table 5 and integrated into the plasmids pMPS, pNRPS-1 and pNRPS-3 with the RedET technology (Genebridges) to give conjugational plasmids. The created plasmids were then transferred into *DSM 22643* and conjugants were analysed for correct genetic integration into the mps-operon, the nrps-1 and the nrps-3 gene, respectively, before being tested for (methyl)griselimycin production.

- 58 -

Table 4: Primer for amplification of homologous DNA-fragments

Mutant	Primer	Sequence	SEQ ID NO
Nrps-1	Nrps-1_for	GCTCTAGAGTCATGCAGATCAACTCG	30
	Nrps-1_rev	GCTCTAGATGGTCACGGACCTGTG	31
Nrps-3	Nrps-3_for	GCTCTAGAATGGAGGGCATCGGAAT	32
	Nrps-3_rev	GCTCTAGAAGCTCTTCCGCCAGGAT	33
mps	Mps_for1	CGGGGTACCCATCTGGTCCGTACCAT	34
	Mps_rev1	TCGCGTCGTTTACCTCA	35
	Mps_for2	GCCTTCACCGGAGCTCA	36
	Mps_rev2	GCTCTAGATCCTGAACTCTCGCATA	37

Underlined: sequences for restriction sites

Table 5: Primer for insertion of AprR-oriT-cassette

Mutant	Primer	Sequence	SEQ ID NO
Nrps-1	Nrps-1_for2	TGAGGTGTGTGCTGCGTCTGTTCCGGGGGC GACTGCGGTATTCCGGGGATCCGTCGACC	38
	Nrps-1_rev2	TGACACGTCGTGGCGTGCACCGTCGTCTCC GTCGGCCCGTGTAGGCTGGAGCTGCTTC	39
Nrps-3	Nrps-3_for2	TCCGGATGCGACGGCAGTCGTCTTTGAGA CACGAGCGTATTCCGGGGATCCGTCGACC	40
	Nrps-3_rev2	AGCATCACGTCGAACGATTGTCTGCGGTG TCCATGTCTTGTAGGCTGGAGCTGCTTC	41
mps	Mps_for3	ATGAGCGTAGGTCTCCGTCTGTGGTAACC CCTGGTGTGATTCCGGGGATCCGTCGACC	42
	Mps_rev3	TCAGGTATCACCGTGGGAAGGTGCCGGCT TTCCAGTACTGTAGGCTGGAGCTGCTTC	43

5   Underlined: sequences homologous to theAprR-oriT cassette

To verify griselimycin production in the wild type and to test the insertional mutants for the abil-
ity of griselimycin synthesis, the different strains were cultured in medium NL5645 (composition

(amount in percentage): lactose 2.0, soy meal 2.5, NaCl 0.5, CaCO$_3$ 0.1 and yeast extract 0.5, pH 6.8) in shaking flasks. At different time points the culture supernatant was collected, filtered and mixed 1:1 with ethanol. After precipitation and separation of unsolved substances the super-natant was analysed by UHPLC (Ultra High Performance Liquid Chromatography) for

5   (methyl)griselimycin production.

As can be taken from Fig. 6 the wildtype strain *DSM 22643* produces griselimycin (see "griseli-mycin" marked peak at 5.29 min in the UHPLC chromatogram (UV210 nm)), whereas the nrps-1, the nrps-3 and the mps-mutants of *DSM 22643* do not produce griselimycin (absence of the peark at 5.29 min). This experiment confirms the importance of the NRPS proteins and Mps pro-

10   teins for the synthesis of griselimycin and methylgriselimycin.

Example 6: Complementation of mps-mutants

To demonstrate that the genes of the mps-operon are relevant for synthesis of (2S, 4R)-4-methylproline the mps-mutant was complemented (1) by feeding (2S, 4R)-4-methylproline and

15   (2) by genetic complementation due to the chromosomal integration of an expression construct, where the mps operon is expressed from the constitutive ermE-promoter.

In the feeding experiment 20 µg/ml (2S, 4R)-4-methylproline were fed to cultures of the mps-mutant and (methyl)griselimycin production was analysed in the culture supernatant as described in Example 5. Whereas in the normal culture medium without (2S, 4R)-4-methylproline feeding

20   the mps-mutant does not produce any (methyl)griselimycin), feeding of 20 µg leads to (methyl)griselimycin production (Fig. 7A; also example 9).

For the genetic complementation of the mps-mutant, a construct, P(ermE)-mps, comprising the mps operon with the genes mps-1, mps-2, mps-3 and mps-4 under the control of the constitutive ermE promoter was created in a conjugational vector with an attP attachment site for integrase

25   mediated integration into the chromosomal attB site (see Fig. 7B for the sequence of the ermE-promoter). After integration of this P(ermE)-mps construct into the mps-mutant created in Ex-ample 5 the complemented mps-mutant was tested for griselimycin production. As shown in Fig. 7C griselimycin production could be restored in the complemented mps-mutant.

Both experiments clearly indicate that the mps operon is necessary for the synthesis of (2S, 4R)-

30   4-methylproline, that is then used for methylgriselimycin synthesis.

Example 7: Heterologous production of methylproline by expression of the *mps1-4* gene set in *Escherichia coli* and *Streptomyces coelicolor*

To confirm that the gene set *mps1-4* encodes the enzymatic machinery for 4-methylproline bio-synthesis constructs for the expression in *E. coli* (pET28-mps) and *S. coelicolor* (pUWL201-mps) were generated.

(a) Generation of pUWL201-mps and heterologous expression in *Streptomyces coelicolor*

A ~ 5.8 kb *Eco*RI fragment from Cosmid B:L23 harbouring the 3` end from the griselimycin biosynthetic gene cluster (54283 - 60163 nt from SEQ ID NO:1) was subcloned into the replica-tive *E. coli/Streptomyces* shuttle vector pUWL201 under control of the constitutive *PermE** promotor. The resulting expression construct pUWL201-mps was subsequently transformed into the host strain *Streptomyces coelicolor* A3(2) using the protoplast method (Kieser et al., 2000). The transformants (*S. coelicolor*/pUWL201-mps) were cultivated in parallel to the wildtype strain in 50 ml TSB medium for 4 days at 30°C in baffled shaking flasks (mutant cultures were supplemented with thiostreptone 20 µg/ml final concentration). Cells were harvested by cen-trifugation and extracted with 50 ml methanol (15 min sonication and 30 min stirring). After filtration the extract was evaporated to dryness and derivatized as described below.

(b) Generation of pET28-mps and heterologous expression in *Escherichia coli*

A ~3.7 kb *Xba*I/*Eco*RI fragment from the Cosmid B:L23-CmR was subcloned into the expres-sion vector pET28b (Novagen) linearized with *Nhe*I/*Eco*RI under control of the inducible T7 promotor. The cloned 3.7 kb DNA fragment harbors the genes *mps1-4* (56434 - 60163 nt from SEQ ID NO:1 extended with the sequence CTAGA at the 5' end). The resulting expression con-struct pET28-mps as well as the empty expression vector pET28b were subsequently trans-formed into the host strain *Escherichia coli* BL21 (DE3)/Codonplus/pL1SL2 by electroporation. The transformants (*E. coli* BL21 (DE3)/Codonplus/pL1SL2/pET28-mps or pET28b) were culti-vated in 50 ml LB medium containing 0.2 % glucose supplemented with kanamycin (50 µg/ml final concentration), ampicillin (50 µg/ml final concentration) and chloramphenicol (25 µg/ml final concentration) at 30°C. Cultures were grown until OD_{600} = 0.6, expression was induced with IPTG (0.3 mM final concentration) and incubation was continued for additional 4 hours at 30°C. Cells were harvested by centrifugation and extracted with 50 ml methanol (15 min sonica-

tion and 20 min stirring). After filtration the extract was evaporated to dryness and derivatized as described below.

(c) Derivatization of the extracts and HPLC-MS analysis

The dried cell extracts as well as (2S,4R)-4-methylproline (as reference standard for retention time and fragmentation spectra) were resuspended in 200 µl coupling buffer (acetonitrile:pyridine:triethylamine:H$_2$O). 50 µl of PITC (phenylisothiocyanat) solution were added and after 5 min reaction at room temperature 50 µl H$_2$O were added to stop the reaction. The solution was evaporated to dryness by speedvac evaporation. The resulting PITC amino acids were dissolved in 1 ml water:acetonitrile (7:2), centrifuged and the supernatant was dried by speedvac evaporation. The derivatized extracts and the derivatized (2S,4R)-4-methylproline reference were then dissolved in 100 µl water/acetonitrile for HPLC/MS analysis.

All measurements were performed on a Dionex Ultimate 3000 RSLC system using a Waters BEH C18, 50 x 2.1 mm, 1.7 µm dp column. Two µl of sample was injected unless otherwise stated. Separation was achieved by a linear gradient with (A) H$_2$O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 µl/min and 45 °C. The gradient was initiated by a 0.33 min isocratic step at 5 % B followed by an increase to 95 % B in 9 min to end up with a 1.6 min flush step at 95 % B before re-equilibration with initial conditions. UV and MS detection were performed simultaneously. Coupling the HPLC to a MS was supported by an Advion Triversa Nanomate nano-ESI system attached to a Thermo Fisher Orbitrap. Mass spectra were acquired in centroid mode ranging from 100-500 m/z at a resolution of R = 30000. Single reaction monitoring using collision induced dissociation was performed on 247.11, 249.12, 265.12, and 267.12 m/z. All parent ions were isolated within a 2 m/z window and fragmented at 35 % CID energy.

The expected derivative of 4-methylproline could be detected in the derivatized extracts from the cultures harboring the *mps* expression constructs (pUWL201-mps and pET28-mps), but not in the derivatized extracts of the negative controls as shown in Fig. 8 exemplified with the heterologous expression in *S. coelicolor* (*S. coelicolor*/pUWL201-mps compared to *S. coelicolor* WT). By comparison of the retention times and high resolution mass data (including fragmentation data) with the derivatized PITC-(2S,4R)-4-methylproline reference substance, this compound could be clearly identified in the extracts of *mps* expression cultures.

This unambiguously proofs that the gene set *mps1-4* encodes all necessary enzymatic activities required for the biosynthesis of the griselimycin precursor 4-methylproline. In addition, deriva-

- 62 -

tives of two proposed intermediates of the pathway (5-hydroxyleucine and γ-methylglutamic acid γ-semialdehyde, see Fig. 8) could be detected providing further proof of the postulated enzymatic activities. The function of the putative leucine hydroxylase could be further verified by *in vitro* studies using the purified enzyme (as described above in Example 5). A derivative of the third postulated intermediate 3-methyl-Δ^1-pyrroline-5-carboxylic acid could not be detected as this compound cannot be derivatized with the applied method.

The successful heterologous expression of the *mps* genes in two different host strains (*E. coli* and *S. coelicolor*) also shows that this strategy can be applied to generate 4-methylproline, e.g. for organic synthesis approaches towards griselimycin or to feed the precursor to griselimycin/methylgriselimycin production cultures to enhance the methylgriselimycin yield.

Example 8: In silico analysis of A domain specificity

An in silico analysis of the amino acid specificity of the A domains of the NRPS proteins was performed according to Challis et al., 2000. Briefly the A domain sequences were aligned with the sequence of the phenylalanine-activating A domain (PheA) of the gramicidin S synthetase GrsA. Based on this alignment, the eight amino acids lining the binding pocket were identified and compared with the specificity-conferring amino acids from A domains with known substrate specificity. In the following Table 6 the specificity of the A domains of NRPS-1, NRPS-2 and NRPS-3 for substrates as predicted in silico is presented. In comparison, the actual substrates as they are activated for introduction into (methyl)griselimycin are listed.

"Residue position" are the eight amino acids of the individual A domains, which are thought to be important for substrate recognition ('specificity-conferring amino acids' or 'specificity-conferring code'). These residues were compared with the 'specificity-conferring codes' of A domains with known substrate specificity (http://nrps.igs.umaryland.edu/nrps/blast.html).

Table 6

A domain	Activated substrate(s)[a]	In silico substrate prediction[b]	Residue position[c]							
			265	266	269	278	299	301	326	330
NRPS-1-A1	Valine	Threonine	D	F	F	C	V	G	I	V
NRPS-1-A2	(4R)Me-Proline	Proline	D	V	Q	F	A	G	H	A
NRPS-1-A3	Threonine	Threonine	D	F	W	N	V	G	M	V
NRPS-1-A4	Leucine	Phenylalanine/Tryptophane	D	A	L	L	L	G	A	V
NRPS-1-A5	(4R)Me-Proline	Proline	D	V	Q	F	A	G	H	A
NRPS-1-A6	Leucine	Phenylalanine/Tryptophane	D	A	L	L	L	G	A	V
NRPS-2-A7	Valine	Threonine	D	F	F	C	V	G	I	V
NRPS-2-A8	Proline/(4R)Me-Pro	Proline	D	V	Q	F	A	A	H	V
NRPS-2-A9	Leucine	Phenylalanine/Tryptophane	D	A	L	L	L	G	A	V
NRPS-3-A10	Glycine	Glycine	D	I	L	Q	L	G	V	I
NosA-A3[d]	(4S)Me-Pro	Proline	D	V	Q	F	I	A	H	L
NcpB-A3[e]	(4S)Me-Pro	Proline	D	V	Q	F	I	A	H	V
NosD-A2[f]	Pro	Proline	D	V	Q	F	I	A	H	V

[a] According to griselimycin and methylgriselimycin primary structures

[b] According to the comparison of the specificity conferring code of other A domains
(http://nrps.igs.umaryland.edu/nrps/blast.html)

[c] The residues correspond to the gramicidin S synthetase PheA numbering (Challis et al., 2000)

[d] Accession number: AF204805_1 (Hoffmann et al., 2003)

[e] Accession number: AAO26334 (Luesch et al., 2003)

[f] Accession number: AAF17281 (Hoffmann et al., 2003)

The results for the A3 domain, the A10 domain and partially the A8 domain are consistent with the amino acids which are actually incorporated into (methyl)griselimycin. The remaining results do not seem to be in alignment with the actually incorporated amino acids, which may be due to the method applied. Based on the eight amino acid residues of the synthetases which were believed to be essential for the specificity of the substrate amino acid to be incorporated into griselimycin, it is not possible to differ between proline and methylproline specific A domains.

Example 9: Phylogenetic analysis of methylproline and proline specific domains

Phylogenetic analysis as presented in Fig. 9 of A domains which are specific for methylproline and proline demonstrate that A domains from methylproline incorporating NRPS modules (NRPS-1-A2 and NRPS-1-A5 and NcpB8(mPro) and NosA-A(mPro)) are more closely related

to each other than to the A domains of proline incorporating domains (NRPS-2-A8 and NosD(mPro)). This suggests, that the specificity is not only determined by the eight amino acid residues as presented in Table 6.

The analysis was performed using the Geneious Tree Builder function from the Geneious Pro
5   5.4.3 software (Tree Alignment Options: Cost Matrix (Blosum 62), Gap open penalty (12), Gap extension penalty (3) and Alignment type (Global alignment); Tree Builder Options: Genetic distance model (Jukes-Cantor), Tree build method (Neighbor-Joining), Outgroup (No outgroup)). The scale bar represents 0.09 substitutions per amino acid site.

10   Example 10: Specificity of A domains of modules 2, 5 and 8

In order to gain insight into the specificity of the A domains of modules 2, 5 and 8, the domains were recombinantly produced and their specificity analysed via an ATP/PPi exchange assay in vitro. Results are shown in Fig. 10. As expected, specificity of the domains A2 and A5 is highest for (2S,4R)-4-methylproline. Interestingly, this is also the case for domain A8, although L-
15   proline is preferably incorporated by module 8 during the biosynthesis of griselimycin. It is shown that there is also a high specificity for L-proline. It is assumed that the C domain of module 8 preferably condenses proline and introduces it into the growing peptide chain. Availability of methylproline during (methyl)griselimycin biosynthesis seems also to play a role as could be shown by feeding experiments. Feeding of 4-methylproline in concentrations from 0.025 to 1.0
20   g/l shows an increase of methylgriselimycin up to fivefold higher than the control. The experiments show also that the incorporation rate is not linear but reaches a saturation at about 0.5 g/l. Also the feeding with trans-hydroxyproline and trans-fluoroproline shows significant incorporation.

25   (a) Production of recombinant A2, A5 and A8 domains

In order to directly investigate A domain substrate specificity, we aimed to express A domains A2, A5 and A8 from the griselimycin pathway as N-terminally His$_6$-tagged proteins from pET28b (+). DNA fragments coding for the adenylation domains of A2, A5 and A8 were amplified from genomic DNA from the *Streptomyces DSM 22643* using the oligos A2_for 5'-
30   CCGACCATATGGATCCGGATGTGACGGTGGG-3' (SEQ ID NO: 44) and A2_rev 5'-ACCGGGAATTCGCCGACGGCGGGCAGGCCGA-3' (SEQ ID NO: 45) (due to the high ho-

mology, all three A domains can be amplified by using the same primer pair), and cloned into pJET2.1 vector by blunt end ligation.

Sequencing revealed that all of the A domains were correct. According to the molecular structure of griselimycin and the *in silico* analysis A domains A2 and A5 were expected to incorporate 2S,4R-methylproline. A domain A8 was expected to show a broader substrate specificity and either incorporate L-proline or 2S,4R-methylproline. To determine this *in vitro* we expressed the A domains in *E. coli* Rosetta pLys(RARE) at 16°C overnight. All three A domains could be obtained in the soluble fractions.

The proteins were then purified with Ni^+-affinity chromatography using a stepwise imidazol gradient. Pure protein could be obtained in fraction 8 for all A domains. Although the amount of protein was very low (0.2 mg/ml), it could be used in the ATP-PP$_i$ exchange assay to determine substrate specificity.

Substrate specificity of the 3 purified adenylation domains was evaluated using the established ATP-PP$_i$ exchange assay (Mootz and Marahiel. 1997, Thomas et al., 2002). Briefly, each protein was incubated with a panel of different amino acids, including the anticipated substrate of each domain. As a control, each protein was incubated in the absence of added amino acid, to determine the radioactivity background reaction. Nine different substrates were tested (see Fig. 10).

(b) Determination of substrate specificity by ATP-[^{32}P]PP$_i$ exchange assay

To determine substrate specificity, ATP-[^{32}P]PPi reactions (100 µl) containing Tris–HCl (pH 7.5, 75 mM), $MgCl_2$ (10 mM), dATP (5 mM), amino acid (5 mM), and protein (2 µg) were performed at 30°C. ^{32}P-tetrasodium pyrophosphate was obtained from Perkin Elmer (NEN #NEX019). The reactions were started by the addition of [^{32}P]PPi (0.1 µCi final amount) for up to 30 min before quenching with charcoal suspensions (500 µl, 1.6% [w/v] activated charcoal, 0.1 M $Na_4P_2O_7$, and 0.35 M perchloric acid in H_2O). The charcoal was pelleted by centrifugation prior to being washed twice with the washing solution 0.1 M $Na_4P_2O_7$ and 0.35 M perchloric acid in H_2O), resuspended in H_2O (500 µl), and counted by liquid scintillation (Beckman LS6500).

Example 11: Transcriptional regulators of the Xre-type encoded by orf-1 and orf-12 of the griselimycin biosynthetic cluster

- 66 -

To elucidate an involvement of two putative transcriptional regulators encoded by orf-1 and orf-12 in production of griselimycin inactivation mutants of the respective ORFs were generated. Subsequently, effects on production of griselimycin were determined by HPLC/MS analysis of extracts of the wild type strain ST105671 and the mutant strains ST105671-orf-1 and ST105671-orf-26, respectively. In order to determine whether effects on production observed in the inactivation mutants are due to direct regulation by the transcription regulators encoded by orf-1 and orf-12, the regulators were heterologously expressed and purified. The purified proteins were used in electrophoretic mobility shift assays (EMSA) to examine direct interaction to putative promoter regions of the griselimycin biosynthetic cluster.

(a) Inactivation of orf-1 and orf-12

The genes orf-1 and orf-12 were disrupted using derivatives of the pKC1132 plasmid integrating in orf-1 and orf-12 by homologous recombination. Two derivatives of pKC1132 (pKC1132-orf-1 and pKC1132-orf-26, respectively) were constructed carrying the central part of the coding region of either orf-1 or orf-12. Therefore, a 623 bp DNA fragment of orf-1 and a 1139 bp DNA fragment of orf-12 were amplified by PCR reactions using the primer combinations dxregrise_for/dxregrise_rev and dxreIIgrise_for/dxreIIgrise_rev, respectively (primer sequences see Table 8). EcoRI restriction sites were thereby attached to both ends of the fragments by the used primers. The DNA fragments and plasmid pKC1132 were digested using EcoRI and subsequently subjected to ligation. The ligations containing the desired plasmids pKC1132-orf-1 and pKC1132-orf-26, respectively, were transformed into electrocompetent E.coli DH10B cells. Transformants were selected on LB agar plates containing 60 μg/ml apramycin. Single colonies were grown in liquid LB medium in the presence of 60 μg/ml apramycin and plasmids were prepared. Plasmids carrying the correct inserts (pKC1132-orf-1 and pKC1132-orf-26) were identified by analytical digestion using EcoRI. E.coli ET12567 cells harboring the non-permissive plasmid pUZ8002 were then transformed using the correct plasmids pKC1132-orf-1 and pKC1132-orf-26. Transformants were selected on LB agar plates containing 60 μg/ml apramycin to yield strains E.coli ET12567pUZ8002pKC1132-orf-1 and E.coli ET12567pUZ8002pKC1132-orf-26. These strains were used to deliver the derivatives of plasmid pKC1132 into strain ST105671 by conjugation. The procedure was performed as follows. 2 ml of 2 day old liquid cultures of Streptomyces strain ST105671 grown in griselimycin production medium 5288 and in TSB medium at 30°C and 2 ml of the respective E.coli strains harboring plasmids pKC1132-orf-

1 or pKC1132-orf-26 (grown in LB containing 25 μg/ml kanamycin, 25 μg/ml chloramphenicol and 60 μg/ml apramycin) were pelleted, washed with sterile water and resuspended in 500 μl TSB each. 250 μl of *Streptomyces* strain ST105671 and the respective *E.coli* strain were mixed and incubated overnight under aeration at 30°C. 1 ml of TSB containing 60 μg/ml apramycin and

5   25 μg/ml nalidixic acid was added and further incubated overnight at 30°C under aeration. After incubation cells were pelleted, washed once with TSB and plated out on MS-agar containing 60 μg/ml apramycin and 25 μg/ml nalidixic acid, followed by further incubation at 30°C until conjugants became visible. Correct integration of the plasmids pKC1132-orf-1 or pKC1132-orf-26 in the coding regions of orf-1 or orf-12, respectively, was verified by performing PCR using ge-

10 nomic DNA of the conjugants and combinations of primers lacZ1, lacZ2 and grixreI_g1, grixreI_g2 in case of disruption of orf-1 or grixreII_g1, grixreII_g2 in case of disruption of orf-12, respectively. The inactivation mutants of ST105671-orf-1 and ST105671-orf-26 generated by this procedure were used to determine effects on griselimycin production compared to the ST105671 predecessor.

15

(b) HPLC/MS analysis of extracts of ST105671-orf-1 and ST105671-orf-26

Liquid cultures of four independent inactivation mutants of orf-1 and of two independent inactivation mutants of orf-12 were grown for two days in TSB liquid medium containing 60 μg/ml apramycin. Using these cultures, 100 ml of fresh TSB medium without antibiotic were inocu-

20 lated 1:50 in duplicates each and grown for 3 days. Wild type ST105671 cells were treated equally without using apramycin in the preculture. Cells were pelleted by centrifugation and the supernatant was used for ethylacetate extraction. To monitor effects on griselimycin production the extracts were subsequently analyzed by HPLC/MS analysis. Quantities of griselimycin of each mutant and the wild type strain ST105671 relative to the total protein amount of each cul-

25 ture were determined. Therefore, the respective peak areas were integrated and the counts were set relative to the total protein amount of the pelleted cells. Mean values and standard deviations were calculated. Inactivation of orf-1 led to a more than 200 fold decrease in griselimycin production. No griselimycin could be detected in the inactivation mutant of orf-12 (see Table 7). The observations show an involvement of the Xre-type transcriptional regulators encoded by orf-

30 1 and orf-12 in regulation of griselimycin production. Whether these effects are a result of direct regulation of promoters within the griselimycin biosynthetic cluster by binding of the Xre-type

- 68 -

regulators or depict an indirect effect should be elucidated by performing DNA protein interaction studies (electrophoretic mobility shift assays, EMSA).

Table 7

strain	no. of replicates	mean (rel. intensity)	std dev
ST105671	4	$7,38^{*}10^{9}$	$4,59^{*}10^{9}$
ST105671-orf-1	8	$3,19^{*}10^{7}$	$1,41^{*}10^{7}$
ST105671-orf-26	8	-	-

5   Table 7: Effects on griselimycin production in mutants ST105671-orf-1 and ST105671-orf-26. Quantities of griselimycin of each mutant and the wild type strain ST105671 are shown as mean values of counts of extracts obtained relative to the whole protein content from independently grown replicate cultures.

10   (c) Verification of specific binding of the transcription regulators Orf-1 and Orf-12 to the promoter regions of nrps1 and nrps2

To examine a direct transcriptional regulation of griselimycin biosynthetic genes by regulators Orf-1 and Orf-12, possible binding to the putative promoter regions upstream of genes nrps1 and nrps2 was determined. Therefore, the coding regions of orf-1 and orf-12 were first amplified in

15   PCR reactions with primer combinations xreI-NdeI-for/xreI-HindIII-rev and xreII-NdeI-for/xreII-HindIII-rev using genomic DNA of strain ST105671 as a template. *Nde*I restriction sites were thereby attached to the 5' ends and *Hind*III restriction sites attached to the 3' ends of the resulting DNA fragments. Plasmid pET22b and the amplified DNA fragments carrying genes orf-1 and orf-12 were digested with *Nde*I and *Hind*III. Subsequently, *Nde*I/*Hind*III digested orf-1

20   and orf-12 were ligated with *Nde*I/*Hind*III digested pET22b in two separate attempts. Ligations were transformed into *E.coli* DH10B and transformants selected on LB agar plates containing 200 µg/ml ampicillin. Plasmids were prepared and correct integration of the inserts (orf-1 and orf-12, respectively) into the plasmids was verified by *Hind*III/*Nde*I digests. Plasmids carrying the inserts were finally subjected to DNA sequencing in order to exclude mutations in the DNA

25   sequence. Due to the cloning strategy, the resulting plasmids pET22b-orf-1 and pET22b-orf-26 encoded for C-terminally 6 fold His tagged (6xHis) Orf-1 and Orf-12, respectively. Both plasmids were transformed into *E.coli* BL21(DE3) by electroporation and transformants were selected on LB agar plates containing 200 µg/ml ampicillin. Expression of the C-terminal 6xHis

tag fusion proteins Orf-1-His and Orf-12-His was performed in LB liquid medium containing 200 µg/ml ampicillin in the presence of 0.1 mM IPTG at 16°C overnight. Cells were lysed by sonication and purification of Orf-1-His (~33kDa) and Orf-12-His (~47kDa) was achieved by using Ni^{2+} affinity columns.

5 Putative promoter region of genes nrps1 (581 bp region upstream of nrps1) and nrps2 (632 bp region upstream of nrps2) were amplified in PCR reactions using primer combinations prom4hex2_rev/prom4hex3_for and prom5hex2_rev/prom5hex3_for, respectively, to yield DNA fragments carrying 18 bp linker regions on both ends. In a second PCR reaction these DNA fragments were labeled with a Hex fluorophor at the 5' end of each strand. At this, primers

10 (Hex2 and Hex3) carrying a 5' Hex fluorophor and being reverse complementary to the 18 bp linker regions attached in the first PCR reactions were used. Hex labeled DNA fragments spanning the above mentioned upstream regions of genes nrps1 and nrps2 were thereby generated and referred to as Pnrps1 and Pnrps2. In a parallel attempt, a 696 bp intergenic region flanked by genes mxan_3950 and mxan_3951 of the annotated genome of *Myxococcus xanthus* DK1622

15 comprising a dual promoter system was amplified with primers EH33951_rev/EH33950_rev using genomic DNA of the respective strain. As the primers carried the same 18 bp linker regions like mentioned above, the DNA fragment could subsequently be labeled with the Hex fluorophor at both 5' ends like described before. As specific binding of regulators Orf-1-His and Orf-12-His to this promoter region out of the genome of *Myxococcus xanthus* DK1622 cannot be

20 expected, the resulting DNA fragment referred to as Pmxan_ctrl served as a negative control in the EMSA assays.

These DNA fragments as well as both purified regulators (Orf-1-His and Orf-12-His) were subsequently used in EMSA assays. In brief, the assays were performed as follows. 4% native PAA-gels were prepared in 1x buffer HG (25 mM HEPES, 192 mM glycin, pH 7.3 with KOH). Pock-

25 ets of the gel were rinsed and electrophoresis was performed for 30 min at 120 V without samples in 1x buffer HG. Samples (100 fMol of each DNA-fragment Pnrps1, Pnrps2 and Pmxan_ctrl with and without 500 fold molar excess of either Orf-1-His or Orf-12-His and 2.5 µg high molecular weight salmon sperm DNA as competitor DNA) were prepared and incubated for 10 min at 30°C. Samples were loaded on the gel and electrophoresis was performed for 2 h at 120 V.

30 Detection of the DNA fragments was done using a Typhoon 9410 imager (Amersham) at the respective wavelength. In the presence of Orf-1-His, a retardation of each Hex labeled DNA fragment compared to the attempts without Orf-1-His could be observed in electrophoresis indi-

cating a non specific binding of regulator Orf-1-His to DNA comprising a promoter region (Fig.11). On the other hand, Orf-12-His exhibited no ability to interact with any of the DNA fragments demonstrating no specificity towards the Hex labeled DNA used in the assays. Taken together these results suggested that neither Orf-1-His nor Orf-12-His are direct transcriptional
5   regulators of the promoters Pnrps1 and Pnrps2.

Xre-type regulators Orf-1 and Orf-12 show remarkable aspects regarding their predicted domain organization which was determined using the SMART tool at http://smart.embl-heidelberg.de/ (Fig. 12). As expected, both regulators possess an N-terminal Xre-type helix-turn-helix (HTH) motif whereby Orf-1 obviously contains a second domain of this kind (HTH2). In addition to
10   this, both regulators are characterized by an extended C-terminus with unknown function.

As it was possible, that these extraordinary features are part of regulatory mechanisms which influence affinity and thereby specificity towards distinct DNA sequences of regulators Orf-1 and Orf-12 the EMSA assays were repeated using C-terminally shortened versions of the regulators. In case of Orf-1, these versions only comprised HTH1, HTH2 or both predicted HTH do-
15   mains. The respective proteins were named ORF1-HTH1 (~9.6 kDa), ORF1-HTH2 (~9.8 kDa) and ORF1-HTH12 (16.4 kDa). In case of Orf-12, the shortened version only contained the N-terminal HTH domain named ORF26-HTH (~9.9kDa). Regulatory effects (e.g. binding of unknown cofactors or protein-protein interactions) resulting in conformational changes of the C-terminus or the whole protein and thereby influencing DNA binding affinity of the regulators
20   should be excluded.

The shortened versions of Orf-1 and Orf-12 were obtained by cloning of the respective 5' regions of orf-1 and orf-12 as *NdeI/Hin*dIII fragments into *NdeI/Hin*dIII digested pET22b to yield the encoded C-terminal 6xHis tagged fusion proteins. Cloning was performed as described in case of the full length Orf-1 and Orf-12. Primer pairs used for amplification of the inserts to be
25   cloned were xreI-NdeI-for/xreIhth1rev (Orf-1-HTH1), xreIhth2for/xreIhth12rev (Orf-1-HTH2), xreI-NdeI-for/xreIhth12rev (Orf-1-HTH12) and xreII-NdeI-for/xreIIhthrev (Orf-12-HTH). Heterologous expression, purification and subsequent verification of binding to the above mentioned promoter regions of genes nrps1 and nrps2 was performed as described for the full length proteins. The EMSA assays performed using these shortened versions of the regulators revealed a
30   specific binding of ORF1-HTH12 to promoter Pnrps1 and of Orf-12-HTH to both promoter regions Pnrps1 and Pnrps2 (Fig. 13).

Only by co-incubation of Orf-1-HTH12 and Pnrps1 a significant retardation in electrophoresis of the DNA could be achieved. This retardation could not be observed using the negative control Pmxan_ctrl. Co-incubation with Pnrps2 gave a less clear result whereby no distinct statement can be made about direct binding of Orf-1-HTH12 to promoter region Pnrps2. Orf-12-HTH co-incubated with Pnrps1 or Pnrps2 resulted in clear distinct retardations of the DNA fragments whereas in case of the negative control Pmxan_ctrl a shift was completely absent.

Based on theses experimental results, roles of the C-termini of both regulators remain to be elucidated but seem to be related to modification of DNA binding specificity of Orf-1 and Orf-12. Nevertheless, based on the effects observed after inactivation of orf-1 or orf-12 on production of griselimycin, the direct specific binding of Orf-1-HTH12 to promoter Pnrps1 and of Orf-12-HTH to promoters Pnrps1 and Pnrps2, the Xre-type transcriptional regulators encoded by orf-1 and orf-12 are clearly functionally related to griselimycin biosynthesis and can be considered as a part of the secondary metabolite cluster.

Table 8

primer name	sequence 5' —> 3'	SEQ ID NO
dxregrise_for	cgGAATTCTGAGAAGCTGCCTGCAC	46
dxregrise_rev	cgGAATTCCTCATAGACCTGAGCGGC	47
dxreIIgrise_for	cgGAATTCCGTAGGACACGGAAACTGTC	48
dxreIIgrise_rev	cgGAATTCGCATCCATCCTGAGCAG	49
EH33950_rev	GGAATGGGCCGGGTACTGCAGCCTGCTATTT CTCAGAATC	50
EH33951_rev	GGAATGGGCCGGGTACTGAGAG-CACGGGGTGTG	51
grixreI_g1	ATCGCCTTCGGCGACAG	52
grixreI_g2	GCTGGGCGTATCCGACTC	53
grixreII_g1	GAAGGGACAGGTCCACCG	54
grixreII_g2	GGTGTCCTGCACGGACAG	55
hex2	GCTCTCGGTGCCCTTGTG	56
hex3	GGAATGGGCCGGGTACTG	57
lacZ1	CTTGGGCTGCAGGTCGAC	58
lacZ2	GTGTGGAATTGTGAGCGG	59
prom4hex2_rev	GCTCTCGGTGCCCTTGTGTTCAC-TATCCCCATGTTGTTTCC	60
prom4hex3_for	GGAATGGGCCGGGTACTGCTGCTGACTGAAC-CAGATCTCAC	61
prom5hex2_rev	GCTCTCGGTGCCCTTGTGGATCGAACAC-GGCTCAGCTC	62
prom5hex3_for	GGAATGGGCCGGGTACTGGCAG-CACTTCATGGTCGC	63

xreI-HindIII-rev	cgtAAGCTTCTCGTTCGCGCCCGTTTC	64
xreIhthI2rev	cgtAAGCTTGGCGATCGACCGGTCTTC	65
xreIhthIrev	cgtAAGCTTATCAGTCAGATCGGGCGG	66
xreIhth2for	gtaaggCATATGGATCTGTTCCCGCGGTCC	67
xreII-HindIII-rev	cgtAAGCTTGGTTGTGCTAATGCCCGTCC	68
xreIIhthrev	cgtAAGCTTCTCCGCTTCCGGCTCAC	69
xreII-NdeI-for	gtaaggCATAT-GAGTGAAAACCTGACGCTGGGAG	70
xreI-NdeI-for	gtaaggCATATGTCTGCTCGTCACTTCGACCG	71

Table 8: List of primers used in the experiments. Name and DNA sequence in 5' to 3' orientation of oligonucleotides are shown. Letters with simple underlining: Restriction sites attached to the primers. Letters with double underlining: 5' linker regions attached to the respective sequence. Small letters: Nucleotides attached to enhance restriction efficiency.

Example 12: *Streptomyces DSM 22643*

Streptomyces DSM 22643 was deposited at 06th of April 2009 according to the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure with the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Inhoffenstraße 7B, 38124 Braunschweig, Germany on June 04, 2009 by Sanofi-Aventis Deutschland GmbH, Industriepark Höchst, 65926 Frankfurt/Main, Germany.

REFERENCES

Altschul S.F. et al., Nucleic Acids Res., 1997, 25, 3389-3402

Barry S.M. and Challis G.L., Curr. Opin. Chem. Biol., 2009, 13, 205-215

Becker D. et al., Eur. J. Biochem. 1997, 249, 739-747

Bibb, M.J. et al., Gene 1985, 38, 215–266

Bibb, M.J. et al., Molecular Microbiology 1994, 14, 533-545

Bierman M. et al., Gene 1992,116, 43-49

Challis G. L. et al., Chem. Biol. 2000, 7, 211-264

Eriko T. et al., Gene, 1995,66, 133- 137

Ferna´ ndez-Moreno M.A. et al., Cell, 1991, 66, 769–780

Finking R. and Marahiel M.A., Annu. Rev. Microbiol., 2004; 58, 453-488

Gaitatzis N. et al., Proc. Natl. Acad. Sci. USA, 2001, 98, 11136-11141

- 73 -

Gust B et al., Proc. Natl. Acad. Sci. USA, 2003, 100, 1541-1546

Hahn M. and Stachelhaus T.; Proc. Natl. Acad. Sci. USA, 2004, 101, 15585-15590;. Epub 2004 Oct 21

Herai S. et al., Proc. .Natl. Acad. Sci. USA 2004, 101, 14031-14035

Hoffmann D. et al., Gene, 2003, 311, 171-180

Kieser et al., *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich 2000

Li W. et al., Appl. Environ. Microbiol., 2009, 75, 2869-2878

Luesch H. et al., J. Org. Chem., 2003, 68, 83-91

Marahiel M.A., J. Pept. Sci., 2009, 15, 799-807

Mervyn J. et al., Molecular Microbiology, 1994, 14(3), 533-545

Mootz H.D. and Marahiel M. A., J. Bacteriol., 1997, 179, 6843-6850

Müller C. et al., Antimicrob. Agents Chemother., 2007, 51, 1028-1037

Murakami, T. et al., J. Bacteriol., 1989, 171, 1459-1466

Peschke U. et al., Mol. Microbiol., 1995, 16, 1137-1156

Sambrook et al., Molecular cloning: A laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2001

Takano, E. et al., Gene, 1995, 166, 133–137

Terlain B. and Thomas J.-P., Bulletin de la Sociétié Chemique de France, 1971, 6, 2657-2662

Thomas M. G. et al., Chem. Biol. 2002, 9, 171-184.

Wehmeier U. F:, Gene, 1995, 165, 149-150

Yin X. et al., Chem. Biol. Chem. 2004, 5, 1274

- 74 -

CLAIMS

1.   Nucleic acid comprising at least one nucleic acid selected from:

(a)   a nucleic acid comprising at least one of the Open Reading Frames (ORFs) 1 to 26 as comprised by SEQ ID NO: 1 encoding proteins of SEQ ID NOs: 2 to 27 or a variant or fragment thereof whereby the variant or fragment encodes a functionally active variant or fragment of a protein of SEQ ID NOs: 2 to 27,

(b)   a nucleic acid encoding at least one of the proteins of SEQ ID NOs: 2 to 27 or a functionally active variant or fragment thereof,

(c)   a nucleic acid encoding a protein that is at least 70 %, 80 %, 90 %, 95 % or 97% identical in amino acid sequence to a protein or fragment thereof encoded by the nucleic acid of (a) or (b),

(d)   a nucleic acid which hybridizes under stringent conditions with a nucleic acid of (a) to (c),

(e)   a nucleic acid which hybridizes under stringent conditions with a nucleic acid of (a) to (d) and consists of 10 to 50 nucleotides in length, or

(f)   a nucleic acid which is complementary to a nucleic acid of (a) to (f).

2.   The nucleic acid of claim 1, wherein said nucleic acid comprises at least two ORFs, preferably at least three ORFs encoding any of the proteins of SEQ ID NOs: 2 to 27 or a functionally active variant or fragment thereof as referred to in items (a) to (d) of claim 1.

3.   The nucleic acid of claim 1 or 2, wherein said nucleic acid comprises at least one of the ORFs 18, 19, 20 and 21 encoding proteins of SEQ ID NO: 19, 20, 21 and 22, respectively, or a functionally active variant or fragment thereof as referred to in items (a) to (d) of claim 1.

4.   The nucleic acid of any one of claims 1 to 3, wherein said nucleic acid comprises at least one of the ORFs 8, 15 and 16 encoding proteins of SEQ ID NO: 9, 16 and 17, respectively, or a functionally active variant or fragment thereof as referred to in items (a) to (d) of claim 1.

5.   The nucleic acid according to any one of claims 1 to 4 comprising or consisting of the griselimycin biosynthesis gene cluster having the sequence of SEQ ID NO: 1 or having a variant sequence of SEQ ID NO: 1 harboring a variant or fragment of at least one of ORFs 1 to 26

- 75 -

whereby the variant or fragment encodes a functionally active variant or fragment of a protein of SEQ ID NOs: 2 to 27.

6.   An expression vector comprising a nucleic acid of any one of claims 1 to 5.

7.   A host cell transformed with an expression vector of claim 6.

8.   A protein comprising at least one amino acid sequence selected from SEQ ID NOs: 2 to 27 or a functionally active variant or fragment thereof or encoded by a nucleic acid of any one of claims 1 to 5.

9.   An antibody specifically directed against at least one of the proteins or a functionally active variant or fragment thereof of claim 8.

10.   Use of the nucleic acid of any one of claims 1 to 5, the expression vector of claim 6, the host cell of claim 7 or the protein of claim 8, in particular of at least one of the proteins of SEQ ID NOs: 9, 16 and 17, for the production of griselimycin and/or methylgriselimycin.

11.   Use of the nucleic acid of claim 3, the expression vector of claim 6 comprising the nucleic acid of claim 3, the host cell of claim 7 transformed with said expression vector or of at least one of the proteins of SEQ ID NOs: 19, 20, 21 and 22 for the production of L-trans-4-methylproline.

12.   A method of producing at least one of the proteins or a functionally active variant or fragment thereof of claim 8 comprising:
a)   expressing the nucleic acid of any one of claims 1 to 5, and
b)   optionally isolating the protein or a functionally active variant or fragment thereof.

13.   A method of determining a variant of ORFs 1 to 26 as comprised by SEQ ID NO: 1 encoding the proteins of SEQ ID NOs: 2 to 27 or of the proteins of SEQ ID NOs: 2 to 27 which modulates production of griselimycin and/or methylgriselimycin comprising
a)   producing a variant of any of ORFs 1 to 26,

b)   expressing the variant into a protein,

c)   allowing the production of griselimycin and/or methylgriselimycin or of a derivative thereof using the protein of b), and

d1)   determining the amount of griselimycin and/or methylgriselimycin,

wherein an increase of the amount of griselimycin and/or methylgriselimycin compared to the amount produced if the protein of b) is not present, but the corresponding invariant form of the protein is present is indicative that the variant is capable of increasing the production of griselimycin and/or methylgriselimycin, or

d2)   determining the antibiotic activity of the derivative of griselimycin and/or methylgriselimycin,

wherein an increase in the antibiotic activity compared to the antibiotic activity if the protein of b) is not present, but the corresponding invariant form of the protein is present is indicative that the variant is capable of producing a derivative of griselimycin and/or methylgriselimycin with increased antibiotic activity,

and wherein the derivative differs in amino acid composition from griselimycin and/or methylgriselimycin and exhibits an enhanced antibiotic activity over griselimycin and/or methylgriselimycin.

14.   A method of determining a variant of ORFs 1 to 26 as comprised by SEQ ID NO: 1 encoding proteins of SEQ ID NOs: 2 to 27 or of the proteins of SEQ ID NOs: 2 to 27 which enhances production of L-trans-4-methylproline comprising

a)   producing a variant of any of ORFs 1 to 26,

b)   expressing the variant into a protein,

c)   allowing the production of L-trans-4-methylproline using the protein of b), and

d)   determining the amount of L-trans-4-methylproline,

wherein an increase of the amount of L-trans-4-methylproline compared to the amount produced if the protein of b) is not present, but the corresponding invariant form of the protein is present is indicative that the variant is capable of increasing the production of L-trans-4-methylproline.

15.   A method of producing griselimycin and/or methylgriselimycin comprising

a)   providing at least one protein of claim 8 or a functionally active variant or fragment thereof or providing a variant protein obtainable by the method of claim 13 or 14, and

- 77 -

b)   incubating the at least one protein of a) with L-N-methylvaline, L-leucine, L-N-methylthreonine, L-proline, L-N-methylleucine and glycine to produce griselimycin and/or methylgriselimycin.

16.   A method of producing L-trans-4-methylproline comprising

a)   providing at least one protein of claim 8 or a functionally active variant or fragment thereof or providing a variant protein obtainable by the method of claim 14, and

b)   incubating the at least one protein of a) with L-leucine, 5-hydroxyleucine, γ-methylglutamic acid γ-semialdehyde and/or 3-methyl-Δ^1-pyrroline-5-carboxylic acid to produce L-trans-4-methylproline.

17.   *Streptomyces DSM 22643.*

(A)
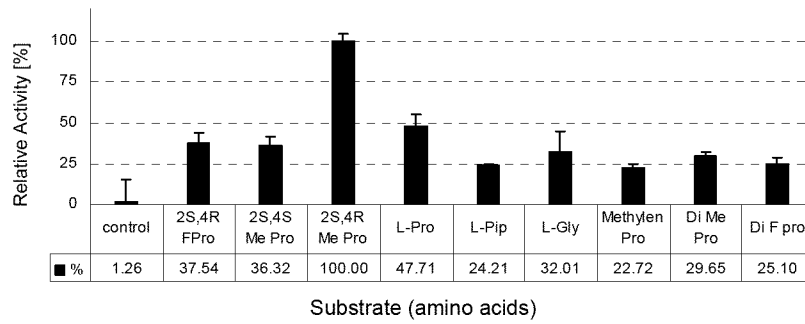

(B)
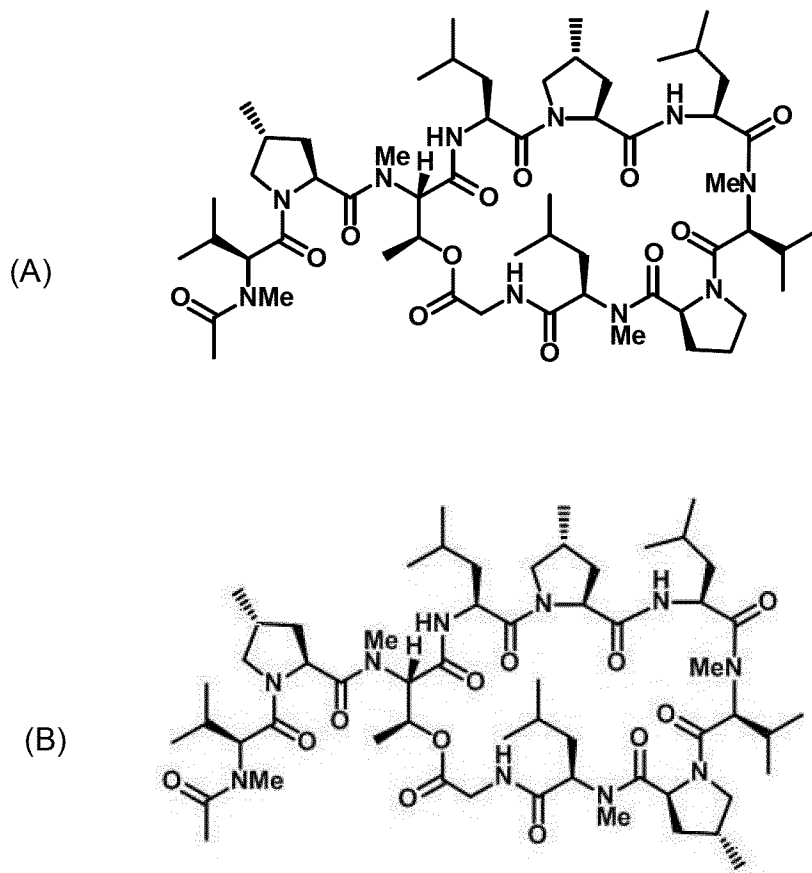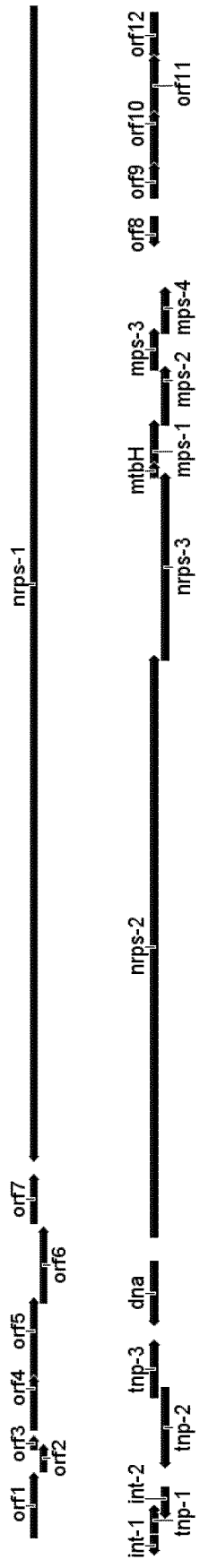
Fig. 1

Fig. 2
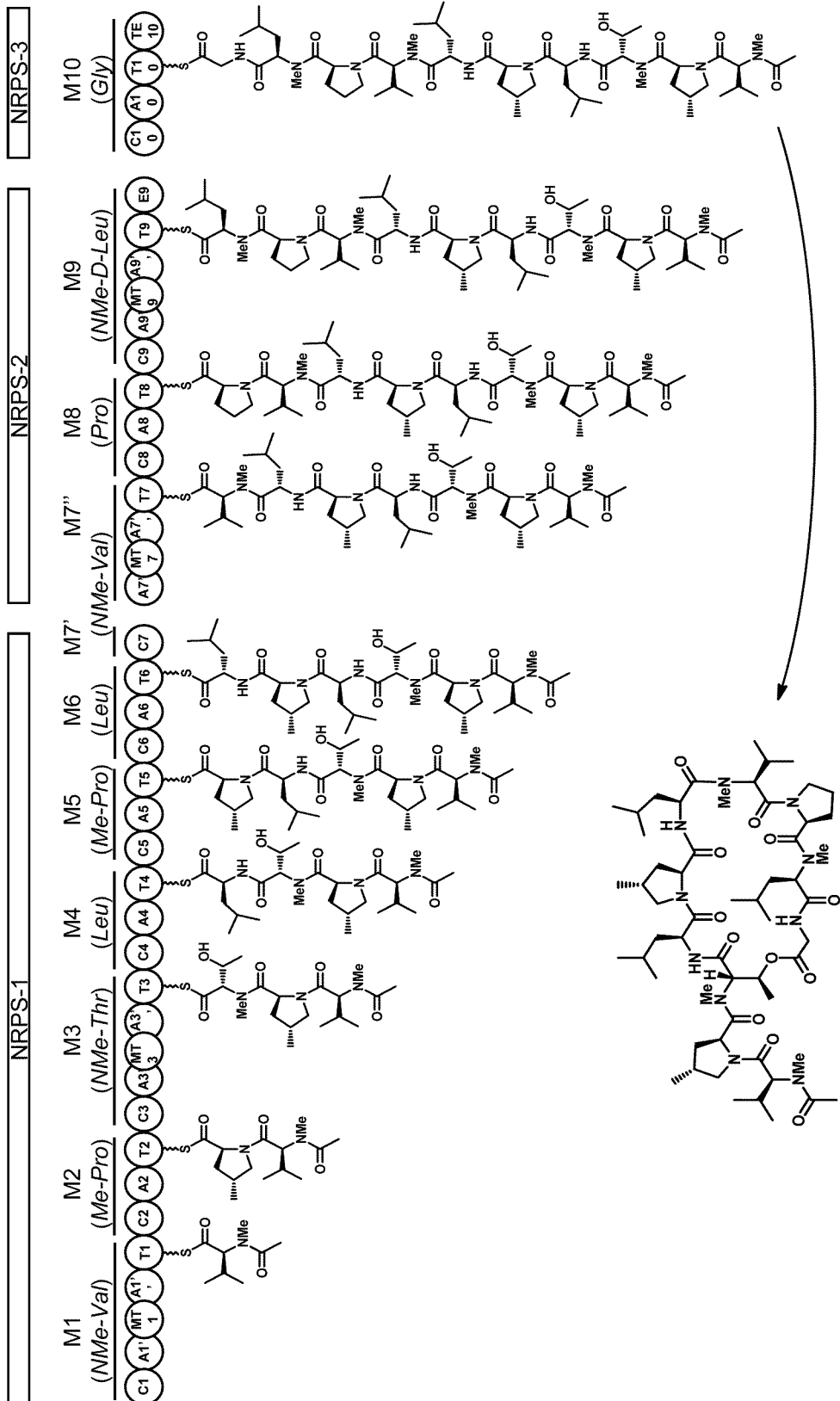
Fig. 3
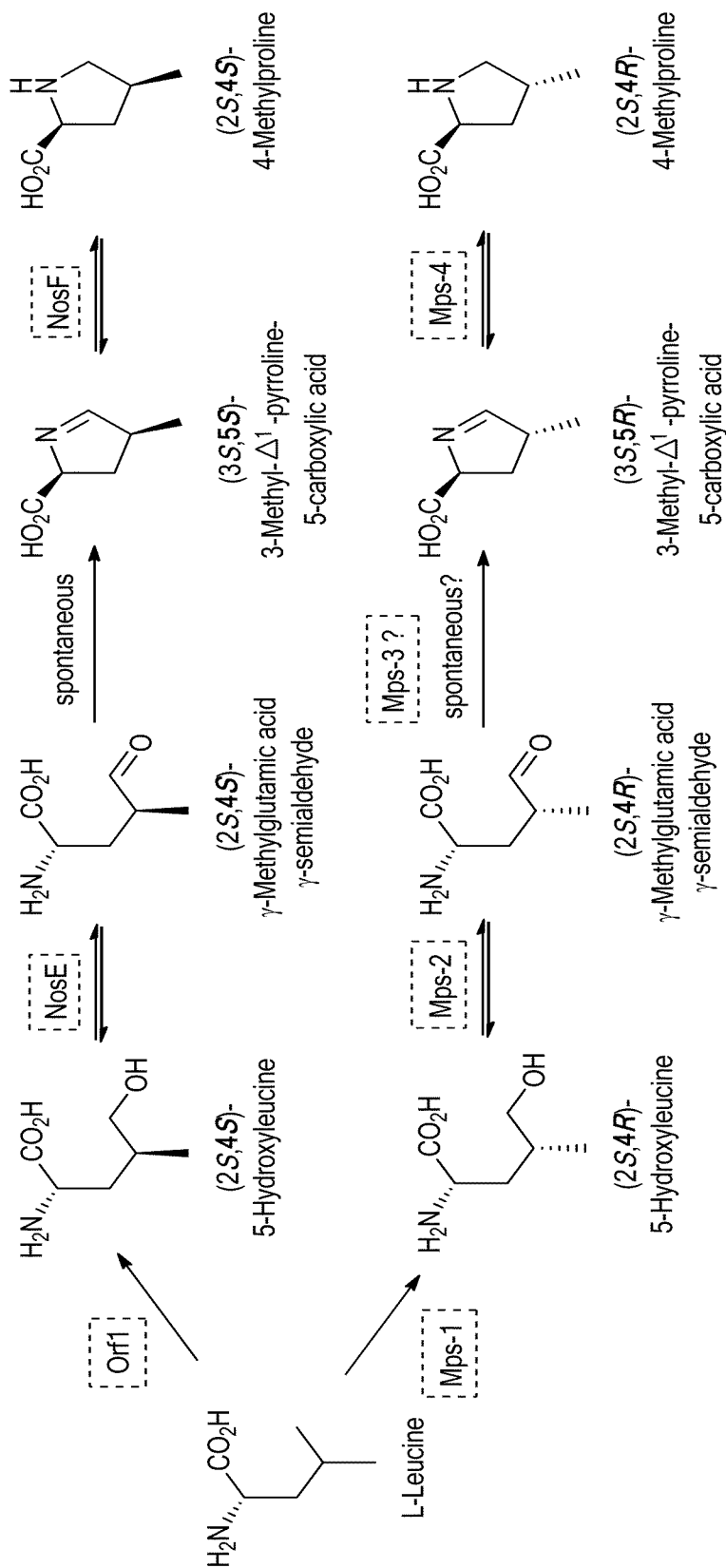
Fig. 4
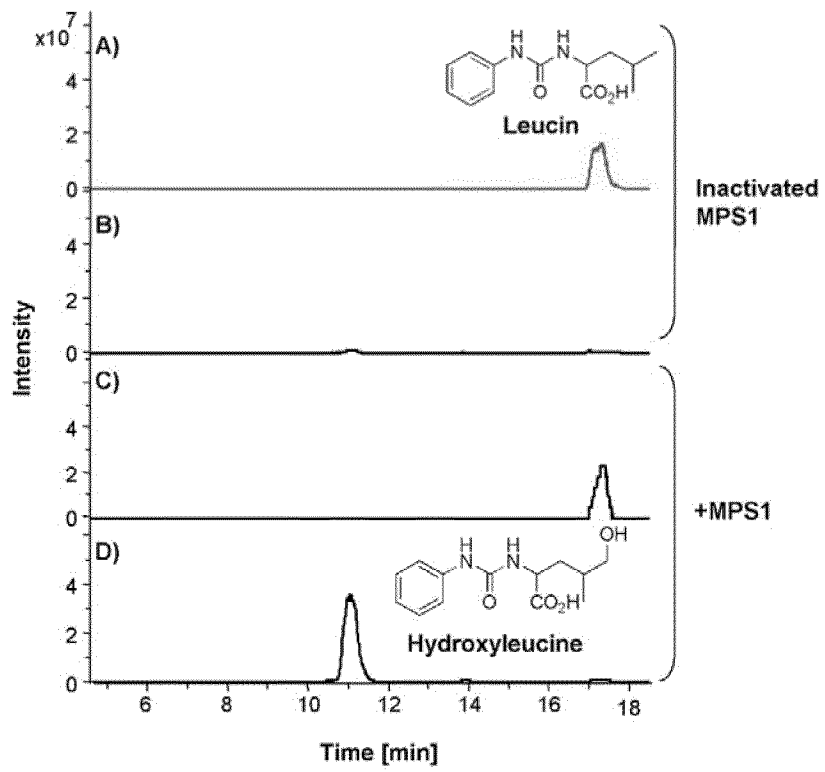
Fig. 5
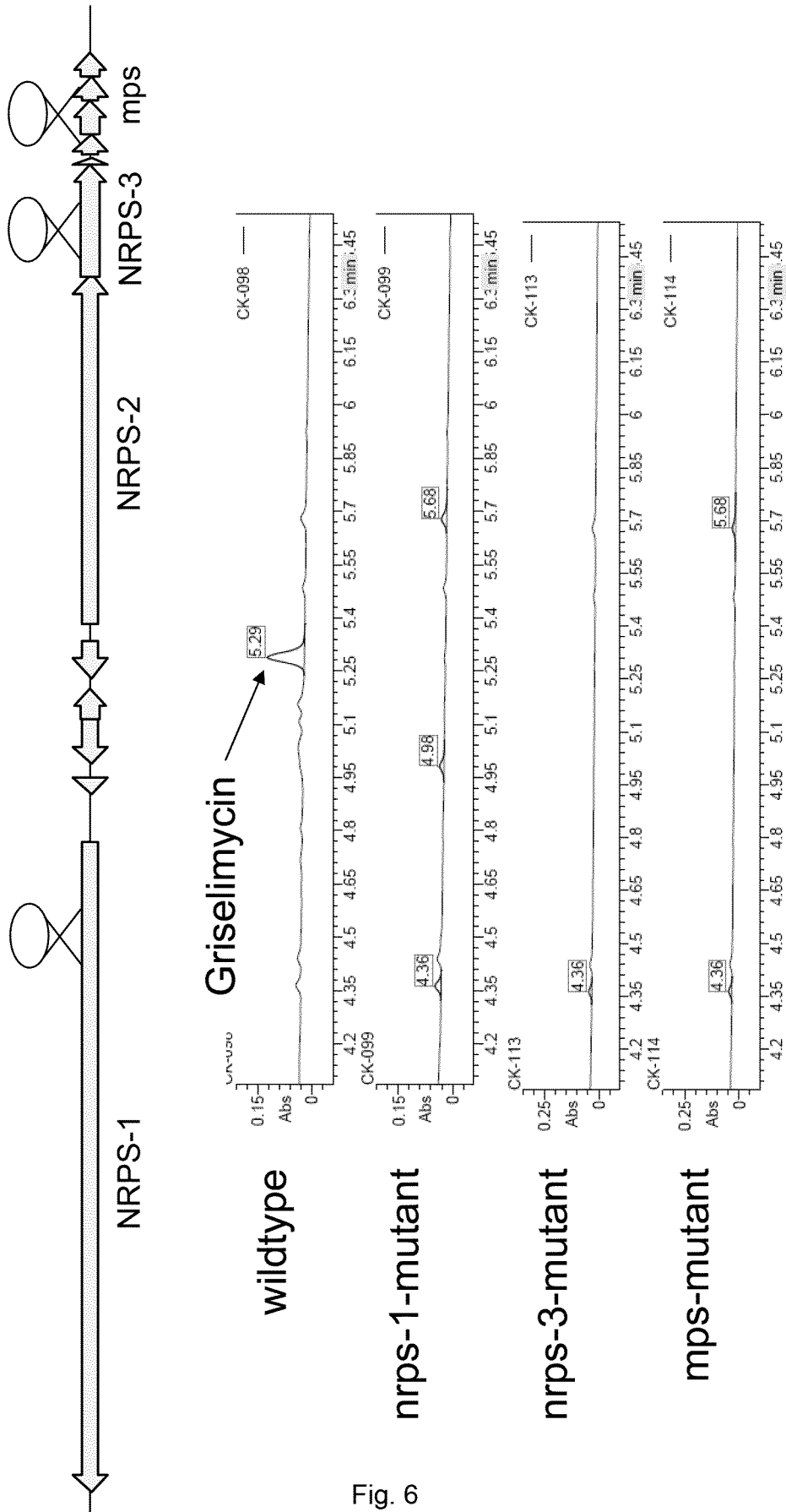
Fig. 6
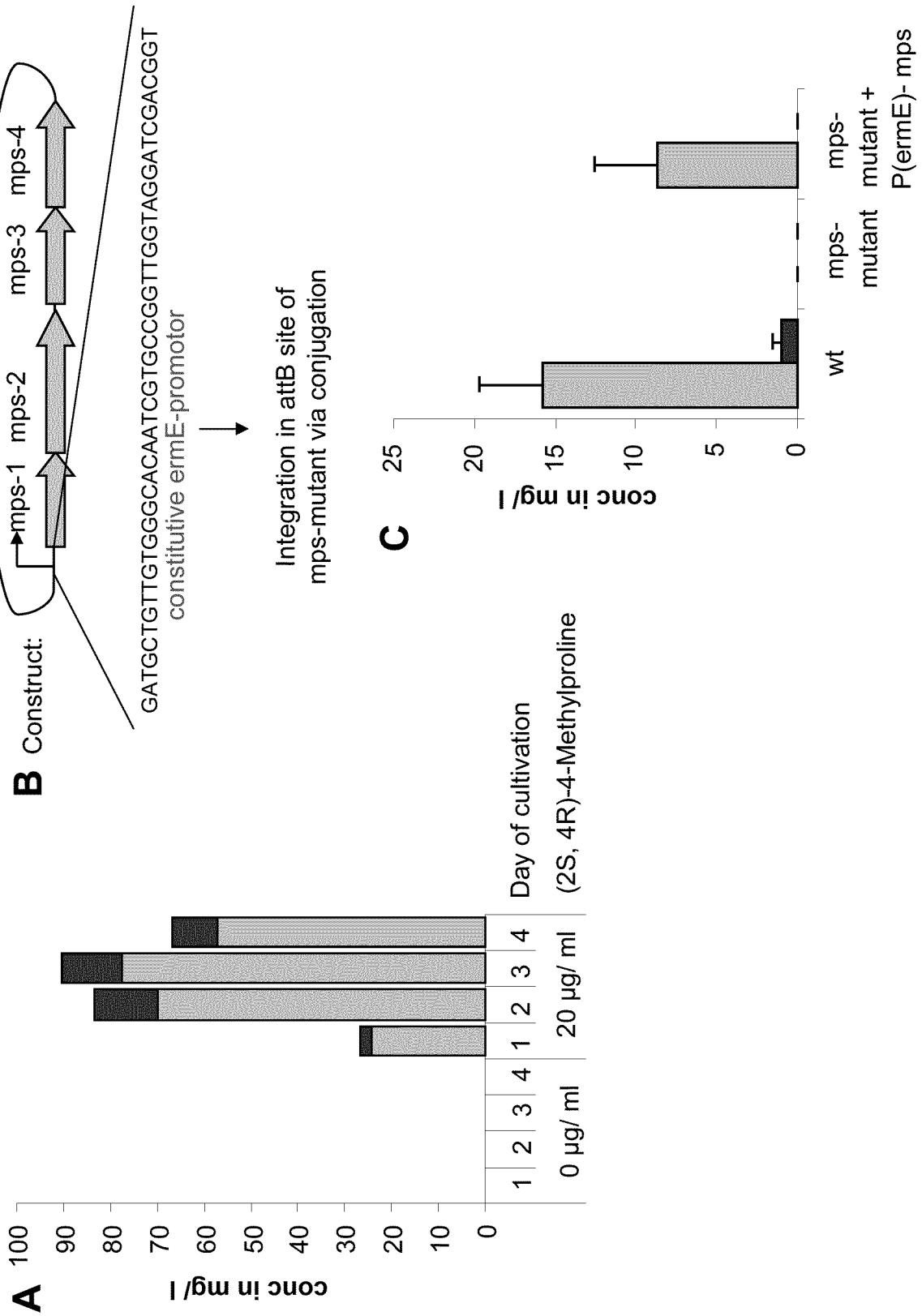
Fig. 7
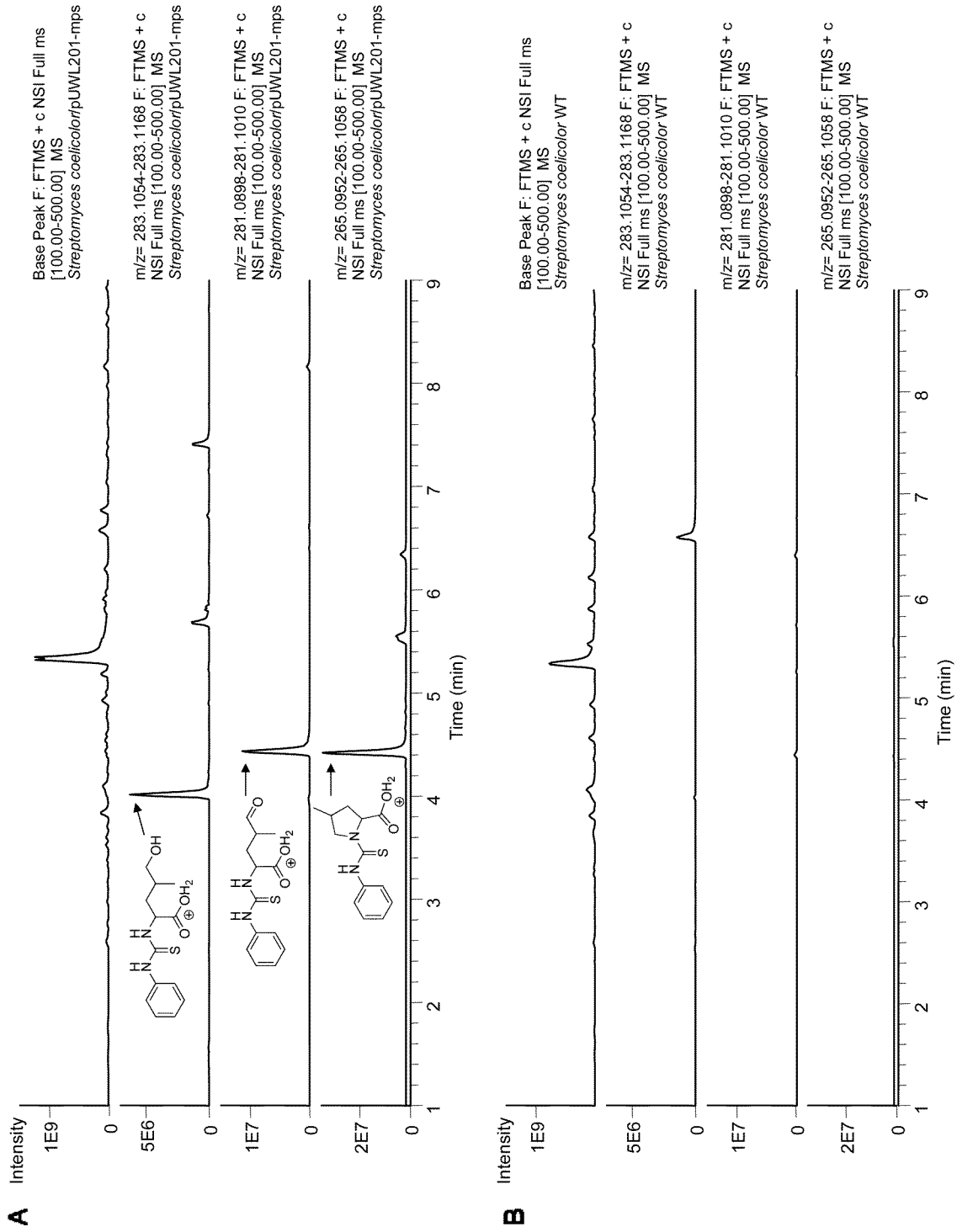
Fig. 8
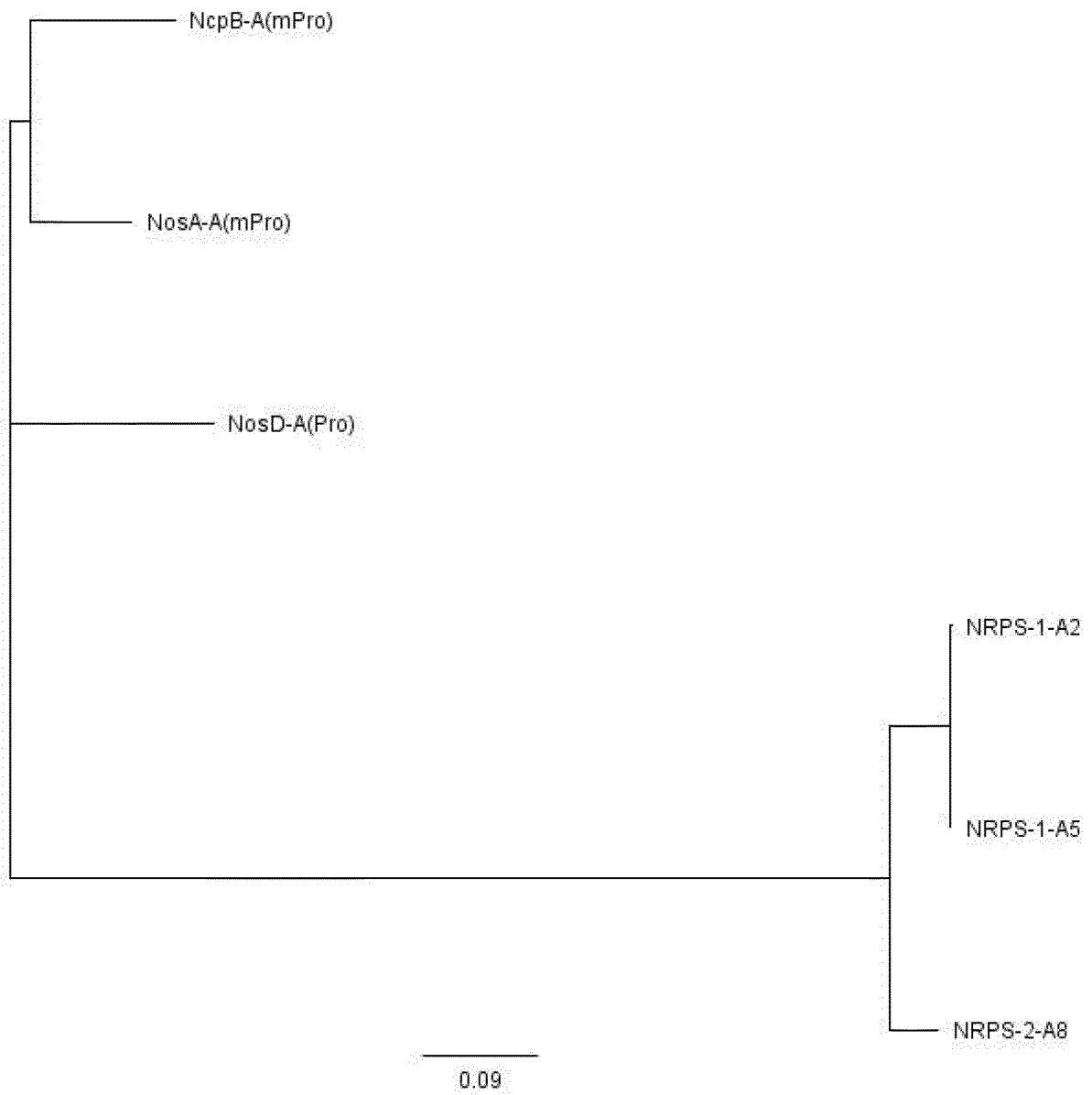
Fig. 9

Fig. 10

Fig.
11

Fig. 12

Fig. 13