



(51) International Patent Classification:
C07K 14/47 (2006.01)

(21) International Application Number:
PCT/GB2011/051395

(22) International Filing Date:
22 July 2011 (22.07.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1012900.5 30 July 2010 (30.07.2010) GB

(71) Applicant (for all designated States except US): **ISIS INNOVATION LIMITED** [GB/GB]; Ewert House, Ewert Place, Summertown, Oxford Oxfordshire OX2 7SG (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **WYSZYNSKI, Filip** [GB/GB]; University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford Oxfordshire OX1 3TA (GB). **DAVIS, Benjamin Guy** [GB/GB]; University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford Oxfordshire OX1 3TA (GB). **BIBB, Mervyn** [GB/GB]; Department of Molecular Microbiology, John Innes Centre, Norwich Norfolk NR4 7UH (GB). **HESKETH, Andrew** [GB/GB]; Department of Biochemistry, University of Cambridge, Building 0, Downing Site, Cambridge Cambridgeshire CB2 1QW (GB).

(74) Agent: **JA Kemp & Co**; 14 South Square, Gray's Inn, London London WC1R 5JJ (GB).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: TUNICAMYCIN GENE CLUSTER

(57) Abstract: The invention relates to a gene cluster for the production of tunicamycins and derivatives thereof. The invention also relates to individual polynucleotides from the gene cluster, variants thereof and the polypeptides they encode. The invention further relates to heterologo-us expressions systems and their use to produce a tunicamycin or a derivative thereof.



TUNICAMYCIN GENE CLUSTER

Field of the invention

The invention relates to a gene cluster for the production of tunicamycins and derivatives thereof. The invention also relates to individual polynucleotides from the gene cluster, variants thereof and the polypeptides they encode. The invention further relates to heterologous expressions systems and their use to produce a tunicamycin or a derivative thereof.

Background of the invention

The tunicamycins are fatty acyl nucleoside antibiotics, first isolated from the soil actinomycete *Streptomyces lysosuperificus* in 1971 and later from *Streptomyces charteusis* (G. Tamura, *Tunicamycin*, Japan Scientific Societies Press, Tokyo, 1982; *U. S. Patent*, 4237225, 1980; and A. Takatsuki, K. Arima and G. Tamura, *J. Antibiot.*, 1971, **24**, 215). Their structures consist of an unusual eleven carbon aminodialdose core (tunicamine) to which uracil and *N*-acetylglucosamine (GlcNAc) are anomerically attached, alongside a range of amide-linked unsaturated fatty acids (Figure 1). A number of related natural products belong to the tunicamycin family – namely streptovirudins, corynetoxins, MM19290, mycospocidin and antibiotic 24010. All share the conserved carbohydrate core and presumably have similar biosynthetic pathways, but the genes required for their production have not been identified.

The tunicamycins are potent inhibitors of bacterial cell wall biosynthesis, targeting MraY which catalyses the formation of the key peptidoglycan precursor undecaprenyl-pyrophosphoryl-*N*-acetylmuramoyl pentapeptide (lipid I), a key peptidoglycan precursor. Unfortunately, these antibiotics have not been used clinically due to cytotoxicity against mammalian cells, associated with inhibition of eukaryotic protein *N*-glycosylation. Specific binding to the active site of UDP-GlcNAc:dolichyl phosphate GlcNAc-1-phosphate transferase (GPT) blocks production of the lipid-linked precursor dolichyl-pyrophosphoryl-*N*-acetyl-glucosamine (Dol-PP-GlcNAc) and teminates asparagine-linked glycoprotein synthesis at the first committed step. This property has also led to the widespread use of tunicamycin as a crucial tool in the study of glycoproteins. A number of synthetic studies towards the tunicamycins have been published, with two full syntheses (A. G. Myers, D. Y. Gin and D. H. Rogers, *J. Am. Chem. Soc.*, 1994, **116**, 4697 and T. Suami, H. Sasai, K. Matsuno and N. Suzuki, *Carbohydr. Res.*, 1985, **143**, 85), and preliminary biosynthetic investigations with labeled precursors have suggested the metabolic origin of some parts of the molecule (B. C. Tsvetanova, D. J. Kiemle and N. P. J. Price, *J. Biol. Chem.*, 2002, **277**, 35289). Despite this interest, the absence of a

sequence for the tunicamycin gene cluster (or any part of it) has hindered understanding of the biosynthetic pathway and it has remained poorly defined 40 years after tunicamycin's first isolation.

Summary of the invention

5 The inventors have surprisingly identified the tunicamycin gene cluster from *Streptomyces chartreusis* (SEQ ID NO: 1). It contains 14 genes (tunA to tunN), each of which encodes a polypeptide enzyme. The inventors have also confirmed that heterologous expression of this gene cluster in a *Streptomyces coelicolor* host confers tunicamycin production.

10 The identification of the tunicamycin gene cluster not only allows tunicamycins to be produced more efficiently, for instance in more efficient host cells, but also allows tunicamycin derivatives to be produced. In particular, variants of one or more of the 14 polypeptide enzymes in the cluster can be designed to have an altered substrate specificity. Such variants can then be used to attach different side groups to the tunicamycin scaffold and thereby form derivatives of tunicamycin. Tunicamycin derivatives having, for instance, altered carbohydrate and/or fatty acid groups can be
15 used to target enzymes involved in bacterial cell wall biosynthesis other than MraY (the enzyme inhibited by tunicamycins). Tunicamycin derivatives can therefore be used to target bacteria that differ from those targeted by tunicamycins. Tunicamycin derivatives which inhibit MraY to a greater degree and/or do not inhibit eukaryotic protein *N*-glycosylation can also be designed.

20 Accordingly, the invention provides a polynucleotide sequence comprising the sequence shown in nucleotides 8616 to 20606 of SEQ ID NO: 1 or a variant having at least 50% homology to nucleotides 8616 to 20606 of SEQ ID NO: 1 over its entire sequence based on nucleotide identity.

The invention further provides a polynucleotide sequence comprising:

the sequence shown in SEQ ID NO: 2 or a variant having at least 75% homology to SEQ ID NO: 2 over its entire sequence based on nucleotide identity;

- 25 (ii) the sequence shown in SEQ ID NO: 4 or a variant having at least 85% homology to SEQ ID NO: 4 over its entire sequence based on nucleotide identity;
- (iii) the sequence shown in SEQ ID NO: 6 or a variant having at least 71% homology to SEQ ID NO: 6 over its entire sequence based on nucleotide identity;
- (iv) the sequence shown in SEQ ID NO: 8 or a variant having at least 72% homology to
30 SEQ ID NO: 8 over its entire sequence based on nucleotide identity;
- (v) the sequence shown in SEQ ID NO: 10 or a variant having at least 76% homology to SEQ ID NO: 10 over its entire sequence based on nucleotide identity;
- (vi) the sequence shown in SEQ ID NO: 12 or a variant having at least 75% homology to SEQ ID NO: 12 over its entire sequence based on nucleotide identity;

- (vii) the sequence shown in SEQ ID NO: 14 or a variant having at least 70% homology to SEQ ID NO: 14 over its entire sequence based on nucleotide identity;
- (viii) the sequence shown in SEQ ID NO: 16 or a variant having at least 71% homology to SEQ ID NO: 16 over its entire sequence based on nucleotide identity;
- 5 (ix) the sequence shown in SEQ ID NO: 18 or a variant having at least 76% homology to SEQ ID NO: 18 over its entire sequence based on nucleotide identity;
- (x) the sequence shown in SEQ ID NO: 20 or a variant having at least 77% homology to SEQ ID NO: 20 over its entire sequence based on nucleotide identity;
- (xi) the sequence shown in SEQ ID NO: 22 or a variant having at least 79% homology to SEQ ID NO: 22 over its entire sequence based on nucleotide identity;
- 10 (xii) the sequence shown in SEQ ID NO: 24 or a variant having at least 66% homology to SEQ ID NO: 24 over its entire sequence based on nucleotide identity;
- (xiii) the sequence shown in SEQ ID NO: 26 or a variant having at least 72% homology to SEQ ID NO: 26 over its entire sequence based on nucleotide identity; or
- 15 (xiv) the sequence shown in SEQ ID NO: 28 or a variant having at least 68% homology to SEQ ID NO: 28 over its entire sequence based on nucleotide identity.

The invention further provides:

- a polynucleotide construct comprising more than one of the polynucleotide sequences of the invention;
- 20 - a vector comprising a polynucleotide sequence of the invention or a polynucleotide construct of the invention operably linked to a control sequence
- a host cell comprising a polynucleotide sequence, polynucleotide construct or a vector of the invention;
- a polypeptide encoded by a polynucleotide of the invention;
- 25 - a heretologous expression system encoded by a polynucleotide of the invention;
- a heterologous expression system comprising more than one of the polypeptides of the invention; and
- a host cell comprising a heterologous expression system of the invention.

The invention additionally provides a polypeptide sequence comprising:

- 30 (a) the sequence shown in SEQ ID NO: 3 or a variant having at least 73% homology to SEQ ID NO: 3 over its entire sequence based on amino acid identity;
- (b) the sequence shown in SEQ ID NO: 5 or a variant having at least 91% homology to SEQ ID NO: 5 over its entire sequence based on amino acid identity;

- (c) the sequence shown in SEQ ID NO: 7 or a variant having at least 61% homology to SEQ ID NO: 7 over its entire sequence based on amino acid identity;
- (d) the sequence shown in SEQ ID NO: 9 or a variant having at least 64% homology to SEQ ID NO: 9 over its entire sequence based on amino acid identity;
- 5 (e) the sequence shown in SEQ ID NO: 11 or a variant having at least 78% homology to SEQ ID NO: 11 over its entire sequence based on amino acid identity;
- (f) the sequence shown in SEQ ID NO: 13 or a variant having at least 77% homology to SEQ ID NO: 13 over its entire sequence based on amino acid identity;
- (g) the sequence shown in SEQ ID NO: 15 or a variant having at least 66% homology to
10 SEQ ID NO: 15 over its entire sequence based on amino acid identity;
- (h) the sequence shown in SEQ ID NO: 17 or a variant having at least 67% homology to SEQ ID NO: 17 over its entire sequence based on amino acid identity;
- (i) the sequence shown in SEQ ID NO: 19 or a variant having at least 78% homology to SEQ ID NO: 19 over its entire sequence based on amino acid identity;
- 15 (j) the sequence shown in SEQ ID NO: 21 or a variant having at least 77% homology to SEQ ID NO: 21 over its entire sequence based on amino acid identity;
- (k) the sequence shown in SEQ ID NO: 23 or a variant having at least 66% homology to SEQ ID NO: 23 over its entire sequence based on amino acid identity;
- (l) the sequence shown in SEQ ID NO: 25 or a variant having at least 53% homology to
20 SEQ ID NO: 25 over its entire sequence based on amino acid identity;
- (m) the sequence shown in SEQ ID NO: 27 or a variant having at least 55% homology to SEQ ID NO: 27 over its entire sequence based on amino acid identity; or
- (n) the sequence shown in SEQ ID NO: 29 or a variant having at least 69% homology to SEQ ID NO: 29 over its entire sequence based on amino acid identity.

25 The invention additionally provides a method of producing a tunicamycin or a derivative thereof, the method comprising culturing a host cell of the invention and isolating the tunicamycin or derivative thereof.

The invention also provides a tunicamycin or a derivative thereof produced using a method of the invention.

30 The invention additionally provides a pharmaceutical composition comprising a tunicamycin or a derivative thereof of the invention and a pharmaceutically acceptable carrier.

Also provided is a tunicamycin or a derivative thereof of the invention for use in a method of treatment of the human or animal body.

The invention additionally provides a tunicamycin or a derivative thereof of the invention for use in a method of treating or preventing a bacterial infection in a subject.

A method of treating or preventing a bacterial infection in a subject comprising administering to said subject a therapeutically or prophylactically effective amount of a tunicamycin or a derivative of the invention.

Description of the Figures

Figure 1 shows the structures of the tunicamycins.

Figure 2 shows the genetic organisation of the tunicamycin biosynthetic gene cluster in *S. chartreusis* and its homologues in *S. clavuligerus* and *A. miriuma*.

Figure 3 shows evidence of heterologous production of tunicamycins in *S. coelicolor*. (A): Bioassay showing heterologous expression of (i) a genomic library-derived cosmid harboring the *tun* gene cluster introduced into a *S. coelicolor* M1152 host (giving recombinant strains *S. coelicolor* M1027 and M1028 derived from library cosmids 6N9 and 7C3, respectively) and control strain *S. coelicolor* M1030 (containing the same cosmid but without any insert sequence) and (ii) the minimal *tun* gene cluster cloned into pRT802 in *S. coelicolor* M1146 (giving recombinant strain *S. coelicolor* M1035) and control strain *S. coelicolor* M1031 (containing the empty pRT802 cosmid); (B): LC/MS analysis of (i) an authentic tunicamycin sample and mycelium extracts of these recombinant *S. coelicolor* strains (ii) M1031, (iii) M1035, (iv) M1027 and (v) M1030. See also Fig. S3 for ¹H NMR analysis of extracts.

Figure 4 shows the proposed biosynthetic pathway for the tunicamycins.

Description of the Sequence Listing

SEQ ID NO: 1 shows the tunicamycin gene cluster from *Streptomyces chartreusis*. It contains 14 genes labeled tunA to tun N. TunA corresponds to nucleotides 8616-9578. TunB corresponds to nucleotides 9581-10594. TunC corresponds to nucleotides 10601-11554. TunD corresponds to nucleotides 11557-12978. TunE corresponds to nucleotides 12978-13679. TunF corresponds to nucleotides 13679-14659. TunG corresponds to nucleotides 14664-15272. TunH corresponds to nucleotides 15272-16816. TunI corresponds to nucleotides 16822-17733. TunJ corresponds to nucleotides 17720-18505. TunK corresponds to nucleotides 18548-18790. TunL corresponds to nucleotides 18790-19476. TunM corresponds to nucleotides 19487-20134. TunN corresponds to nucleotides 20151-20606.

SEQ ID NO: 2 shows the polynucleotide sequence of tunA.

SEQ ID NO: 3 shows the amino acid sequence of tunA.

SEQ ID NO: 4 shows the polynucleotide sequence of tunB.

SEQ ID NO: 5 shows the amino acid sequence of tunB.

SEQ ID NO: 6 shows the polynucleotide sequence of tunC.

SEQ ID NO: 7 shows the amino acid sequence of tunC.

5 SEQ ID NO: 8 shows the polynucleotide sequence of tunD.

SEQ ID NO: 9 shows the amino acid sequence of tunD.

SEQ ID NO: 10 shows the polynucleotide sequence of tunE.

SEQ ID NO: 11 shows the amino acid sequence of tunE.

SEQ ID NO: 12 shows the polynucleotide sequence of tunF.

10 SEQ ID NO: 13 shows the amino acid sequence of tunF.

SEQ ID NO: 14 shows the polynucleotide sequence of tunG.

SEQ ID NO: 15 shows the amino acid sequence of tunG.

SEQ ID NO: 16 shows the polynucleotide sequence of tunH.

SEQ ID NO: 17 shows the amino acid sequence of tunI.

15 SEQ ID NO: 18 shows the polynucleotide sequence of tunI.

SEQ ID NO: 19 shows the amino acid sequence of tunJ.

SEQ ID NO: 20 shows the polynucleotide sequence of tunK.

SEQ ID NO: 21 shows the amino acid sequence of tunK.

SEQ ID NO: 22 shows the polynucleotide sequence of tunL.

20 SEQ ID NO: 23 shows the amino acid sequence of tunL.

SEQ ID NO: 24 shows the polynucleotide sequence of tunM.

SEQ ID NO: 25 shows the amino acid sequence of tunM.

SEQ ID NO: 26 shows the polynucleotide sequence of tunN.

SEQ ID NO: 27 shows the amino acid sequence of tunN.

25

Detailed description of the invention

It is to be understood that different applications of the disclosed products and methods may be tailored to the specific needs in the art. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not
30 intended to be limiting.

In addition as used in this specification and the appended claims, the singular forms “a”, “an”, and “the” include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to “a polynucleotide” includes “polynucleotides”, reference to “a polypeptide”

includes two or more such polypeptides, reference to “a host cell” includes two or more such host cells, and the like.

All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

5

Polynucleotides

The inventors have surprisingly identified the tunicamycin gene cluster from *Streptomyces chartreusis* (SEQ ID NO: 1). They have also confirmed that heterologous expression of this gene cluster in a *Streptomyces coelicolor* host confers tunicamycin production.

10 The identification of the tunicamycin gene cluster in *Streptomyces chartreusis* was not straightforward. Natural product gene clusters, particularly those of polyketide or non-ribosomal peptide origin, have often been identified by PCR amplification of highly conserved signature genes using degenerate primers, followed by screening of genomic libraries for the presence of these sequences. However, tunicamycin is unlikely to require a large number of genes for its production
15 and few of its biosynthetic genes can be predicted with enough precision to confidently assign a particular genetic homologue as a highly conserved probe sequence for degenerate primer design. For this reason, *de novo* genome scanning of a known tunicamycin producer together with ‘filtered’ genome mining was used instead as a rapid, more direct method to identify the genes for tunicamycin biosynthesis. The specific method used by the inventors is clearly set out in the Example below.
20 Essentially, the inventors sequenced the entire genome of the *Streptomyces chartreusis* bacterium and identified the tunicamycin gene cluster *in silico*. They then prepared numerous cosmids from the genomic library (each containing about 0.1% of the bacterium’s genome) and isolated the cosmid containing the tunicamycin gene cluster. The ability of the cluster to confer tunicamycin production was then confirmed by transfecting a different bacterium (*Streptomyces coelicolor*) with the cluster.

25 It is surprising that the tunicamycin gene cluster contains only 14 genes and about 12 kilobases. This is much smaller than other clusters known in the art.

The invention provides a polynucleotide sequence comprising the sequence shown in nucleotides 8616 to 20606 of SEQ ID NO: 1 or a variant having at least 50% homology to nucleotides 8616 to 20606 of SEQ ID NO: 1 over its entire sequence based on nucleotide identity.
30 Nucleotides 8616 to 20606 of SEQ ID NO: 1 corresponds to the 14 genes within the tunicamycin gene cluster, namely tunA to tunN. The polynucleotide sequence may comprise the whole of SEQ ID NO: 1 or a variant having at least 50% homology to SEQ ID NO: 1 over its entire sequence based on nucleotide identity. The polynucleotide sequence may consist of the sequence shown in

nucleotides 8616 to 20606 of SEQ ID NO: 1 or a variant having at least 50% homology to nucleotides 8616 to 20606 of SEQ ID NO: 1 over its entire sequence based on nucleotide identity.

A variant having at least 50% homology to (1) nucleotides 8616 to 20606 of SEQ ID NO: 1 or (2) the whole of SEQ ID NO: 1 over its entire sequence based on nucleotide identity retains the ability to express a system of enzymes that is capable of producing a tunicamycin or a derivative thereof. The variant preferably expresses at least 12 enzymes, most preferably SEQ ID NOs: 3, 5, 7, 9, 11, 13, 15, 17, 23, 25, 27, 29 or variants thereof as discussed below (i.e. do not express tunI and tunJ or variants thereof). The variant preferably expresses all of SEQ ID NOs: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 and 29 or variants thereof as discussed below. The ability of a system of enzymes to produce a tunicamycin or a derivative thereof can be assayed using any method known in the art. For instance, the ability of a variant to express a system of enzymes that is capable of producing a tunicamycin or a derivative thereof can be assayed as described in the Example. In particular, the variant is expressed in a host cell and the ability of the host cell to produce a tunicamycin or a derivative thereof is assayed.

The variant of nucleotides 8616 to 20606 of SEQ ID NO: 1 or the whole of SEQ ID NO: 1 typically includes modifications that alter the substrate sensitivity of one or more, such as 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or all, of the 14 enzymes. Such modifications are discussed in more detail below.

The polynucleotide sequence may be a naturally occurring variant which is expressed by an organism, for instance by a bacterium different from *Streptomyces chartreusis*. Variants also include non-naturally occurring variants produced by recombinant technology. Over the entire length of the polynucleotide sequence of nucleotides 8616 to 20606 of SEQ ID NO: 1 or the whole of SEQ ID NO: 1, a variant is at least 50% homologous to that sequence based on nucleotide identity. More preferably, the variant may be at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90% and more preferably at least 95%, 97% or 99% homologous based on nucleotide identity to nucleotides 8616 to 20606 of SEQ ID NO: 1 or the whole of SEQ ID NO: 1 over the entire sequence. The variant preferably comprises 14 regions having at least 70%, for example at least 75%, 80%, 85%, 90% or 95%, nucleotide identity with nucleotides 8616-9578, 9581-10594, 10601-11554, 11557-12978, 12978-13679, 13679-14659, 14664-15272, 15272-16816, 16822-17733, 17720-18505, 18548-18790, 18790-19476, 19487-20134 and 20151-20606 of SEQ ID NO: 1.

Standard methods in the art may be used to determine homology. For example the UWGCG Package provides the BESTFIT program which can be used to calculate homology, for example used on its default settings (Devereux *et al* (1984) *Nucleic Acids Research* **12**, p387-395). The PILEUP

and BLAST algorithms can be used to calculate homology or line up sequences (such as identifying equivalent residues or corresponding sequences (typically on their default settings)), for example as described in Altschul S. F. (1993) *J Mol Evol* 36:290-300; Altschul, S.F *et al* (1990) *J Mol Biol* 215:403-10.

5 Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pair (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighbourhood word score
10 threshold (Altschul *et al*, supra). These initial neighbourhood word hits act as seeds for initiating searches to find HSP's containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extensions for the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of
15 one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

20 The BLAST algorithm performs a statistical analysis of the similarity between two sequences; see e.g., Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90: 5873-5787. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two amino acid sequences would occur by chance. For example, a sequence is considered similar to another sequence if the
25 smallest sum probability in comparison of the first sequence to the second sequence is less than about 1, preferably less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

 Any number of nucleotide substitutions may be made to the sequence of nucleotides 8616 to 20606 of SEQ ID NO: 1 or the whole of SEQ ID NO: 1, for example up to 100, 500, 300, 1000 or
30 5000 substitutions. Codons within the sequence may be replaced with different codons that encode the same amino acid (i.e. with (i.e. degenerate codons).

 Codons may be replaced such that conservative substitutions are introduced into the expressed polypeptides. Conservative substitutions replace amino acids with other amino acids of similar chemical structure, similar chemical properties or similar side-chain volume. The amino

acids introduced may have similar polarity, hydrophilicity, hydrophobicity, basicity, acidity, neutrality or charge to the amino acids they replace. Alternatively, the conservative substitution may introduce another amino acid that is aromatic or aliphatic in the place of a pre-existing aromatic or aliphatic amino acid. Conservative amino acid changes are well-known in the art and may be selected in accordance with the properties of the 20 main amino acids as defined in Table 1 below. Where amino acids have similar polarity, this can also be determined by reference to the hydropathy scale for amino acid side chains in Table 2.

Table 1 – Chemical properties of amino acids

Ala	aliphatic, hydrophobic, neutral	Met	hydrophobic, neutral
Cys	polar, hydrophobic, neutral	Asn	polar, hydrophilic, neutral
Asp	polar, hydrophilic, charged (-)	Pro	hydrophobic, neutral
Glu	polar, hydrophilic, charged (-)	Gln	polar, hydrophilic, neutral
Phe	aromatic, hydrophobic, neutral	Arg	polar, hydrophilic, charged (+)
Gly	aliphatic, neutral	Ser	polar, hydrophilic, neutral
His	aromatic, polar, hydrophilic, charged (+)	Thr	polar, hydrophilic, neutral
Ile	aliphatic, hydrophobic, neutral	Val	aliphatic, hydrophobic, neutral
Lys	polar, hydrophilic, charged(+)	Trp	aromatic, hydrophobic, neutral
Leu	aliphatic, hydrophobic, neutral	Tyr	aromatic, polar, hydrophobic

Table 2 - Hydropathy scale

Side Chain	Hydropathy
Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8
Cys	2.5
Met	1.9
Ala	1.8
Gly	-0.4
Thr	-0.7
Ser	-0.8
Trp	-0.9
Tyr	-1.3
Pro	-1.6
His	-3.2
Glu	-3.5

	Gln	-3.5
	Asp	-3.5
	Asn	-3.5
	Lys	-3.9
5	Arg	-4.5

One or more nucleotides may additionally be deleted from of nucleotides 8616 to 20606 of SEQ ID NO: 1 or SEQ ID NO: 1. Up to 10, 20, 30, 40, 50, 100, 200 or 300 residues may be deleted, or more.

10 Variants may be fragments of nucleotides 8616 to 20606 of SEQ ID NO: 1 or SEQ ID NO: 1. Such fragments retain the ability discussed above. A fragment preferably comprises nucleotides 8616-9578, 9581-10594, 10601-11554, 11557-12978, 12978-13679, 13679-14659, 14664-15272, 15272-16816, 16822-17733, 17720-18505, 18548-18790, 18790-19476, 19487-20134 and 20151-20606 of SEQ ID NO: 1.

15 One or more nucleotides may be alternatively or additionally added to the polynucleotides described above. An extension may be provided at the 5' or 3' of 8616 to 20606 of SEQ ID NO: 1 or SEQ ID NO: 1. The extension may be quite short, for example from 3 to 30 nucleotides in length. Alternatively, the extension may be longer, for example up to 150 or 300 nucleotides. The extension may be a control sequence at the 5' end.

20 As discussed above, a variant of nucleotides 8616 to 20606 of SEQ ID NO: 1 or the whole of SEQ ID NO: 1 retains the ability to express a system of enzymes that is capable of producing a tunicamycin or a derivative thereof. A variant typically comprises nucleotides 8616-9578, 9581-10594, 10601-11554, 11557-12978, 12978-13679, 13679-14659, 14664-15272, 15272-16816, 16822-17733, 17720-18505, 18548-18790, 18790-19476, 19487-20134 and 20151-20606 of SEQ ID
25 NO: 1 that encode the 14 enzymes. A variant typically includes one or more modifications, such as substitutions, additions or deletions, outside these regions.

The invention also provides a polynucleotide sequence comprising the sequence of one of the 14 tun genes present in SEQ ID NO: 1 or a variant thereof. In particular, the invention provides:

- (i) the sequence shown in SEQ ID NO: 2 or a variant having at least 75% homology
30 (preferably at least 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 2 over its entire sequence based on nucleotide identity;
- (ii) the sequence shown in SEQ ID NO: 4 or a variant having at least 85% homology (preferably at least 90% 95%, 97% or 99% homology) to SEQ ID NO: 4 over its entire sequence based on nucleotide identity;

- (iii) the sequence shown in SEQ ID NO: 6 or a variant having at least 71% homology (preferably at least 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 6 over its entire sequence based on nucleotide identity;
- (iv) the sequence shown in SEQ ID NO: 8 or a variant having at least 72% homology (preferably at least 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 8 over its entire sequence based on nucleotide identity;
- (v) the sequence shown in SEQ ID NO: 10 or a variant having at least 76% homology (preferably at least 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 10 over its entire sequence based on nucleotide identity;
- (vi) the sequence shown in SEQ ID NO: 12 or a variant having at least 75% homology (preferably at least 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 12 over its entire sequence based on nucleotide identity;
- (vii) the sequence shown in SEQ ID NO: 14 or a variant having at least 70% homology (preferably at least 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 14 over its entire sequence based on nucleotide identity;
- (viii) the sequence shown in SEQ ID NO: 16 or a variant having at least 71% homology (preferably at least 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 16 over its entire sequence based on nucleotide identity;
- (ix) the sequence shown in SEQ ID NO: 18 or a variant having at least 76% homology (preferably at least 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 18 over its entire sequence based on nucleotide identity;
- (x) the sequence shown in SEQ ID NO: 20 or a variant having at least 77% homology (preferably at least 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 20 over its entire sequence based on nucleotide identity;
- (xi) the sequence shown in SEQ ID NO: 22 or a variant having at least 79% homology (preferably at least 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 22 over its entire sequence based on nucleotide identity;
- (xii) the sequence shown in SEQ ID NO: 24 or a variant having at least 66% homology (preferably at least 70%, 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 24 over its entire sequence based on nucleotide identity;
- (xiii) the sequence shown in SEQ ID NO: 26 or a variant having at least 72% homology (preferably at least 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 26 over its entire sequence based on nucleotide identity; or

- (xiv) the sequence shown in SEQ ID NO: 28 or a variant having at least 68% homology (preferably at least 70%, 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 28 over its entire sequence based on nucleotide identity.

The polynucleotide sequence may consist of the relevant SEQ ID NO: or variant thereof.

5 A variant of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 or 28 encodes a polypeptide which retains the enzymatic activity of the corresponding wild-type polypeptide. A variant preferably encodes a polypeptide which retains the enzymatic activity of the corresponding wild-type polypeptide, but has a substrate specificity that differs from the corresponding wild-type polypeptide. For instance, a variant of SEQ ID NO: 8 preferably encodes a polypeptide which
10 retains the enzymatic activity of SEQ ID NO: 9 (i.e. glycosyltransferase activity), but has a substrate specificity that differs from SEQ ID NO: 9 (e.g. has a different carbohydrate specificity). For those enzymes that catalyse a reaction between two or more substrates, it is preferred that the specificity for only one of the enzyme's substrates is altered. Preferred ways in which the substrate specificity is altered is discussed below with reference to the polypeptides of the invention. In preferred
15 embodiments:

- the variant of SEQ ID NO: 8 encodes a polypeptide having a carbohydrate specificity that differs from that of SEQ ID NO: 9;
- the variant of SEQ ID NO: 6 encodes a polypeptide having a fatty acid specificity that differs from that of SEQ ID NO: 7;
- 20 - the variant of SEQ ID NO: 22 encodes a polypeptide having a fatty acid specificity that differs from that of SEQ ID NO: 23;
- the variant of SEQ ID NO: 24 encodes a polypeptide having a phospholipid specificity that differs from that of SEQ ID NO: 25.

A variant must also retain its ability to be expressed in a host cell as described below. The
25 enzymatic activity and/or substrate specificity of a polypeptide can be assayed as described below with reference to the polypeptides of the invention.

The polynucleotide sequence may be a naturally occurring variant which is expressed by an organism, for instance by a bacterium different from *Streptomyces chartreusis*. Variants also include non-naturally occurring variants produced by recombinant technology. Over the entire length of the
30 polynucleotide sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 or 28, a variant has a specific percentage homology to that sequence based on nucleotide identity as discussed above. Homology may be determined as discussed above with reference to SEQ ID NO: 1.

Any number of nucleotide substitutions may be made to the sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 or 28, for example up to 10, 50, 30 or 100 substitutions. Codons

within the sequence may be replaced with codons that encode the same amino acid (i.e. degenerate codons). Codons may be replaced such that conservative substitutions are introduced in the expressed polypeptide sequences as discussed above with reference to SEQ ID NO: 1.

One or more nucleotides may additionally be deleted from of SEQ ID NO: 2, 4, 6, 8, 10, 12,
5 14, 16, 18, 20, 22, 24, 26 or 28. Up to 1, 2, 3, 4, 5, 10, 20 or 30 nucleotides may be deleted, or more.

Variants may be fragments of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 or 28. Such fragments retain the enzymatic ability discussed above.

One or more nucleotides may be alternatively or additionally added to the polynucleotides described above. An extension may be provided at the 5' or 3' of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14,
10 16, 18, 20, 22, 24, 26 or 28. The extension may be quite short, for example from 3 to 30 nucleotides in length. Alternatively, the extension may be longer, for example up to 150 or 300 nucleotides. The extension may be a control sequence at the 5' end.

The invention also provides a polynucleotide construct comprising more than one, such as 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or all, of the polynucleotide sequences defined in (i) to (xiv) above.
15 A construct of the invention preferably comprises the polynucleotide sequences necessary to encode any of the heterologous expression systems of the invention discussed below. The construct preferably comprises the twelve polynucleotide sequences as defined in (i) to (viii) and (xi) to (xiv) above. The construct more preferably comprises all fourteen of the polynucleotide sequences as defined in (i) to (xiv) above. Both constructs can be used to confer to a host cell the ability to
20 produce tunicamycin or a derivative thereof. The latter (more preferred construct) provides the transporters necessary to transport the tunicamycin or derivative thereof out of the host cell (tunI and tunJ).

The construct more preferably comprises at least five polynucleotide sequences as defined in (i) to (xiv) above. The construct most preferably comprises five polynucleotide sequences as defined
25 in (iii), (iv), (v), (xi) and (xii) above. This most preferred embodiment confers upon a host cell the ability to produce tunicamycin or a derivative thereof from tunicaminyI-uracil.

The construct can include one or more, such as 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14, of the variants defined in (i) to (xiv) above. The inclusion of one or more variants, particularly those that express polypeptides with an altered substrate specificity, allows the production of tunimycin
30 derivatives as discussed below. The variants may be any of those defined above. The constructs of the invention preferably comprise one of more of the preferred variants described above, i.e. those that express polypeptides having a carbohydrate specificity or fatty acid specificity that differs from that of the corresponding wild-type polypeptide. This confers upon a host cell the ability to produce

tunicamycin derivatives having altered carbohydrate and/or fatty acid groups. Again, this is discussed in more detail below with reference to the heterologous expression systems of the invention.

Polynucleotide sequences and constructs may be isolated and replicated using standard methods in the art. Genomic DNA may be extracted from an organism, such as *Streptomyces chartreusis*. The relevant sequence(s) may be amplified using PCR involving specific primers. The amplified sequence(s) may then be incorporated into a recombinant replicable vector such as a cloning vector. The vector may be used to replicate the polynucleotide sequence or construct in a compatible host cell. Thus polynucleotide sequences or constructs may be made by introducing a gene into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells for cloning of polynucleotides are known in the art and described in more detail below.

Polynucleotide sequences extracted from an organism, such as *Streptomyces chartreuses*, can of course be modified to form any of the variants described above. Methods for doing this are well-known in the art.

The polynucleotide sequence or construct may be cloned into suitable expression vector. In an expression vector, the polynucleotide sequence or construct is typically operably linked to at least one control sequence which is capable of providing for the expression of the polynucleotide sequence or construct by the host cell. Multiple control sequences can be used to express a polynucleotide construct of the invention. Such expression vectors can be used to express one or more of the polypeptides of the invention.

The term “operably linked” refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence “operably linked” to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. Multiple copies of the same or different polynucleotide sequence or construct may be introduced into the vector.

The expression vector may then be introduced into a suitable host cell. Thus, one or more of the polypeptides of the invention can be produced by inserting a polynucleotide sequence or a construct of the invention into an expression vector, introducing the vector into a compatible bacterial host cell, and growing the host cell under conditions which bring about expression of the polynucleotide sequence or construct.

The vectors may be for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide sequence or construct and optionally a regulator of the promoter. The vectors may contain one or more selectable marker

genes, for example an ampicillin resistance gene. Promoters and other expression regulation signals may be selected to be compatible with the host cell for which the expression vector is designed. A T7, *trc*, *lac*, *ara* or λ_L promoter is typically used.

The host cell typically expresses the one or more polypeptides at a high level. The host cell preferably expresses the one or more polypeptides to a greater degree than *Streptomyces chartreuses*. Host cells transformed with a polynucleotide sequence or a construct will be chosen to be compatible with the expression vector used to transform the cell. The host cell is typically bacterial and preferably *Streptomyces lividans*, *Streptomyces coelicolor*, *Streptomyces chartreusis*, *Streptomyces lysosuperificus* or *Escherichia coli*. Any cell with a λ DE3 lysogen, for example C41 (DE3), BL21 (DE3), JM109 (DE3), B834 (DE3), TUNER, Origami and Origami B, can express a vector comprising the T7 promoter.

Alternatively, the polynucleotide sequences or constructs of the invention may be isolated, substantially isolated, purified or substantially purified. A polynucleotide sequence or construct is isolated or purified if it is completely free of any other components, such as lipids or other polynucleotides. A polynucleotide sequence or construct is substantially isolated if it is mixed with carriers or diluents which will not interfere with its intended use. For instance, a polynucleotide sequence or construct is substantially isolated or substantially purified if it present in a form that comprises less than 10%, less than 5%, less than 2% or less than 1% of other components, such as lipids or other polynucleotides.

Polypeptides

The invention provides a polypeptide encoded by any of the polynucleotides of the invention. Polypeptides may be expressed from the polynucleotides of the invention as discussed above.

The invention also provides a polypeptide comprising the sequence of one of the 14 tun enzymes encoded by SEQ ID NO: 1 or a variant thereof. In particular, the invention provides:

- (a) the sequence shown in SEQ ID NO: 3 or a variant having at least 73% homology (preferably at least 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 3 over its entire sequence based on amino acid identity;
- (b) the sequence shown in SEQ ID NO: 5 or a variant having at least 91% homology (preferably at least 95%, 97% or 99% homology) to SEQ ID NO: 5 over its entire sequence based on amino acid identity;

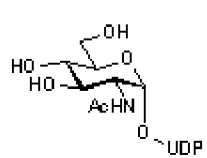
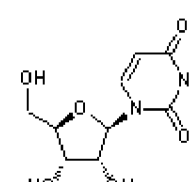
- (c) the sequence shown in SEQ ID NO: 7 or a variant having at least 61% homology (preferably at least 65%, 70%, 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 7 over its entire sequence based on amino acid identity;
- (d) the sequence shown in SEQ ID NO: 9 or a variant having at least 64% homology (preferably at least 65%, 70%, 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 9 over its entire sequence based on amino acid identity;
- (e) the sequence shown in SEQ ID NO: 11 or a variant having at least 78% homology (preferably at least 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 11 over its entire sequence based on amino acid identity;
- (f) the sequence shown in SEQ ID NO: 13 or a variant having at least 77% homology (preferably at least 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 13 over its entire sequence based on amino acid identity;
- (g) the sequence shown in SEQ ID NO: 15 or a variant having at least 66% homology (preferably at least 70%, 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 15 over its entire sequence based on amino acid identity;
- (h) the sequence shown in SEQ ID NO: 17 or a variant having at least 67% homology (preferably at least 70%, 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 17 over its entire sequence based on amino acid identity;
- (i) the sequence shown in SEQ ID NO: 19 or a variant having at least 78% homology (preferably at least 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 19 over its entire sequence based on amino acid identity;
- (j) the sequence shown in SEQ ID NO: 21 or a variant having at least 77% homology (preferably at least 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 21 over its entire sequence based on amino acid identity;
- (k) the sequence shown in SEQ ID NO: 23 or a variant having at least 66% homology (preferably at least 70%, 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 23 over its entire sequence based on amino acid identity;
- (l) the sequence shown in SEQ ID NO: 25 or a variant having at least 53% homology (preferably at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 97% or 99% homology) to SEQ ID NO: 25 over its entire sequence based on amino acid identity;
- (m) the sequence shown in SEQ ID NO: 27 or a variant having at least 55% homology (preferably at least 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 97% or

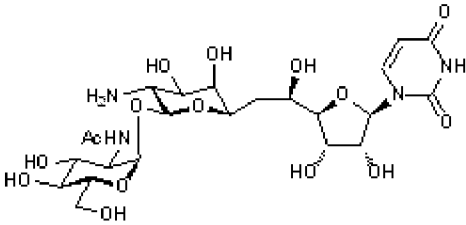
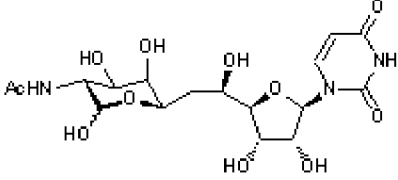
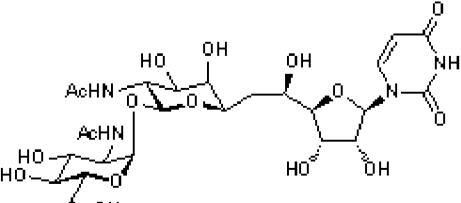
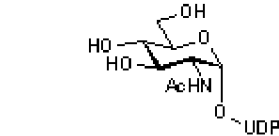
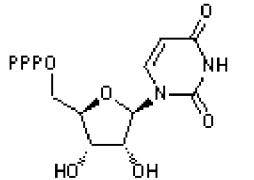
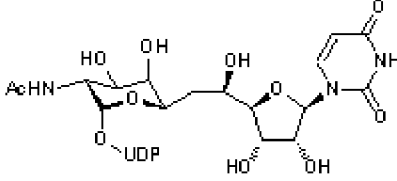
99% homology) to SEQ ID NO: 27 over its entire sequence based on amino acid identity; or

- (n) the sequence shown in SEQ ID NO: 29 or a variant having at least 69% homology (preferably at least 70%, 75%, 80%, 85%, 90%, 95%, 97% or 99% homology) to SEQ ID NO: 29 over its entire sequence based on amino acid identity.

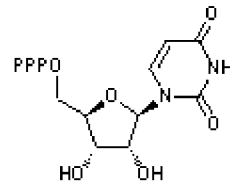
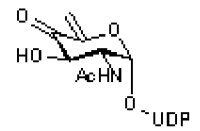
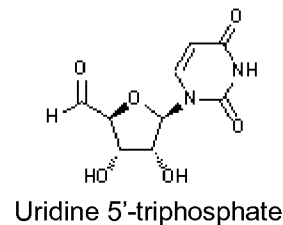
A variant of SEQ ID NO: 2, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or 29 retains the enzymatic activity of the wild-type polypeptide. A variant preferably retains the enzymatic activity of the corresponding wild-type polypeptide, but has a substrate specificity that differs from the corresponding wild-type polypeptide. For instance, a variant of SEQ ID NO: 9 preferably retains the enzymatic activity of SEQ ID NO: 9 (i.e. glycosyltransferase activity), but has a substrate specificity that differs from SEQ ID NO: 9 (e.g. has a different carbohydrate specificity). A “different” substrate specificity means that the polypeptide specifically catalyzes a reaction involving a different substrate or different substrates compared with the corresponding wild-type enzyme. Table 3 below summarizes the enzymatic activity and substrates of the 14 wild-type enzymes. A variant may catalyze a reaction involving a substrate or substrates that differ from those in columns A and B. For those enzymes that catalyse a reaction between two or more substrates, it is preferred that the specificity for only one of the enzyme’s substrates is altered (i.e. column A or B).

Table 3 – Tun enzymes activity and substrates

SEQ ID NO:	Gene	Amino acids	Enzymatic activity	A	B
3	tunA	321	epimerase/dehydratase	UDP-N-acetylglucosamine (UDP-GlcNAc)	-
					
5	tunB	338	Fe-S oxidoreductase	Uridine	-
					

7	tunC	318	<i>N</i> -acyltransferase		Fatty acid
9	tunD	474	Glycosyltransferase		UDP-GlcNAc (a nucleotide carbohydrate)
11	tunE	234	<i>N</i> -deacetylase		-
13	tunF	327	Epimerase		-
15	tunG	203	Phosphatase		-
17	tunH	515	Nucleotide pyrophosphatase		-
19	tunI	304	ABC transport ATP-binding subunit	Tunicamycin (see Figure 4)	-
21	tunJ	262	ABC transport permease subunit	Tunicamycin (see Figure 4)	-
23	tunK	81	Acyl carrier protein	Fatty acid	-
25	tunL	229	Phospholipid phosphatase	Phospholipid	-
27	tunM	216	Methyltransferase	Uridine 5'-aldehyde	UDP-4-keto-5,6-ene-GlcNAc (a nucleotide carbohydrate)

29 tunN 152

UDP-ribose
pyrophosphatase

For those enzymes that use carbohydrates as a substrate, the variants preferably have a different carbohydrate specificity from the wild-type polypeptide. In particular, for those enzymes that use UDP-N-acetylglucosamine (UDP-GlcNAc) or UDP-4-keto-5,6-ene-GlcNAc as substrates (tunA, SEQ ID NO: 3; tunD, SEQ ID NO: 9; and tunM, SEQ ID NO: 27), the variant preferably uses any of the following carbohydrates as a substrate: a hexose, a 6-deoxyhexose, a hexosamine a pentose, a hexuronic acid, a glucuronic acid, a monosaccharide and an oligosaccharide. The hexose is preferably glucose, galactose, mannose, allose, altrose, gulose, idose, talose, psicose, fructose, sorbose or tagatose. The 6-deoxyhexose is preferably fucose or rhamnose. The hexosamine is preferably glucosamine, N-acetylglucosamine, galactosamine, N-acetylgalactosamine, mannosamine, N-acetylmannosamine or N-acetylquinovosamine. The pentose is preferably arabinose, lyxose, ribose or xylose. The hexuronic acid is preferably glucuronic acid, iduronic acid or galacturonic acid. The monosaccharide is preferably sialic acid, neuraminic acid, or another hexose derivative present in natural products. The oligosaccharide is preferably a linear or branched chain of

The tunicamycins are based on a tunicaminy-uracil scaffold (see Figure 4). Part of this scaffold is derived from UPP-GlcNAc as a result of the action of tunA and tunM. Variants of SEQ ID NO: 3 (tunA) or SEQ ID NO: 27 (tunM) that have a different carbohydrate specificity will produce derivatives of tunicmycin that are based on a scaffold containing a different carbohydrate group.

As can be seen from Figure 4, an N-acetylglucosamine group is added to the tunicamycin-uracil scaffold by the action of tunD (SEQ ID NO: 9). In a preferred embodiment, the variant of SEQ ID NO: 9 has a carbohydrate specificity that differs from that of SEQ ID NO: 9. The variant's carbohydrate specificity may differ in any of the ways disclosed above. Variants of SEQ ID NO: 9 having a different carbohydrate specificity will add a different carbohydrate group to the tunicaminy-uracil scaffold.

A fatty acid is added to the tunicamycin-uracil scaffold by the action of tunC (SEQ ID NO: 7). These fatty acids are sequestered from phospholipids by tunL (SEQ ID NO: 25) and activated by tunK (SEQ ID NO: 23). In a preferred embodiment, the variant of SEQ ID NO: 7 has a fatty acid specificity that differs from that of SEQ ID NO: 7. In another preferred embodiment, the variant of SEQ ID NO: 23 has a fatty acid specificity that differs from that of SEQ ID NO: 23. The variants of SEQ ID NO: 7 or 23 more preferably use any of the following fatty acids as a substrate: a cis-unsaturated fatty acid, a trans-unsaturated fatty acid, a saturated fatty acid, a polyunsaturated fatty acid, a mycolic acid, an isoprenoid fatty acid, a branched fatty acid, a cyclic fatty acid, a hydroxy fatty acid, an epoxy fatty acid, a furanoid fatty acid and a glycolipid.

In another preferred embodiment, the variant of SEQ ID NO: 25 has a phospholipid specificity that differs from that of SEQ ID NO: 25. The variant of SEQ ID NO: 25 more preferably uses as a substrate a phospholipid containing any of the specific fatty acids listed above. Preferably the variants of SEQ ID NOs: 7, 23 and 29 are all specific for the same fatty acid and phospholipids containing that fatty acid. Systems in which the variants of SEQ ID NOs: 7, 23 and 29 have fatty acid and phospholipid specificities that differ from the wild-type polypeptides will add a different fatty acid group to the tunicaminy-uracil scaffold.

Other groups can be added to the tunicaminy-uracil scaffold by variants of tun D and/or tunC which specifically use those groups as substrates. In a preferred embodiment, the variant of SEQ ID NO: 9 specifically uses any of the following as a substrate instead of Glc-NAc: an amino acid, glycerol and glycerol 3'-phosphate. In a preferred embodiment, the variant of SEQ ID NO: 7 specifically uses any of the following as a substrate instead of a fatty acid: an amino acid, glycerol and glycerol 3'-phosphate. The amino acid may be any of the naturally-occurring amino acids, but is preferably serine or threonine.

A variant may include modifications that facilitate its handling or expression in a particular host cell.

The enzymatic activity of a variant can be assayed using any method known in the art. For instance, the ability of a variant to catalyze a reaction can be assayed by expressing the variant in a host cell and determining whether or not the host cell is capable of catalyzing the reaction. Substrate specificity can also be tested in this way. A host cell expressing the variant is contacted with specific molecules to determine which, if any, it is capable of using as a substrate. The substrate specificity of a variant can be compared with that of a wild-type polypeptide by comparing the substrate specificity of host cells expressing the variant and wild-type polypeptides respectively.

The polypeptide may be a naturally occurring variant which is expressed by an organism, for instance by a bacterium different from *Streptomyces chartreusis*. Variants also include non-naturally

occurring variants produced by recombinant technology. Over the entire length of the amino acid sequence of SEQ ID NO: 2, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or 29, a variant will have a specific homology to that sequence based on amino acid identity as described above. Preferred levels of homology based on amino acid identity are also described above. There may be at least 80%, for example at least 85%, 90% or 95%, amino acid identity over a stretch of 40 or more, for example 50, 100, 150, 200, 250, 270 or 280 or more, contiguous amino acids ("hard homology"). Methods for determining homology are described above.

Amino acid substitutions may be made to the amino acid sequence of SEQ ID NO: 2, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or 29 in addition to those discussed above, for example up to 1, 2, 3, 4, 5, 10, 20 or 30 substitutions. Conservative substitutions may be made, for example, according to Tables 1 and 2 above. Modifications may be made to the active site of the variants to alter its substrate specificity.

One or more amino acid residues of the amino acid sequence of SEQ ID NO: 2, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or 29 may additionally be deleted from the polypeptide. Up to 1, 2, 3, 4, 5, 10, 20 or 30 residues may be deleted, or more.

Variants may be fragments of SEQ ID NO: 2, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or 29. Such fragments retain enzyme activity. Fragments may be at least 50, 100, 200 or 250 amino acids in length. A fragment preferably comprises the active site of SEQ ID NO: 2, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or 29.

One or more amino acids may be alternatively or additionally added to the polypeptides described above. An extension may be provided at the amino terminus or carboxy terminus of the amino acid sequence of SEQ ID NO: 2, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or 29 or a variant or fragment thereof. The extension may be quite short, for example from 1 to 10 amino acids in length. Alternatively, the extension may be longer, for example up to 50 or 100 amino acids.

The variant may be modified for example by the addition of histidine or aspartic acid residues to assist its identification or purification or by the addition of a signal sequence to promote their secretion from a cell where the polypeptide does not naturally contain such a sequence.

The polypeptides of the invention may be labelled with a revealing label. The revealing label may be any suitable label which allows the pore to be detected. Suitable labels include, but are not limited to, fluorescent molecules, radioisotopes, e.g. ^{125}I , ^{35}S , ^{14}C , enzymes, antibodies, antigens, polynucleotides and ligands such as biotin.

The polypeptides of the invention may be isolated from *Streptomyces chartensis*, or made synthetically or by recombinant means. For example, the polypeptide may be synthesised by *in vitro* translation and transcription. The amino acid sequence of the polypeptide may be modified to

include non-naturally occurring amino acids or to increase the stability of the protein. When the polypeptide is produced by synthetic means, such amino acids may be introduced during production. The polypeptide may also be altered following either synthetic or recombinant production.

The polypeptide may also be produced using D-amino acids. For instance, the polypeptide may comprise a mixture of L-amino acids and D-amino acids. This is conventional in the art for producing such proteins or peptides.

The polypeptide may also contain other non-specific chemical modifications as long as they do not interfere with its enzymatic activity. A number of non-specific side chain modifications are known in the art and may be made to the polypeptides. Such modifications include, for example, reductive alkylation of amino acids by reaction with an aldehyde followed by reduction with NaBH₄, amidination with methylacetimidate or acylation with acetic anhydride.

The polypeptides of the invention can be produced using standard methods known in the art. Polynucleotide sequences encoding the polypeptide may be isolated and replicated using standard methods in the art. Such sequences are discussed in more detail above. Polynucleotide sequences encoding a polypeptide of the invention may be expressed in a bacterial host cell using standard techniques in the art. The polypeptide may be produced in a cell by *in situ* expression of the polypeptide from a recombinant expression vector. The expression vector optionally carries an inducible promoter to control the expression of the polypeptide.

A polypeptide may be produced in large scale following purification by any protein liquid chromatography system from organisms that naturally express the protein or after recombinant expression as described below. Typical protein liquid chromatography systems include FPLC, AKTA systems, the Bio-Cad system, the Bio-Rad BioLogic system and the Gilson HPLC system.

The polypeptides of the invention may be isolated, substantially isolated, purified or substantially purified. A polypeptide is isolated or purified if it is completely free of any other components, such as lipids or other polypeptides. A polypeptide is substantially isolated if it is mixed with carriers or diluents which will not interfere with its intended use. For instance, a polypeptide is substantially isolated or substantially purified if it present in a form that comprises less than 10%, less than 5%, less than 2% or less than 1% of other components, such as lipids or other polypeptides.

Heterologous expression systems

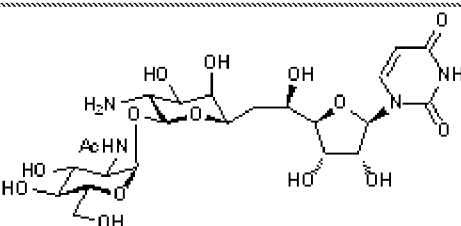
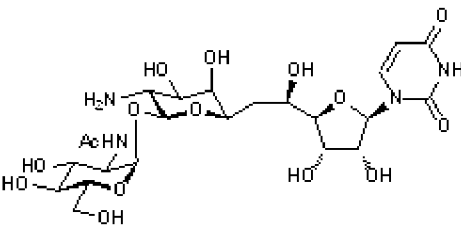
The invention also provides a heretologous expression system encoded by a polynucleotide sequence comprising the sequence shown in nucleotides 8616 to 20606 of SEQ ID NO: 1 or a variant having at least 50% homology to nucleotides 8616 to 20606 of SEQ ID NO: 1 over its entire

sequence based on nucleotide identity. A “heterologous expression system” is a group of heterologous enzymes present in a host cell that is capable of producing a tunicamycin or a derivative thereof. Enzymes are heterologous if they are not native to the host cell. Host cells can be transformed with a polynucleotide of the invention or a polynucleotide construct of the invention as discussed above.

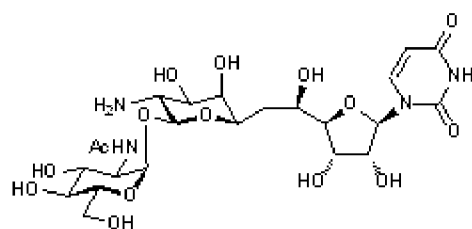
The invention also provides a heterologous expression system comprising more than one, such as 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or all, of the polypeptides defined in (a) to (n) above. The heterologous expression system preferably comprises the twelve polynucleotides sequences as defined in (a) to (h) and (k) to (n) above. The heterologous expression system more preferably comprises all fourteen of the polypeptides as defined in (i) to (xiv) above, i.e. the system comprises a polypeptide defined in (a), a polypeptide defined in (b), a polypeptide defined in (c), a polypeptide defined in (d), a polypeptide defined in (e), a polypeptide defined in (f), a polypeptide defined in (g), a polypeptide defined in (h), a polypeptide defined in (i), a polypeptide defined in (j), a polypeptide defined in (k), a polypeptide defined in (l), a polypeptide defined in (m) and a polypeptide defined in (n). Both systems can be used to to produce tunicamycin or a derivative thereof. The latter (more preferred system) comprises the transporters necessary to transport the tunicamycin or derivative thereof out of the host cell (tunI and tunJ).

Table 4 below summarizes some of the different heterologous systems envisaged by the invention and the starting molecule(s) that may be used to produce a tunicamycin or a derivative thereof.

Table 4 – Heterologous expression systems of the invention

No.	Polypeptide(s) defined in	Starting molecule(s)	Comments
1	(c) (k)		-
2	(c) (l)		The acids produced by (l) will need to be activated

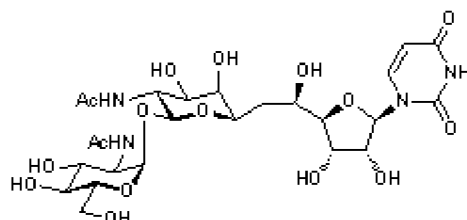
3 (c)
(l)
(k)



-

4 (c)
(l)
(k)
(e)

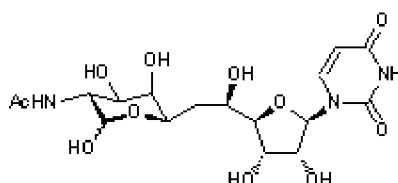
Tunicaminyl-uracil



-

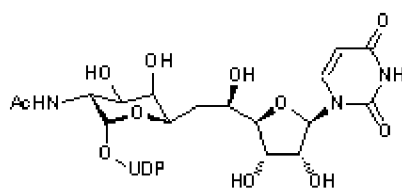
5 (c)
(l)
(k)
(e)
(d)

Tunicaminyl-uracil



-

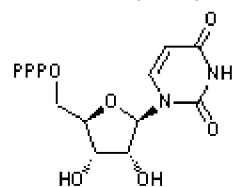
6 (c)
(l)
(k)
(e)
(d)
(h)



-

7 (c)
(l)
(k)
(e)
(d)
(h)

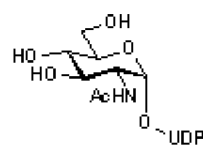
Uridine 5'-triphosphate



-

8 (c)
(l)
(k)
(e)
(d)
(h)
(f)

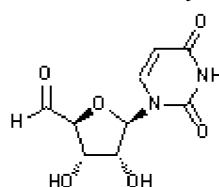
UDP-GlcNAc



-

9 (c)
(l)
(k)
(e)
(d)
(h)
(f)
(m)

Uridine 5'-aldehyde



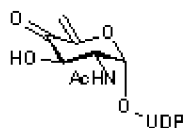
-

and

UDP-4-keto-5,6-ene-GlcNAc (a nucleotide)

26

carbohydrate)

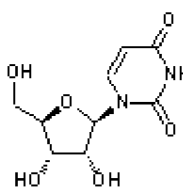


10

(c)
(l)
(k)
(e)
(d)
(h)
(f)
(m)
(b)

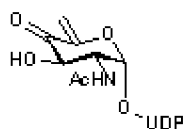
Uridine

-



and

UDP-4-keto-5,6-ene-GlcNAc (a nucleotide
carbohydrate)

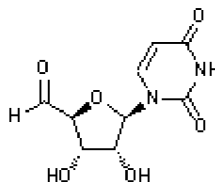


11

(c)
(l)
(k)
(e)
(d)
(h)
(f)
(m)
(a)

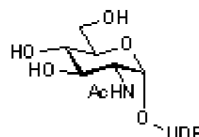
Uridine 5'-aldehyde

-



and

UDP-N-acetylglucosamine (UDP-GlcNAc)

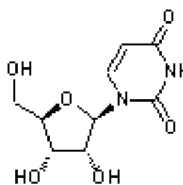


12

(c)
(l)
(k)
(e)
(d)
(h)
(f)
(m)
(a)
(b)

Uridine

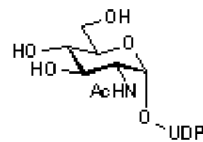
-



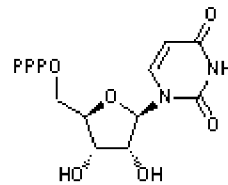
and

UDP-N-acetylglucosamine (UDP-GlcNAc)

27



Uridine 5'-triphosphate



13 (c)
(l)
(k)
(e)
(d)
(h)
(f)
(m)
(a)
(b)
(n)
(g)

Any of heterologous expression systems 1 to 13 may further comprise the polypeptides defined in (i) and (j). These embodiments are herein referred to as heterologous expression systems 14 to 26. The polypeptides defined in (i) and (j) transport the tunicamycin or derivative thereof out of the host cell.

The heterologous expression system preferably comprises at least five polypeptides as defined in (a) to (n) above. The heterologous expression system most preferably comprises five polypeptides as defined in (c), (d), (e), (k) and (l) above (i.e. heterologous expression systems 5 to 13 and 18 to 26). This most preferred embodiment allows a host cell to produce tunicamycin or a derivative thereof from tunicaminy-uracil.

The heterologous expression system can comprise one or more, such as 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14, of the variants defined in (a) to (n) above. As discussed above, the inclusion of one or more variants, particularly those that express polypeptides with an altered substrate specificity, allows the production of tunimycin derivatives. The variants may be any of those defined above. All of heterologous expression systems 1 to 26 preferably comprise a variant of SEQ ID NO: 7 that has an altered substrate specificity, most preferably an altered fatty acid specificity. All of heterologous expression systems 1 to 26 preferably comprise a variant of SEQ ID NO: 25 that has an altered substrate specificity, most preferably an altered phospholipids specificity. The variant of SEQ ID NO: 25 most preferably specifically reacts with phospholipids that contain the fatty acid used as a substrate by the variant of SEQ ID NO: 7.

Heterologous expression systems 2 to 13 and 15 to 26 preferably comprise a variant of SEQ ID NO: 23 that has an altered substrate specificity, most preferably an altered fatty acid specificity. In the most preferred embodiment of heterologous expression systems 2 to 13 and 15 to 26, the variant of SEQ ID NO: 25 specifically reacts with phospholipids that contain the fatty acid used as a substrate by the variants of SEQ ID NOs: 7 and 23. These embodiments allow the production

tunicamycin derivatives in which the fatty acid group (attached to the scaffold) is replaced by another group, preferably a different fatty acid.

Heterologous expression systems 5 to 13 and 18 to 26 preferably comprise a variant of SEQ ID NO: 9 that has an altered substrate specificity, most preferably an altered carbohydrate specificity.

5 These embodiments allow the production tunicamycin derivatives in which the N-acetylglucosamine group (attached to the scaffold) is replaced by another group, preferably a different carbohydrate or sugar group.

Heterologous expression systems 9 to 13 and 22 to 26 preferably comprise a variant of SEQ ID NO: 27 that has an altered substrate specificity, most preferably an altered carbohydrate
10 specificity. These embodiments allow the production tunicamycin derivatives in which the sugar group in the scaffold is replaced by different groups, preferably a different sugar or carbohydrate.

Heterologous expression systems 9 to 13 and 22 to 26 preferably comprise (a) a variant of SEQ ID NO: 7 that has an altered substrate specificity, most preferably an altered fatty acid specificity, (b) a variant of SEQ ID NO: 25 that has an altered substrate specificity, most preferably
15 an altered phospholipids specificity, (c) a variant of SEQ ID NO: 23 that has an altered substrate specificity, most preferably an altered fatty acid specificity, and (d) a variant of SEQ ID NO: 9 that has an altered substrate specificity, most preferably an altered carbohydrate specificity. This embodiment allows the production of tunicamycin derivatives in which the N-acetylglucosamine and fatty acid groups (attached to the scaffold) are replaced by different groups, preferably a different
20 carbohydrate or sugar group and a different fatty acid group.

Heterologous expression systems 9 to 13 and 22 to 26 preferably comprise (a) a variant of SEQ ID NO: 7 that has an altered substrate specificity, most preferably an altered fatty acid specificity, (b) a variant of SEQ ID NO: 25 that has an altered substrate specificity, most preferably
25 an altered phospholipids specificity, (c) a variant of SEQ ID NO: 23 that has an altered substrate specificity, most preferably an altered fatty acid specificity, (d) a variant of SEQ ID NO: 9 that has an altered substrate specificity, most preferably an altered carbohydrate specificity, and (e) a variant of SEQ ID NO: 27 that has an altered substrate specificity, most preferably an altered carbohydrate specificity. This embodiment allows the production of tunicamycin derivatives in which (1) the N-acetylglucosamine and fatty acid groups (attached to the scaffold) are replaced by different groups,
30 preferably a different carbohydrate or sugar group and a different fatty acid group, and (2) the sugar group in the scaffold is replaced by different groups, preferably a different sugar or carbohydrate.

Methods of producing tunicamycin or derivatives thereof

The invention further provides a method of producing a tunicamycin or a derivative thereof. The method comprising culturing a host cell comprising a polynucleotide of the invention, a polynucleotide construct of the invention, a vector of the invention or a heterologous expression system of the invention and isolating the tunicamycin or derivative thereof. The derivative can be designed as discussed above. The method may produce a derivative in which (1) the N-acetylglucosamine (attached to the tunicamycin scaffold) is replaced by a different group, preferably a different carbohydrate or sugar group, (2) the fatty acid group (attached to the tunicamycin scaffold) is replaced by a different group, preferably a different fatty acid group, or (3) the sugar group in the tunicamycin scaffold is replaced by different groups, preferably a different sugar or carbohydrate. The method may also produce derivatives in which (1) and (2), (2) and (3), (1) and (3) and (1), (2) and (3).

The method preferably produces derivatives that are highly selective for MraY. This means that the derivatives do not inhibit any other enzymes to any measureable or significant degree.

The method also preferably produces derivatives that do not bind to the active site of UDP-GlcNAc:dolichyl phosphate GlcNAc-1-phosphate transferase (GPT). These embodiments allow the derivatives to be used as antibiotics in humans.

Tunicamycins, derivatives and medical uses

The invention also provides a tunicamycin or a derivative thereof produced using a method of the invention. Any of the methods described above may be used.

The invention also provides a tunicamycin or a derivative thereof of the invention for use in a method of treatment of the human or animal body.

The invention further provides a tunicamycin or a derivative thereof of the invention for use in a method of treating or preventing a bacterial infection in a subject. The invention also provides use of a tunicamycin or a derivative thereof in the manufacture of a medicament for treating or preventing a bacterial infection in a subject. The invention also provides a method of treating or preventing a bacterial infection in a subject comprising administering to said subject a pharmaceutically acceptable amount of a tunicamycin or a derivative thereof of the invention.

Typically, the subject is human. However, it may be non-human. Preferred non-human animals include, but are not limited to, primates, such as marmosets or monkeys, commercially farmed animals, such as horses, cows, sheeps or pigs, and pets, such as dogs, cats, mice, rats, guinea pigs, ferrets, gerbils or hamsters. The subject can be any animal that is capable of being infected by a bacterium.

The bacterium causing the infection may be any bacterium expressing MraY.

The bacterium may, for instance, be any bacterium that has a peptidoglycan component in the cell wall. The bacterium may be Gram-positive or Gram-negative. In a preferred instance the bacterium is Gram-positive. The bacterium may in particular be a pathogenic bacterium.

In one preferred instance, the bacterium may be one selected from a bacterium of the following Gram-positive bacteria families: *Streptococcus*, *Staphylococcus* (including MRSA), *Corynebacterium*, *Listeria*, *Bacillus* (including *Enterococcus*) and *Clostridium*. Examples of Gram-positive bacteria include *Clostridium botulinum*, *Clostridium difficile*, *Clostridium perfringens*, *Clostridium tetani*, *Corynebacterium diphtheriae*, *Enterococcus faecalis*, *Enterococcus faecium*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Staphylococcus saprophyticus*, *Streptococcus agalactiae*, *Streptococcus pneumoniae* and *Streptococcus pyogenes*.

In another instance, the bacterium may be a Gram-negative bacteria, with preferred examples including: *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Moraxella catarrhalis*, *Hemophilus influenzae*, *Klebsiella pneumoniae*, *Legionella pneumophila*, *Pseudomonas aeruginosa*, *Escherichia coli*, *Proteus mirabilis*, *Enterobacter cloacae*, *Serratia marcescens*, *Helicobacter pylori*, *Salmonella enteritidis*, *Salmonella typhi*, *Acinetobacter baumannii*, *Clostridium*, *Brucella*, *Shigella* and *Vibrio cholerae*. The Gram-negative bacterium may be, for instance, *Bordetella pertussis*, *Borrelia burgdorferi*, *Brucella abortus*, *Brucella canis*, *Brucella melitensis*, *Brucella suis*, *Campylobacter jejuni*, *Escherichia coli*, *Francisella tularensis*, *Haemophilus influenzae*, *Helicobacter pylori*, *Legionella pneumophila*, *Leptospira interrogans*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Pseudomonas aeruginosa*, *Rickettsia rickettsii*, *Salmonella typhi*, *Salmonella typhimurium*, *Shigella sonnei*, *Treponema pallidum*, *Vibrio cholerae* and *Yersinia pestis*.

In one preferred instance, the bacterium is a *Streptococcus* or *Pseudomonas*, particularly where the condition to be treated is pneumonia. In another instance the bacterium may be *Shigella*, *Campylobacter* or *Salmonella*, particularly where the condition to be treated is a food borne infection. The bacterium may be one responsible for tetanus, typhoid fever, diphtheria, syphilis or leprosy. For instance, the bacterium may be one selected from the group *Clostridium tetani*, *Salmonella typhi*, *Corynebacterium diphtheriae*, *Treponema pallidum* and *Mycobacterium leprae*. In another instance, the bacterium may be an opportunistic pathogen, and in a preferred instance may be selected from *Pseudomonas aeruginosa*, *Burkholderia cenocepacia*, and *Mycobacterium avium*. In a further instance, the bacterium may be Chlamydia, *Mycobacterium* or *Brucella*. In a further preferred instance, the bacterium is a *Mycobacterium* (including tuberculosis and leprae). In one instance the bacterium is *Mycobacterium tuberculosis*, particularly where tuberculosis is being treated. In further instances, the bacterium may be *Chlamydia pneumoniae*, *Chlamydia trachomatis*, *Chlamydophila psittaci*, *Mycobacterium leprae*, *Mycobacterium tuberculosis* or *Mycoplasma pneumoniae*.

In a further instance, the bacterium may be *Clostridium tetani*, *Clostridium difficile*, *Clostridium perfringens*, *Chlamydophila psittaci*, *Clostridium botulinum*, *Enterococcus faecalis*, *Enterococcus faecium*, *Helicobacter pylori*, *Legionella pneumophila*, *Leptospira interrogans*, *Listeria monocytogenes*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Mycoplasma pneumoniae*,
5 *Pseudomonas aeruginosa*, *Shigella sonnei*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Staphylococcus saprophyticus*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Treponema pallidum*, *Vibrio cholerae* or *Yersinia pestis*.

The invention may be used to treat infections and conditions caused by any of the above-mentioned bacteria.

10 The tunicamycin or derivative thereof can be administered to the subject in order to prevent the onset of one or more symptoms of the bacterial infection. This prophylaxis. In this embodiment, the subject can be asymptomatic. The subject is typically one that has been exposed to the bacterium. A prophylactically effective amount of the tunicamycin or derivative thereof is administered to such a subject. A prophylactically effective amount is an amount which prevents the
15 onset of one or more symptoms of the bacterial infection.

The tunicamycin or derivative thereof can be administered to the subject in order to treat one or more symptoms of the bacterial infection. In this embodiment, the subject is typically symptomatic. A therapeutically effective amount of the tunicamycin or derivative thereof is administered to such a subject. A therapeutically effective amount is an amount effective to
20 ameliorate one or more symptoms of the disorder.

The tunicamycin or derivative thereof can be administered to the subject by any suitable means. The compound can be administered by enteral or parenteral routes such as via oral, buccal, anal, pulmonary, intravenous, intra-arterial, intramuscular, intraperitoneal, intraarticular, topical or other appropriate administration routes.

25 The formulation of the tunicamycin or derivative thereof will depend upon factors such as the nature of the compound and the disorder to be treated. The tunicamycin or derivative thereof may be administered in a variety of dosage forms. It may be administered orally (e.g. as tablets, troches, lozenges, aqueous or oily suspensions, dispersible powders or granules), parenterally, subcutaneously, intravenously, intramuscularly, intrasternally, transdermally or by infusion
30 techniques. The compound may also be administered as a suppository. A physician will be able to determine the required route of administration for each particular subject.

Typically, the tunicamycin or derivative thereof is formulated for use with a pharmaceutically acceptable carrier or diluent and this may be carried out using routine methods in the pharmaceutical art. The pharmaceutical carrier or diluent may be, for example, an isotonic

solution. For example, solid oral forms may contain, together with the active compound, diluents, e.g. lactose, dextrose, saccharose, cellulose, corn starch or potato starch; lubricants, e.g. silica, talc, stearic acid, magnesium or calcium stearate, and/or polyethylene glycols; binding agents; e.g. starches, arabic gums, gelatin, methylcellulose, carboxymethylcellulose or polyvinyl pyrrolidone; 5 disaggregating agents, e.g. starch, alginic acid, alginates or sodium starch glycolate; effervescing mixtures; dyestuffs; sweeteners; wetting agents, such as lecithin, polysorbates, laurylsulphates; and, in general, non-toxic and pharmacologically inactive substances used in pharmaceutical formulations. Such pharmaceutical preparations may be manufactured in known manner, for example, by means of mixing, granulating, tableting, sugar-coating, or film coating processes.

10 Liquid dispersions for oral administration may be syrups, emulsions and suspensions. The syrups may contain as carriers, for example, saccharose or saccharose with glycerine and/or mannitol and/or sorbitol.

Suspensions and emulsions may contain as carrier, for example a natural gum, agar, sodium alginate, pectin, methylcellulose, carboxymethylcellulose, or polyvinyl alcohol. The suspensions or 15 solutions for intramuscular injections may contain, together with the tunicamycin or derivative thereof, a pharmaceutically acceptable carrier, e.g. sterile water, olive oil, ethyl oleate, glycols, e.g. propylene glycol, and if desired, a suitable amount of lidocaine hydrochloride.

Solutions for intravenous or infusions may contain as carrier, for example, sterile water or preferably they may be in the form of sterile, aqueous, isotonic saline solutions.

20 For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1% to 2%.

Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium 25 carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10% to 95% of active ingredient, preferably 25% to 70%. Where the pharmaceutical composition is lyophilised, the lyophilised material may be reconstituted prior to administration, e.g. a suspension. Reconstitution is preferably effected in buffer.

30 Capsules, tablets and pills for oral administration to a patient may be provided with an enteric coating comprising, for example, Eudragit "S", Eudragit "L", cellulose acetate, cellulose acetate phthalate or hydroxypropylmethyl cellulose.

Pharmaceutical compositions suitable for delivery by needleless injection, for example, transdermally, may also be used.

A therapeutically or prophylactically effective amount of the compound is administered. The dose may be determined according to various parameters, especially according to the compound used; the age, weight and condition of the subject to be treated; the route of administration; and the required regimen. Again, a physician will be able to determine the required route of administration and dosage for any particular subject. A typical daily dose is from about 0.1 to 50mg per kg, preferably from about 0.1mg/kg to 10mg/kg of body weight, according to the activity of the specific inhibitor, the age, weight and conditions of the subject to be treated, the type and severity of the disease and the frequency and route of administration. Preferably, daily dosage levels are from 5mg to 2g.

The invention provides a pharmaceutical composition comprising a tunicamycin or a derivative thereof of the invention and a pharmaceutically acceptable carrier. Such pharmaceutical compositions comprise a therapeutically or prophylactically effective amount of the tunicamycin or a derivative thereof and may further comprise instructions to enable the kit to be used in the method of the invention or details regarding which subjects the method may be used for.

The following Example illustrates the invention:

Example

1 Materials and Methods

1.1 Materials and DNA Manipulation Methods

DNA manipulations were performed according to standard procedures for *E. coli* (K. F. Chater, *Philos. Trans. R. Soc., B*, 2006, **361**, 761 and B. K. Leskiw, R. Mah, E. J. Lawlor and K. F. Chater, *J. Bacteriol.*, 1993, **175**, 1995) and *Streptomyces* (J. Sambrook and D. Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 3rd edn., 2000). Unless otherwise stated, all *Streptomyces* media are listed in T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood, *Practical Streptomyces Genetics*, John Innes Foundation, Norwich, 2000. Chemical reagents, DNA oligonucleotides and media components were purchased from Sigma-Aldrich or BD Biosciences and used without further purification. Restriction endonucleases were purchased from New England Biolabs and remaining enzymes from Invitrogen.

1.2 Bacterial strains, plasmids and culture conditions

All *S. coelicolor* strains were propagated on MS agar, *S. chartreusis* NRL3882 on OB agar and *S. lysosuperificus* ATCC31396 on MYM agar, at 30°C. *E. coli* strains and *B. subtilis* EC1524

were routinely grown in Luria-Bertani broth (LB) or on 1.5% LB agar plates supplemented with appropriate antibiotics. For recombinant strain selection, antibiotics were used in the following concentrations: carbenicillin (100 µg/mL), kanamycin (50 µg/mL), apramycin (50 µg/mL), chloramphenicol (25 µg/mL) or nalidixic acid (25 µg/mL). For the isolation of genomic DNA, 10 µL of dense *S. chartreusis* or *S. lysosuperificus* spore preparations were inoculated into 50 mL TSB/YEME (1:1) in a 250 mL flask and incubated with shaking at 250 rpm and 30°C for 24h. For heterologous production, the growth medium was replaced with TYD¹³ and inoculated with spores from appropriate recombinant strains.

1.3 Genome scanning of *S. chartreusis* NRL3882

Isolation of high molecular weight genomic DNA from *S. chartreusis* NRL3882 and *S. lysosuperificus* AATCC31396 was accomplished by published procedures (T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood, *Practical Streptomyces Genetics*, John Innes Foundation, Norwich, 2000). Genomic DNA *S. chartreusis* NRL3882 was sequenced and assembled by the University of Liverpool Advanced Genomics Facility using a Roche 454 Titanium pyrosequencing platform and the Roche Newbler (v2.0.00.20) assembler software. A BLAST database of the *S. chartreusis* genome was constructed and genome scanning, sequence analysis and functional annotation were performed using BLAST search tools (S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucl. Acids Res.*, 1997, **25**, 3389) and the Artemis v12.0 (K. Rutherford, J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.-A. Rajandream and B. Barrell, *Bioinformatics*, 2000, **16**, 944) software package. A putative *tun* gene cluster was located spanning two non-overlapping contigs and the constituent genes were labeled *tunA-N*. Using primers based on internal fragments from both ends of *tunD*, generation and sequencing of a PCR product spanning the gap between the two contigs showed their sequences to be contiguous on the bacterial chromosome.

1.4 Generation and screening of *S. chartreusis* genomic library

Genomic DNA was partially digested with *Sau3AI* to yield 30-60 kb restriction fragments. Size-fractionated fragments were cloned into SuperCos1 (Stratagene), packaged using the Gigapack III Gold Packaging Extract Kit (Stratagene) and transduced into *E. coli* XL1 Blue MR (Stratagene), in each case following the manufacturers' instructions. At the Genome Analysis Centre (Norwich, UK), 3073 individual cosmid clones were picked and transferred to 96-well microtiter plates containing LB medium and ampicillin. DNA from these colonies was fixed onto nylon membrane filters according to published methods (J. Sambrook and D. Russell, *Molecular Cloning: A*

Laboratory Manual, Cold Spring Harbor Laboratory Press, New York, 3rd edn., 2000). Internal fragments of genes *tunA* and *tunN* were amplified by PCR from genomic DNA, labeled with ³²P using the Rediprime II DNA Labelling System (Amersham) and used as hybridization probes for library screening. Eight clones hybridizing to both probes were subjected to restriction analysis with BamHI and *Xho*I, resulting in the selection of four cosmids (4H8, 5K7, 6N9 and 7C3) containing a complete, centrally located *tun* gene cluster.

1.5 Preparation of recombinant *S. coelicolor* strains harbouring the *tun* gene cluster

The SuperCos1-derived library cosmids 4H8, 5K7, 6N9 and 7C3 were made conjugative and integrative by λ Red-mediated recombination of the vector sequence with a 5.2 kb *Ssp*I fragment from pMJCOS1 that contained an apramycin resistance cassette *aac3(IV)*, *oriT* and ϕ C31 integrase *int* and attachment site *attB*, as well as flanking sequences with identity to corresponding regions of the SuperCos1 backbone. The resultant constructs, named pIJ12315-8, were introduced into *S. coelicolor* M1152 via *E. coli* ET12567/pUZ8002 by conjugation according to published procedures (B. Gust, T. Kieser and K. F. Chater, *REDIRECT technology: PCR-targeting system in Streptomyces coelicolor*, John Innes Foundation, Norwich, 2002) and analysed for tunicamycin production. In addition, a 12.9 kb *Sac*I fragment from cosmid 4H8 was cloned directly into the *Sac*I site pRT802 yielding pIJ12003a. The *Sac*I fragment contained the complete predicted *tun* gene cluster plus 427 bp upstream of *tunA* and 500 bp downstream of *tunN*; neither of these flanking DNA sequences is predicted to possess an entire ORF. Subsequently, pIJ12003a was conjugated into *S. coelicolor* M1146 via triparental mating using *E. coli* S17-1 and *E. coli* ET12567/pUZ8002.

1.6 Analysis of tunicamycin production by recombinant *S. coelicolor* Strains

Recombinant *S. coelicolor* M1025-1028, M1031 and their vector-only control strains were grown on R5 agar at 30°C for 48 h. Agar cores were transferred to empty plates, which were then flooded with 50°C soft nutrient agar preinfected with tunicamycin-sensitive *B. subtilis* EC1524. Additionally, sterile filter disks spotted with 15 μ L of tunicamycin stock solution (1 mg/mL) were placed atop the solidified agar. The plates were grown at 30°C for 18 h and examined for growth inhibition of the reporter strain. Furthermore, recombinant strains were grown in liquid TYD medium for 5 days at 30°C and the suspected tunicamycin metabolites were isolated and characterized by methanolic mycelium extraction and LC/MS analysis, as described previously (B. C. Tsvetanova, D. J. Kiemle and N. P. J. Price, *J. Biol. Chem.*, 2002, **277**, 35289) using a Micromass LCT (ESI-TOFMS) coupled to an Agilent 1200 Series LC System. Crude extracts were purified by flash chromatography using Fluka Kiegselgel 60 220-440 mesh silica gel (mobile phase:

water/isopropanol/ethyl acetate 1:3:6) and subjected to ^1H NMR spectroscopy along with an authentic sample to further confirm the presence of tunicamycin.

2 Results

2.1 Identification of the tunicamycin biosynthetic gene cluster

High molecular weight genomic DNA was isolated from the two bacterial strains known to produce tunicamycin – *S. chartreusis* NRL3882 and *S. lysosuperificus* ATCC31396. The latter was shown to contain a high copy number plasmid and since this would bias subsequent sequencing data towards plasmid sequences, we elected to work with *S. chartreusis* NRL3882. The genome was sequenced to 36 $\times$ coverage, generating 3112 contigs with a maximum size of 53.9 kb and an N50 average size of 4.6 kb. The contigs covered 7.95 Mb of the *S. chartreusis* chromosome with a G+C content of 70%, consistent with existing data for *Streptomyces* genomes.

Using bioinformatics tools tBLASTn and Artemis, these contigs were scanned for the presence of candidate *tun* genes. The tunicamycins contain many unique structural motifs (Figure 1) created by proteins with functions for which no similar examples or parallels are known. We therefore selected from a wide range of existing gene products with possible, chemically-similar function to putative members of the tunicamycin biosynthetic cluster. These focused on unique features and included: 1 $\rightarrow$ 1-linking glycosyltransferases (such as those from 3,3'-neotrehalosadiamine biosynthesis²⁴ and from the OtsA-OtsB, TreY-TreZ and TreS pathways of trehalose biosynthesis²⁵), *N*-acetyl-hexosamine *N*-deacetylases (e.g., from mycothiol, glycerophosphatidylinositol, neomycin and teicoplanin biosynthesis.), lipid-processing proteins and numerous examples from the abundant NDP-hexose epimerase and dehydratase families. Using the combined presence of these reactivities as a 'filter' to mine the genome sequence data, a single contig was identified containing homologues of inositol-GlcNAc deacetylase, hexose epimerase/dehydratase and acyl carrier protein, in a cluster of eleven open reading frames (ORFs). The 3'-end of this operon coincided with the end of the contig and terminated with the partial sequence of a putative GT-1 family glycosyltransferase (termed *tunD*), another key element of our reactivity filter. The full length sequences of glycosyltransferases homologous to *tunD* were screened against the *S. chartreusis* genome and an additional contig detected, containing the remaining portion of *tunD* and three further ORFs. In this way, the use of a bioinformatics filter, based on chemical logic, uniquely suggested the unification of an operon fractured across two distinct contigs. This unification was confirmed experimentally by

generating a PCR product to bridge the gap between the two contigs, using primers matching the 5' and 3' ends of *tunD* and a genomic DNA template; its sequence confirmed that the two contigs were indeed adjacent on the chromosome and revealed that there were no bases between them. This joined contig contained 14 ORFs that appear to lie in a single operon, with many translationally coupled to the preceding gene (Figure 2). Further bioinformatic analysis revealed that these genes are likely to constitute a tunicamycin biosynthetic gene cluster and the predicted function of genes *tunA-N* and flanking ORFs are presented in Table 5.

Table 5 - Deduced functions of *tun* genes and comparison with homologues in *S. clavuligerus* and *A. mirium*

ORF	aa ^a	Proposed function in tunicamycin biosynthesis	Closest Protein Homologue ^d , Origin, (%Id/Si) ^b , Acc. ^c	Homologue in <i>S. clavuligerus</i> ATCC27064, aa, (%Id/Si), Acc.	Homologue in <i>A. mirum</i> DSM43827, aa, (%Id/Si), Acc.
ORF-2	406		Integrase, <i>Streptomyces sviveus</i> ATCC 29083, (52/65), ZP_05020140		
ORF-1	297		Transposase, <i>Rhodococcus jostii</i> RHA1, (50/64), YP_700005		
TunA	321	UDP-GlcNAc epimerase/dehydrata se	NAD-dependent epimerase/dehydratase, <i>Streptomyces</i> sp. Mg1, (42/58), ZP_04996782	SCLAV_4287, 276, (72/80), ZP_06773762	Amir_2816, 322, (54/65), YP_003100592
TunB	338	Uridine oxidoreductase	Radical SAM domain protein, <i>Haloterrigena turkmenica</i> DSM5511, (32/48), YP_003405396	SCLAV_4286, 338, (90/95), ZP_06773761	Amir_2817, 340, (78/87), YP_003100593
TunC	318	<i>N</i> -acyltransferase	GCN5-related <i>N</i> -acetyltransferase, <i>Fervidobacterium nodosum</i> , (33/50), YP_001410548	SCLAV_4285, 322, (60/72), ZP_06773760	Amir_2818, 318, (43/57), YP_003100594
TunD	474	Glycosyltransferase	Group 1 family glycosyltransferase, <i>Thermococcus barophilus</i> , (28/45), ZP_04876510	SCLAV_4284, 461, (63/77), ZP_06773759	Amir_2819, 451, (47/58), YP_003100595
TunE	234	<i>N</i> -deacetylase	GlcNAc-phosphatidylinositol de- <i>N</i> -acetylase, <i>Cylindrospermopsis raciborskii</i> , (43/57), ZP_06309433	SCLAV_4283, 236, (77/85), ZP_06773758	Amir_2820, 230, (63/76), YP_003100596
TunF	327	UDP-GlcNAc- 4-epimerase	UDP-glucose 4-epimerase, <i>Paenibacillus</i> sp. oral taxon 786, (46/63), ZP_04852226	SCLAV_4282, 327, (76/83), ZP_06773757	Amir_2821, 332, (58/70), YP_003100597
TunG	203	UMP phosphatase	Phosphoglycerate mutase, <i>Frankia</i> sp. CcI3, (29/47), YP_481446	SCLAV_4281, 208, (65/73), ZP_06773756	Amir_2822, 223, (50/60), YP_003100598
TunH	515	UDP- tunicaminylluracil pyrophosphatase	Type I nucleotide pyrophosphatase, <i>Burkholderia</i> sp. 383, (35/50), YP_370731	SCLAV_4280, 518, (66/76), ZP_06773755	Amir_2823, 510, (53/65), YP_003100599
TunI	304	ABC transporter ATP-binding subunit	Putative ABC transporter ATP- binding subunit, <i>Streptomyces scabiei</i> 87.22, (41/61), YP_003492364	SCLAV_4279, 302, (77/88), ZP_06773754	Amir_2824, 302, (60/73), YP_003100600
TunJ	262	ABC transporter permease subunit	ABC-2 type transporter, <i>Thermobaculum terrenum</i> , (32/51), YP_003322218	SCLAV_4278, 261, (76/83), ZP_06773753	Amir_2825, 253, (61/78), YP_003100601

TunK	81	Acyl carrier protein	Phosphopantetheine-binding protein, <i>Catenulispora acidiphila</i> DSM 44928, (32/61), YP_003117493	SCLAV_4277, 81, (65/87), ZP_06773752	Amir_2826, 79, (34/54), YP_003100602
TunL	229	Phospholipid phosphatase	Phosphoesterase PA-phosphatase, <i>Micromonospora aurantiaca</i> , (33/42), ZP_06217896	SCLAV_4276, 223, (52/67), ZP_06773751	-
TunM	216	Radical SAM protein	Methyltransferase family protein, <i>Saccharomonospora viridis</i> , (48/63), YP_003133112	SCLAV_4274, 212, (54/67), ZP_06773749	Amir_2815, 232, (30/54), YP_003100591
TunN	152	UTP pyrophosphatase	NUDIX hydrolase, <i>Nakamurella multipartita</i> , (36/55), YP_003200035	SCLAV_4275, 170, (68/77), ZP_06773750	-
ORF1	213		Secreted protein, <i>Streptomyces viridochromogenes</i> , (81/90), ZP_05533938		
ORF2	573		Secreted protein, <i>Streptomyces viridochromogenes</i> , (90/94), ZP_05533937		
ORF3	606		Secreted protein, <i>Streptomyces viridochromogenes</i> , (82/89), ZP_05533936		

2.2 Isolation and heterologous expression of the *tun* cluster

To confirm the involvement of the putative *tun* gene cluster in tunicamycin biosynthesis, it was introduced into *S. coelicolor* and the resulting recombinant strains screened for tunicamycin production. To do this, a cosmid library of *S. chartreusis* NRL3882 was assembled in *Escherichia coli* and probed with ³²P-labelled PCR amplicons from within *tunA* and *tunN*, which are located at either terminus of the putative *tun* operon (Figure 2). Cosmids hybridizing to both PCR probes were isolated and restriction analysis revealed four that contained the entire 12 kb *tun* gene cluster positioned centrally in the cosmid insert. The backbones of these SuperCos1-based cosmids were subsequently exchanged for that of the conjugative and integrative vector pMJCOS1 through λ -RED-mediated recombination. These modified cosmids were transferred to *S. coelicolor* M1152 by conjugation, and integration into the chromosomal ϕ C31 phage attachment site achieved by selecting for apramycin resistance. In addition, a 12.9 kb *SacI* fragment from one of the four cosmids that contained the complete putative *tun* gene cluster plus 427 bp upstream of *tunA* and 500 bp downstream of *tunN* (neither of these additional DNA sequences is predicted to possess an entire ORF) was cloned into the conjugative and integrative vector pRT802. The resulting clone was similarly conjugatively transferred into an *S. coelicolor* M1146 host.

Heterologous expression of the putative *tun* gene cluster was monitored using an agar-diffusion bioassay; all five recombinant strains produced zones of inhibition when assayed against *Bacillus subtilis*, whereas control strains containing the relevant vectors alone did not (Figure 3A). This key observation shows the *tun* cluster codes for a secondary metabolite with bactericidal activity against the *B. subtilis* reporter strain. To confirm the identity of this bactericidal metabolite, recombinant strains containing the putative *tun* gene cluster and the relevant controls were grown in liquid culture for five days and the pelleted mycelium extracted with methanol. LC/MS analysis of the putative *tun*-containing clones revealed a mass distribution and fragmentation pattern identical to that of tunicamycin; this metabolite was absent from the control strains (Figure 3B). The presence of tunicamycin product was further confirmed by ¹H NMR spectroscopy (data now shown). The transfer of tunicamycin production to *S. coelicolor* M1146 by the *SacI*-cloned *tun* operon is particularly poignant, since it likely delineates the boundaries of the *tun* gene cluster and defines the minimal biosynthetic gene cluster necessary for tunicamycin production and for future mechanistic and redesign studies.

3 Discussion

3.1 Description of the *tun* gene cluster and identification of homologous clusters in *S. clavuligerus* and *A. mirum*

The likely minimal *tun* gene cluster identified by heterologous expression comprises a contiguous 12.0 kb stretch of DNA containing a total of 14 ORFs, all of which are oriented in the same direction with many translationally coupled to the preceding gene, presumably to ensure equivalent levels of synthesis of each enzyme. This suggests that the entire cluster is contained within a single polycistronic transcript, which is feasible given its small overall size. The overall G+C content of this region is 65.0%, well below that of a typical *Streptomyces* genome or indeed the rest of the *S. chartreusis* genome. This suggests the *tun* cluster was acquired from another, lower G+C-content, organism at some point during its evolution.

The ORFs flanking the proposed *tun* gene cluster are clearly not required for tunicamycin biosynthesis. The three flanking genes downstream of the *tun* cluster (ORF1-3) have close homologues in many *Streptomyces* genomes and encode conserved housekeeping genes. The two upstream flanking genes (ORF⁻¹ and ORF⁻²) are homologous to transposase and integrase genes respectively, lending support to the hypothesis that *S. chartreusis* acquired the *tun* gene cluster by lateral gene transfer. The 1.9 kb region between ORF⁻¹ and *tunA* contains a putative ORF with multiple frameshifts ("junk DNA"), again consistent with recent evolutionary

acquisition.

Bioinformatic analysis revealed other potential tunicamycin producers. Homologous gene clusters were identified in *Actinosynnema mirum* DSM43827 and *Streptomyces clavuligerus* ATCC27064; the latter has been reported to produce the closely related antibiotic MM19290. Only minor differences were observed between the three gene clusters, suggesting a recently shared evolutionary heritage (Figure 2).

Proteins encoded by the *S. chartreusis* gene cluster exhibited greatest similarity to those from *S. clavuligerus*, with amino acid sequence identities ranging from 52 to 90%. It is highly likely that genes annotated SCLAV_4274 to SCLAV_4287 are responsible for MM19290 biosynthesis in *Streptomyces clavuligerus* ATCC27064. Although the structure of MM19290 has not been reported, the high degree of homology with the *tun* genes from *S. chartreusis* strongly suggests that – like the streptovirudins and corynetoxins – this compound shares its core carbohydrate skeleton with tunicamycin.

Proteins encoded by the *A. mirum* cluster exhibit amino acid sequence identities with Tun proteins from *S. chartreusis* that range from 30 to 78%. While no full-length homologues of TunN or TunL were found, closer inspection of the *A. mirum* sequence revealed a truncated version TunL that contained a number of frameshift mutations. Since this organism has not been reported to produce any antibiotics structurally related to the tunicamycins, it is probable that we have uncovered a silent gene cluster that has lost the ability to produce its tunicamycin-like metabolite.

3.2 Proposed biosynthetic pathway for the tunicamycins

Bioinformatic analysis of the *tun* gene cluster with BLAST and Artemis, together with identification of conserved active-site residues, allowed us to predict the functions of the products of *tunA* through to *tunN* to be predicted (Table 1). These assignments reconcile the genetic insight gained here with previous feeding experiments using labeled precursors that together allow us to propose a detailed biosynthetic pathway to the tunicamycins (Figure 4). Construction of the tunicaminy-uracil core proceeds via the tail-to-tail coupling of uridine and galactosamine derivatives through a C–C linkage. The involvement of the UDP-4-keto-5,6-ene-GlcNAc intermediate is supported by the presence of *tunF* and *tunA*, coding for a UDP-hexose-4-epimerase and a UDP-GlcNAc epimerase/dehydratase respectively and acting on UDP-GlcNAc – previously established as a metabolic precursor. The presence of two enzymes of similar/related function may suggest that since C-4 of the α,β -unsaturated intermediate has lost all stereochemical information, its subsequent reduction after a coupling event may be enzymatically stereocontrolled. Uridine-5'-aldehyde is also likely to feature as an intermediate

and has been implicated in the biosynthesis of nikkomycin, polyoxin, liposidomycin and capuramycin nucleoside antibiotic families, although its formation and mechanistic role has not been studied in detail and remains poorly understood.

For tunicamycin, we propose the TunB-mediated formation, from uridine, of a radical SAM protein containing a 4Fe4S redox centre. The requisite uridine is obtained from UTP by the sequential action of TunN (a nucleotide pyrophosphatase) and TunG (a nucleotide monophosphate phosphatase) respectively. The coupling event is likely to be mediated by TunB in combination with TunM, a methyltransferase homologue. Both of these enzymes catalyse radical processes, and thus the coupling of the two activated carbohydrate intermediates may proceed via a radical mechanism either by addition to an α,β -unsaturated ketone or through a Barbier-type mechanism. Subsequent tailoring of the pseudodisaccharide tunicaminy-uracil core involves the formation of an α,β -1,1-trehalose linkage. Nucleotide-sugar pyrophosphatase TunH likely catalyses the hydrolysis of UDP from this sugar for subsequent transfer of GlcNAc to the liberated anomeric position. This step is catalysed by GT-1 family glycosyltransferase TunD, yielding the core pseudotrisaccharide skeleton of the tunicamycins with concurrent formation of two new stereocentres in the α,β -1,1-glycosidic bond. The final modification to this skeleton involves the introduction of a range of acyl chains to form each of the up to eighteen tunicamycin homologues that have been described. Since the heterologously expressed *tun* gene cluster produced fully acylated tunicamycins despite lacking a fatty acid synthase gene, the constituent acyl chains are most likely derived from the cellular pool of fatty acids, as previously observed in teicoplanin biosynthesis. The most likely function of TunL, a putative type 2 phosphatidic acid phosphatase (PAP2), is in the regulation of lipid synthesis in the producing bacterium. By down-regulating the levels of cellular phosphatidic acid and up-regulating levels of its cleavage product diacylglycerol, phospholipid biosynthesis is repressed and cellular pools of fatty acids can be diverted for use in tunicamycin biosynthesis via β -oxidative degradation pathways. Tunicamycin-producing organisms appear to have evolved an efficient way of perturbing the complex regulatory pathways of lipid metabolism regulation, allowing increased tunicamycin biosynthesis without negatively affecting vital cellular processes. Acyl carrier protein TunK next activates these sequestered fatty acids for subsequent acylation, presumably through the action of a fatty acyl-ACP ligase from primary metabolism, since no such ligase is present in the *tun* gene cluster. The tunicamycin core skeleton is prepared for amide bond formation by *N*-deacetylation with TunE, a member of the GlcNAc *N*-deacetylase family. TunC subsequently functions as an *N*-acyltransferase to install the sequestered and activated fatty acids, yielding the full range of tunicamycin homologues.

The *tun* gene cluster described is relatively small in size, although a previous suggestion that as few as five genes would be necessary for the biosynthesis of tunicamycin has proved too conservative. Of the nine additional genes not originally predicted, two are involved in the generation of free uridine from UTP, contrary to suggestions that uridine would be obtained directly from primary metabolism. Two further genes are implicated in formation of UDP-tunicaminy-uracil – one coding for a sugar epimerase supplementary to the dehydratase catalyzing UDP-4-keto-5,6-ene-GlcNAc formation and one which mediates the radical coupling event alongside the gene responsible for uridine oxidation. Hydrolysis of UDP from the undecose intermediate has also been shown to require enzyme catalysis. Although the acyl side chains are likely to originate from cellular pools of fatty acids – consistent with the lack of a fatty acid synthase – the *tun* gene cluster still encodes two enzymes that provide sufficient fatty acid flux and are involved in sequestering lipids and processing them prior to attachment. Finally, the last two additional *tun* genes are not directly involved in tunicamycin biosynthesis, but are likely to be crucial in conferring self-resistance to the producing organism. *tunI* and *tunJ* together encode for an ABC transporter, homologues of which are responsible for rapid ATP-driven efflux of antibiotics from cells in a large number of antibiotic-producing organisms.

No regulatory genes were found in the *tun* gene cluster, suggesting that tunicamycin production may be subject to global control associated with growth rate reduction. The presence of rare TTA leucine codons (only 2% of *S. coelicolor* genes contain a TTA codon) in *tunA* and *tunM* may well reflect an element of translational regulation. In *S. coelicolor*, the accumulation of $_{\text{Leu}}\text{tRNA}^{\text{UUA}}$ is temporally regulated, and translation of mRNAs containing this codon may be largely confined to later stages of growth.

4 Conclusions

With well over 8000 literature citations, the tunicamycins have attracted a great deal of attention for many years thanks to their unique structure and function, and their potent and specific inhibition of *N*-acetyl-D-hexosamine-1-phosphate translocases involved in important cellular processes – particularly eukaryotic protein *N*-glycosylation and bacterial peptidoglycan biosynthesis. In this application, we have identified the biosynthetic genes of a tunicamycin-family antibiotic for the first time, offering rich insights into the poorly understood biosynthetic pathway of this fascinating family of nucleoside antibiotics. Through molecular cloning and heterologous expression of the *tun* gene cluster in a *S. coelicolor* host, we have identified the minimal set of genes required for tunicamycin production. Additionally, we have identified

close homologues of the *tun* gene cluster in *A. mirum* DSM43827 and *S. clavuligerus* ATCC27064. The latter organism is known to produce MM19290 – an antibiotic closely related to tunicamycin – and based on the close similarity of its homologous cluster with the *tun* genes, we suggest that this cluster is likely responsible for MM19290 biosynthesis in *S. clavuligerus* ATCC27064. Furthermore, we propose MM19290 shares the core structure of the tunicamycins and differs only in the nature of its acyl side chains. The availability of relatively inexpensive high throughput sequencing, combined with the genome scanning approach guided by the chemical logic described here provides a rapid and efficient way of identifying natural product gene clusters. The exponential increase in publically available gene sequences in recent years has dramatically expanded the possibilities afforded by bioinformatic analysis. Our results suggest that such *in silico* mining of partially assembled genome sequences will constitute an increasingly effective tool during the early stages of dissecting a bacterial biosynthetic pathway.

The findings presented here will allow detailed studies of tunicamycin biosynthesis. Functional characterization of individual enzymes will provide insights into how some of the unique linkages in tunicamycin are constructed. In addition, armed with this comprehensive toolbox of biosynthetic machinery, tunicamycin derivatives with altered selectivity for bacterial *MraY* *versus* human GPT can now be sought, potentially leading to future therapeutic antibiotics with improved antibacterial activity and reduced cytotoxicity. Importantly, the mode of action of tunicamycin is orthogonal to all existing antibiotic drugs. Tunicamycin also provides a unique natural product template for inhibition of carbohydrate processing enzymes. It represents a likely transition state mimic and hence transition state mimics of other important nucleotide sugar-dependant carbohydrate processing enzymes might also be targeted by precursor-driven biosynthesis or chemoenzymatic methods, exchanging terminal functionalities of the tunicamycin structure.

CLAIMS

1. A polynucleotide sequence comprising the sequence shown in nucleotides 8616 to 20606 of SEQ ID NO: 1 or a variant having at least 50% homology to nucleotides 8616 to 20606 of SEQ ID NO: 1 over its entire sequence based on nucleotide identity.
2. A polynucleotide sequence comprising:
 - (i) the sequence shown in SEQ ID NO: 2 or a variant having at least 75% homology to SEQ ID NO: 2 over its entire sequence based on nucleotide identity;
 - (ii) the sequence shown in SEQ ID NO: 4 or a variant having at least 8 homology to SEQ ID NO: 4 over its entire sequence based on nucleotide identity;
 - (iii) the sequence shown in SEQ ID NO: 6 or a variant having at least 71% homology to SEQ ID NO: 6 over its entire sequence based on nucleotide identity;
 - (iv) the sequence shown in SEQ ID NO: 8 or a variant having at least 72% homology to SEQ ID NO: 8 over its entire sequence based on nucleotide identity;
 - (v) the sequence shown in SEQ ID NO: 10 or a variant having at least 76% homology to SEQ ID NO: 10 over its entire sequence based on nucleotide identity;
 - (vi) the sequence shown in SEQ ID NO: 12 or a variant having at least 75% homology to SEQ ID NO: 12 over its entire sequence based on nucleotide identity;
 - (vii) the sequence shown in SEQ ID NO: 14 or a variant having at least 70% homology to SEQ ID NO: 14 over its entire sequence based on nucleotide identity;
 - (viii) the sequence shown in SEQ ID NO: 16 or a variant having at least 71% homology to SEQ ID NO: 16 over its entire sequence based on nucleotide identity;
 - (ix) the sequence shown in SEQ ID NO: 18 or a variant having at least 76% homology to SEQ ID NO: 18 over its entire sequence based on nucleotide identity;
 - (x) the sequence shown in SEQ ID NO: 20 or a variant having at least 77% homology to SEQ ID NO: 20 over its entire sequence based on nucleotide identity;
 - (xi) the sequence shown in SEQ ID NO: 22 or a variant having at least 79% homology to SEQ ID NO: 22 over its entire sequence based on nucleotide identity;
 - (xii) the sequence shown in SEQ ID NO: 24 or a variant having at least 66% homology to SEQ ID NO: 24 over its entire sequence based on nucleotide identity;
 - (xiii) the sequence shown in SEQ ID NO: 26 or a variant having at least 72% homology to SEQ ID NO: 26 over its entire sequence based on nucleotide identity; or

(xiv) the sequence shown in SEQ ID NO: 28 or a variant having at least 68% homology to SEQ ID NO: 28 over its entire sequence based on nucleotide identity.

3. A polynucleotide sequence according to claim 2, wherein the variant encodes a polypeptide having a substrate specificity that differs from the corresponding wild-type polypeptide.
4. A polynucleotide sequence according to claim 3, wherein the variant of SEQ ID NO: 8 encodes a polypeptide having a carbohydrate specificity that differs from that of SEQ ID NO: 9.
5. A polynucleotide sequence according to claim 3, wherein the variant of SEQ ID NO: 6 encodes a polypeptide having a fatty acid specificity that differs from that of SEQ ID NO: 7.
6. A polynucleotide sequence according to claim 3, wherein the variant of SEQ ID NO: 22 encodes a polypeptide having a fatty acid specificity that differs from that of SEQ ID NO: 23.
7. A polynucleotide sequence according to claim 3, wherein the variant of SEQ ID NO: 24 encodes a polypeptide having a phospholipid specificity that differs from that of SEQ ID NO: 25.
8. A polynucleotide construct comprising more than one of the polynucleotide sequences defined in any one of claims 2 to 7.
9. A polynucleotide construct according to claim 8, wherein the construct comprises fourteen polynucleotide sequences as defined in parts (i) to (xiv) of claim 2.
10. A polynucleotide construct according to claim 8, wherein the construct comprises at least five polynucleotide sequences as defined in parts (i) to (xiv) of claim 2.
11. A polynucleotide construct according to claim 10, wherein the construct comprises five polynucleotide sequences as defined in parts (iii), (iv), (v), (xi) and (xii) of claim 2.
12. A polynucleotide construct according to any one of claims 8 to 12, wherein the variant is as defined in any one of claims 3 to 8.

13. A vector comprising a polynucleotide sequence according to any one of claims 1 to 7 or a polynucleotide construct according to any one of claims 8 to 12 operably linked to a control sequence.
14. A host cell comprising a polynucleotide sequence according to any one of claims 1 to 7, a polynucleotide construct according to any one of claims 8 to 12 or a vector according to claim 13.
15. A polypeptide encoded by a polynucleotide according to any one of claims 2 to 7.
16. A polypeptide sequence comprising:
- (a) the sequence shown in SEQ ID NO: 3 or a variant having at least 73% homology to SEQ ID NO: 3 over its entire sequence based on amino acid identity;
 - (b) the sequence shown in SEQ ID NO: 5 or a variant having at least 91% homology to SEQ ID NO: 5 over its entire sequence based on amino acid identity;
 - (c) the sequence shown in SEQ ID NO: 7 or a variant having at least 61% homology to SEQ ID NO: 7 over its entire sequence based on amino acid identity;
 - (d) the sequence shown in SEQ ID NO: 9 or a variant having at least 64% homology to SEQ ID NO: 9 over its entire sequence based on amino acid identity;
 - (e) the sequence shown in SEQ ID NO: 11 or a variant having at least 78% homology to SEQ ID NO: 11 over its entire sequence based on amino acid identity;
 - (f) the sequence shown in SEQ ID NO: 13 or a variant having at least 77% homology to SEQ ID NO: 13 over its entire sequence based on amino acid identity;
 - (g) the sequence shown in SEQ ID NO: 15 or a variant having at least 66% homology to SEQ ID NO: 15 over its entire sequence based on amino acid identity;
 - (h) the sequence shown in SEQ ID NO: 17 or a variant having at least 67% homology to SEQ ID NO: 17 over its entire sequence based on amino acid identity;
 - (i) the sequence shown in SEQ ID NO: 19 or a variant having at least 78% homology to SEQ ID NO: 19 over its entire sequence based on amino acid identity;
 - (j) the sequence shown in SEQ ID NO: 21 or a variant having at least 77% homology to SEQ ID NO: 21 over its entire sequence based on amino acid identity;
 - (k) the sequence shown in SEQ ID NO: 23 or a variant having at least 66% homology to SEQ ID NO: 23 over its entire sequence based on amino acid identity;
 - (l) the sequence shown in SEQ ID NO: 25 or a variant having at least 53% homology to SEQ ID NO: 25 over its entire sequence based on amino acid identity;

(m) the sequence shown in SEQ ID NO: 27 or a variant having at least 55% homology to SEQ ID NO: 27 over its entire sequence based on amino acid identity; or

(n) the sequence shown in SEQ ID NO: 29 or a variant having at least 69% homology to SEQ ID NO: 29 over its entire sequence based on amino acid identity.

17. A polypeptide according to claim 16, wherein the variant has a substrate specificity that differs from the corresponding wild-type polypeptide.

18. A polypeptide according to claim 17, wherein the variant of SEQ ID NO: 9 has a carbohydrate specificity that differs from that of SEQ ID NO: 9.

19. A polypeptide according to claim 17, wherein the variant of SEQ ID NO: 7 has a fatty acid specificity that differs from that of SEQ ID NO: 7.

20. A polypeptide according to claim 17, wherein the variant of SEQ ID NO: 23 has a fatty acid specificity that differs from that of SEQ ID NO: 23.

21. A polypeptide according to claim 17, wherein the variant of SEQ ID NO: 25 has a phospholipid specificity that differs from that of SEQ ID NO: 25.

22. A heterologous expression system encoded by a polynucleotide according to claim 1.

23. A heterologous expression system comprising more than one of the polypeptides defined in any one of claims 16 to 21.

24. A heterologous expression system according to claim 23, wherein the construct comprises fourteen polypeptides as defined in parts (a) to (n) of claim 16.

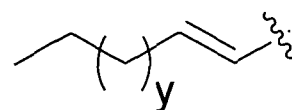
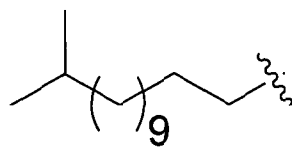
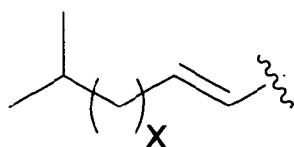
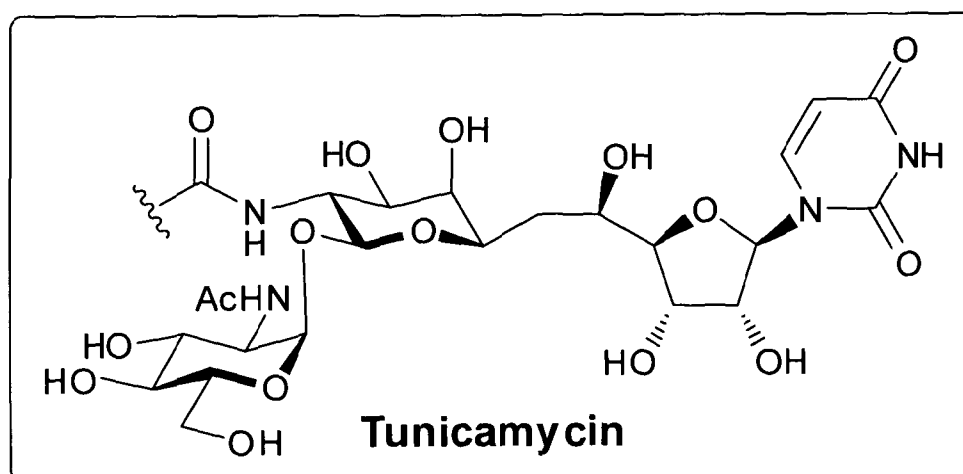
25. A heterologous expression system according to claim 23, wherein the construct comprises at least five polypeptides as defined in parts (a) to (n) of claim 16.

26. A heterologous expression system according to claim 25, wherein the construct comprises five polypeptides as defined in parts (c), (d), (e), (k) and (l) of claim 16.

27. A heterologous expression system according to any one of claims 22 to 26, wherein the variant is as defined in any one of claims 16 to 21.
28. A host cell comprising a heterologous expression system according to any one of claims 22 to 27.
29. A method of producing a tunicamycin or a derivative thereof, the method comprising culturing a host cell according to claim 14 or 28 and isolating the tunicamycin or derivative thereof.
30. A method according to claim 29, wherein the derivative inhibits a different enzyme involved in bacterial cell wall biosynthesis from that inhibited by a tunicamycin.
31. A tunicamycin or a derivative thereof produced using a method according to claim 29 or 30.
32. A derivative according to claim 31, wherein the derivative is highly selective for MraY.
33. A derivative according to claim 31 or 32, wherein the derivative does not bind to the active site of UDP-GlcNAc:dolichyl phosphate GlcNAc-1-phosphate transferase (GPT).
34. A pharmaceutical composition comprising a tunicamycin or a derivative thereof according to any one of claims 31 to 33 and a pharmaceutically acceptable carrier.
35. A tunicamycin or a derivative thereof according to any one of claims 31 to 33 for use in a method of treatment of the human or animal body.
36. A tunicamycin or a derivative thereof according to any one of claims 31 to 33 for use in a method of treating or preventing a bacterial infection in a subject.
37. A method of treating or preventing a bacterial infection in a subject comprising administering to said subject a therapeutically or prophylactically effective amount of a tunicamycin or a derivative according to any one of claims 31 to 33.

1/4

Fig. 1



$$x = \begin{bmatrix} 7 & \text{Tun13:1} \\ 8 & \text{Tun14:1}_A \\ 9 & \text{Tun15:1}_B \\ 10 & \text{Tun16:1}_A \\ 11 & \text{Tun17:1}_B \end{bmatrix}$$

Tun15:0

$$y = \begin{bmatrix} 9 & \text{Tun14:1}_B \\ 10 & \text{Tun15:1}_A \\ 11 & \text{Tun16:1}_B \\ 12 & \text{Tun17:1}_A \end{bmatrix}$$

Fig. 2

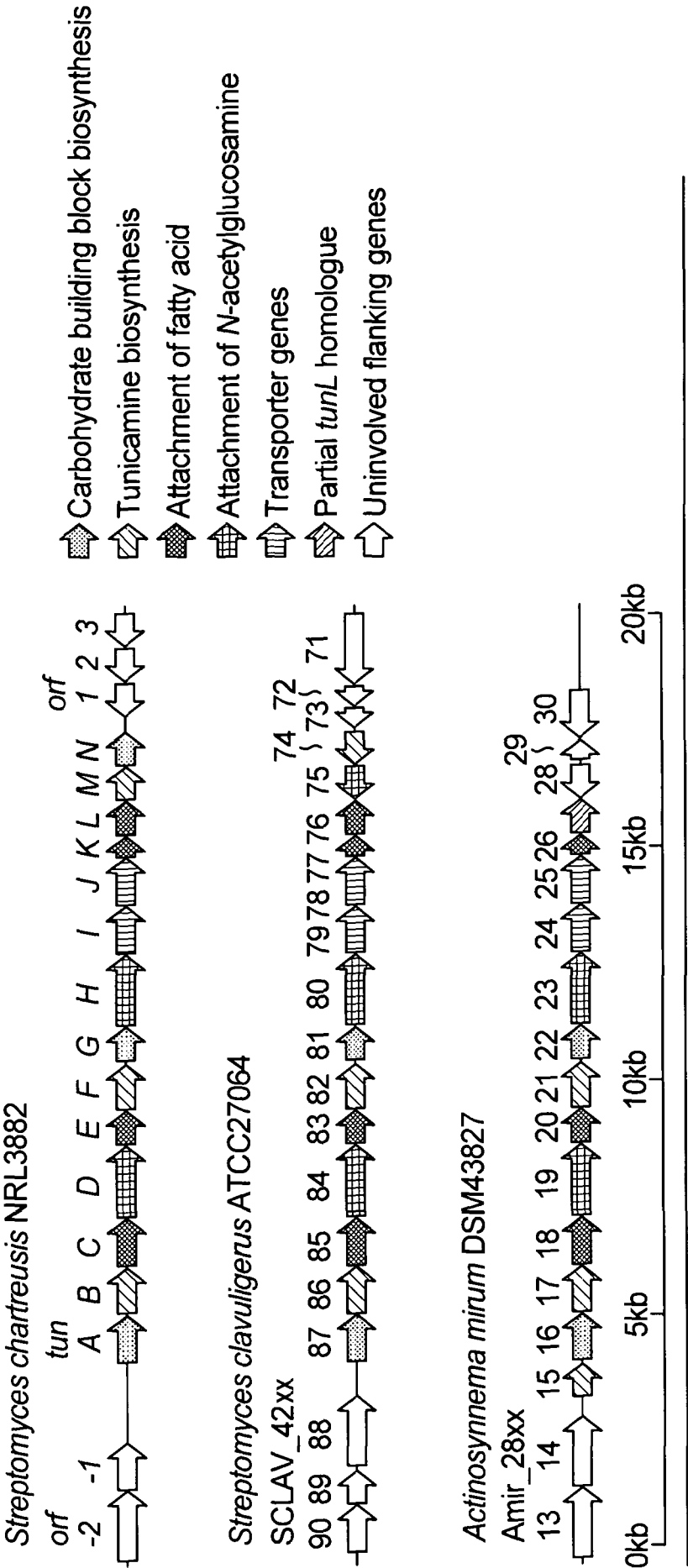


Fig. 3

3/4

