Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic

André MA van Wageningen[1], Peter N Kirkpatrick[1], Dudley H Williams[1], Barbara R Harris[2], Jo K Kershaw[2], Nicola J Lennard[2], M Jones[2], Steven JM Jones[2] and Patricia J Solenberg[3]

Background: The emergence of resistance to vancomycin, the drug of choice against methicillin-resistant *Staphylococcus aureus*, in enterococci has increased the need for new antibiotics. As chemical modification of the antibiotic structure is not trivial, we have initiated studies towards enzymatic modification by sequencing the DNA coding for the biosynthesis of chloroeremomycin (also known as A82846B and LY264826).

Results: Analysis of 72 kilobases of genomic DNA from *Amycolatopsis orientalis*, the organism that produces chloroeremomycin, revealed the presence of 39 putative genes, including those coding for the biosynthesis of the antibiotic. Translation and subsequent comparison with known proteins in public databases identified enzymes responsible for the biosynthesis of the heptapeptide backbone and 4-*epi*-vancosamine, as well as those for chlorination and oxidation reactions involved in the biosynthesis of chloroeremomycin.

Conclusions: The genes responsible for the biosynthesis of chloroeremomycin have been identified, and selective expression of these genes could lead to the synthesis of new potent glycopeptide antibiotics.

Addresses: [1]Cambridge Centre for Molecular Recognition, Department of Chemistry, Lensfield Road, Cambridge, CB2 1EW, UK. [2]The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK. [3]Lilly Research Laboratories, Eli Lilly & Co, Indianapolis, IN 46285, USA.

Correspondence: Dudley H Williams
E-mail: dhw1@cam.ac.uk

Key words: antibiotic biosynthesis, chloroeremomycin, DNA sequence, glycopeptides

Received: **5 December 1997**
Revisions requested: **9 January 1998**
Revisions received: **26 January 1998**
Accepted: **30 January 1998**

Published: **10 March 1998**

Chemistry & Biology March 1998, **5**:155–162
http://biomednet.com/elecref/1074552100500155

© Current Biology Ltd ISSN 1074-5521

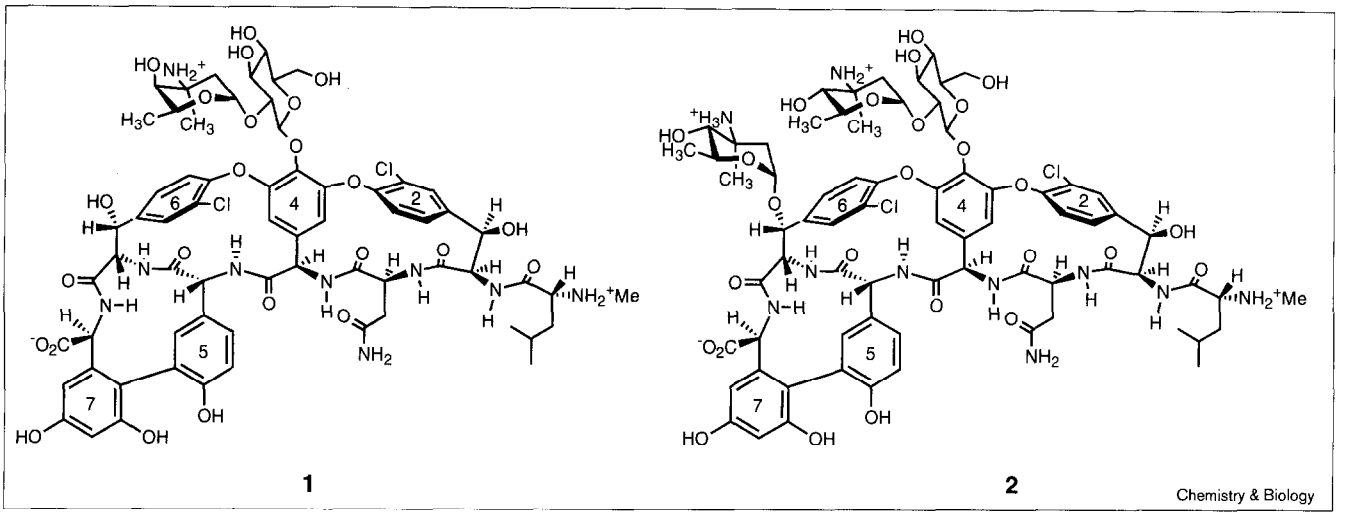
Introduction

Vancomycin (**1**, Figure 1) and teicoplanin are the antibiotics of choice against methicillin-resistant *Staphylococcus aureus* (MRSA) [1]. Vancomycin-resistant enterococci (VRE) currently cause problems as hospital pathogens, increasing the need for antibiotics active against VRE [2–4]. Our research has been directed towards a better understanding of the molecular mode of action of vancomycin group antibiotics and has led to a generally accepted model. The model includes the molecular basis of binding to mucopeptide cell wall precursors terminating in –Lys–D-Ala–D-Ala, the importance of dimerisation of the antibiotic, which is promoted by the presence of sugar residues and chlorine atoms, and the importance of membrane anchoring in the case of teicoplanin [5]. It appears that these findings, taken collectively, can account for the remarkable activity of a new semi-synthetic glycopeptide antibiotic, derived from chloroeremomycin (**2**, also known as A82846B and LY264826), against VRE [6].

As modifications of the glycopeptide antibiotics via standard organic synthetic pathways are somewhat restricted, we decided to sequence the DNA that is thought to code for the biosynthesis of chloroeremomycin (**2**), with a view to using the identified enzymes to produce new glycopeptide antibiotics (e.g. by introducing desirable structural features onto the heptapeptide backbone). The peptide backbone is a conserved feature of the glycopeptide antibiotics, although different amino acids are frequently found for different antibiotics at residues 1 and 3 (numbering from the amino terminus). For chloroeremomycin, the backbone is based on H_2N-*R*-Leu-*R*-Tyr-*S*-Asn-*R*-HPG-R-HPG-*S*-Tyr-*S*-DHPG-OH (HPG, 4-hydroxyphenylglycine; DHPG, 3,5-dihydroxyphenylglycine; *R* and *S* indicate the absolute configuration at the α carbon of the amino acids, Figure 1) [5]. Previous research in our group has shown that for vancomycin (and hence, by analogy, probably also for chloroeremomycin) and ristocetin A, the DHPG moieties are derived from acetate, probably via a polyketide mechanism [7], whereas the HPG residues are derived from tyrosine [7,8].

Recently, identification and expression of three glycosyl transferases, involved in the introduction of the sugar moieties onto chloroeremomycin, have been reported [9]. It was shown that these enzymes could be used to produce new glycopeptide antibiotics. In this paper, we describe the sequencing and analysis of 72 kilobases (kb)

Figure 1



Structures of vancomycin (1) and chloroeremomycin (2).

of genomic DNA from *Amycolatopsis orientalis*, the organism that produces chloroeremomycin. This sequence includes not only the genes coding for the above-mentioned glycosyl transferases, but also most of the remaining genes coding for the biosynthesis of chloroeremomycin. This is the first report of a relatively complete biosynthetic pathway to a vancomycin group antibiotic. Analysis of the identified genes not only leads to a better insight into the steps involved in the biosynthesis, but expression of suitable enzymes of the pathway may allow the synthesis of new antibiotics that could help to fight VRE, and possibly also vancomycin-resistant *S. aureus* (VRSA). Furthermore, the work will facilitate the characterisation of enzymes responsible for the introduction of special structural features present in vancomycin group antibiotics (such as aromatic ether and biaryl linkages, and chlorine substituents) that are themselves of wide interest.

Results and discussion

Compositional organisation and location of the proteins

Both cosmids used for the present sequencing work (pCZA361 and pCZA363) [9] are rich in GC, with a GC content of ~68%. The most commonly found start codon in the two cosmids is ATG (90%); TTG seems to be the start codon for one enzyme (ORF8) and GTG is found for the remaining enzymes.

Analysis of the 72 kb genomic DNA led to the location of 39 putative proteins; 9 on cosmid pCZA363 and 30 on cosmid pCZA361. The results are summarized in Figure 2. The genes on cosmid pCZA363 are all found on the same DNA strand, whereas six of the genes on cosmid pCZA361 are located on the opposite strand.

Enzymes and reactions involved in the biosynthesis of chloroeremomycin

Peptide synthetases

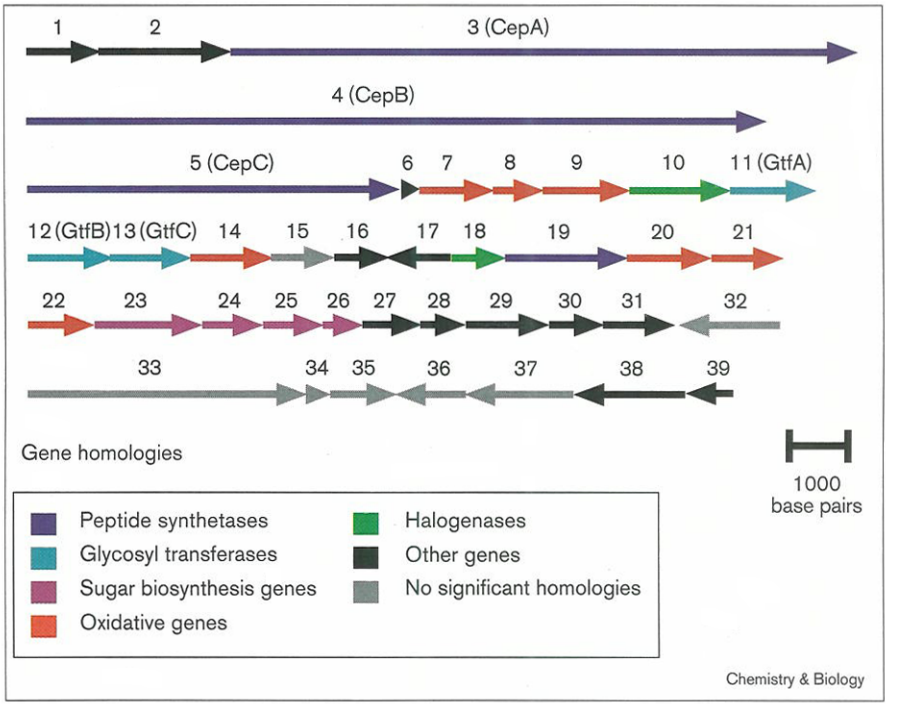
Analysis of the sequence revealed that three putative proteins have very significant homology to existing peptide synthetases, a class of large multienzymes that catalyse the formation of small peptides by a non-ribosomal route [10]. Examination of the protein sequences of these enzymes has indicated that they consist of structural modules, each of which is responsible for activating and binding one of the constituent amino acids. These modules are organised in an order that is colinear with the amino acid sequence of the peptide [10–12].

The process by which the peptide is synthesised is known as the multienzyme thiotemplate mechanism [10,13,14]. In the first step, the component amino acids are activated as aminoacyl adenylates. Each amino acid is then covalently linked to an enzyme-attached 4′-phosphopantetheine cofactor by thioesterification [14,15]. Peptide synthesis occurs by amide bond formation at the contact sites of the activating domains. Here, the growing peptide chain is transferred from the 4′-phosphopantetheine of one enzyme module to the next, each time increasing in length by one amino acid. Finally, the completed peptide is released from the last domain by thioesterase scission.

The coding sequence for the three peptide synthetases CepA, CepB, and CepC spans 27 kb at the start of the region of the bacterial genome linked with antibiotic biosynthesis. These putative enzymes have 3158, 4077 and 1860 amino acids, respectively. The protein sequence of each was analysed for the presence of standard motifs present in

Figure 2

Representation of the open reading frames (ORFs) located on cosmids pCZA361 (10–39) and pCZA363 (1–9).



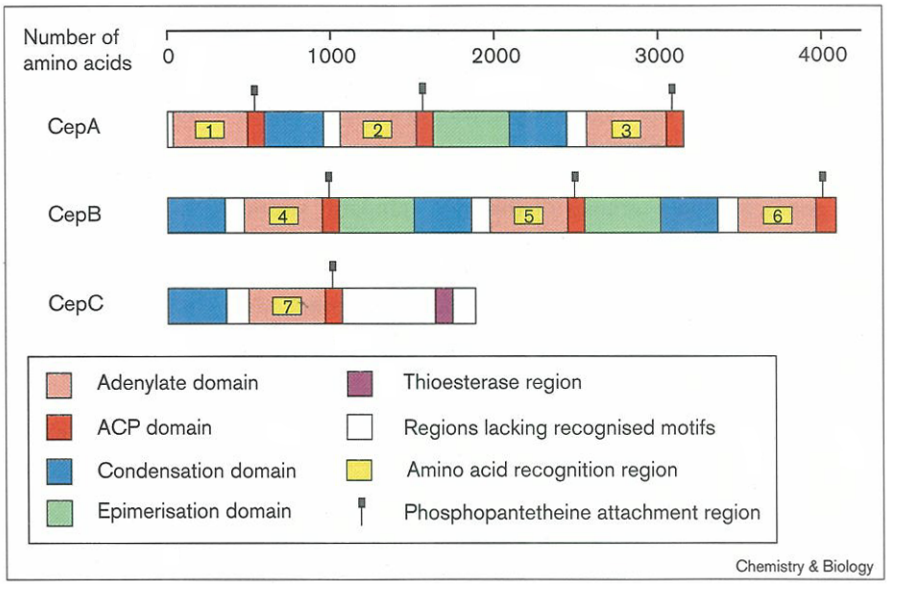
other peptide synthetases (for a summary see Kleinkauf and Von Döhren [10]), and homology searches were performed to allow the different functional domains to be identified. The results of this analysis are shown in Figure 3.

It can be seen that there are seven modules of similar functional organisation (three in CepA, three in CepB and one in CepC), as would be expected for a set of multienzymes responsible for assembling a heptapeptide. This organisation closely resembles that of previously discovered peptide synthetases such as those for surfactin (SrfA1, SrfA2 and SrfA3) and gramicidin (GrsA and GrsB) [12,16,17].

Four of the seven constituent amino acids (1, 2, 4 and 5) of chloroeremomycin have the *R*-configuration, so it might

Figure 3

Schematic representation of the functional organisation of the three peptide synthetases, CepA, CepB and CepC. The numbers indicate sequentially the amino-acid-activating domains on each peptide synthetase.



be expected that four domains with epimerisation character could be identified in the three peptide synthetases. This is the case for modules 2, 4 and 5, but not for module 1 (Figure 3), suggesting the possibility that, in fact, it is *R*-leucine that is accepted by CepA.

A further facet of the organisation of these peptide synthetases is the unusual nature of the transfer of the growing peptide chain between the individual multienzymes. For both the transfer of tripeptide between CepA and CepB, and hexapeptide between CepB and CepC, the carboxy-terminal amino acid of the peptide being transferred has the *S*-configuration. This contrasts with all but one of the previously characterised peptide synthetase systems, such as those for surfactin and gramicidin [12,16,17], where interenzymatic transfers occur with peptides in which the carboxy-terminal amino acid has the *R*-configuration. The only other example of the unusual *S*-peptide transfer is provided by the recently characterised bacitracin peptide synthetases from the *bac* operon of *Bacillus licheniformis* [18]. The internal epimerisation domains of CepA and CepB are also a feature shared with these multienzymes.

It was proposed by Cosmina *et al.* [16] that an ~180 amino acid region in the adenylation domain of each module is responsible for specific amino acid recognition. For each of the seven modules, these regions were compared both with each other, and with corresponding regions in other peptide synthetases. Significant results from these comparisons are shown in Table 1.

The evidence provided by these comparisons, taken in conjunction with the feeding experiments mentioned previously [7,8], allows us to infer that the amino acids used for residues 2 to 7 are, respectively, *S*-tyrosine, *S*-asparagine, *S*-4-hydroxyphenylglycine, *S*-4-hydroxyphenylglycine, *S*-tyrosine and *S*-3,5-dihydroxyphenylglycine. Interestingly, the region that recognises the amino-terminal acid does not show good homology with other *S*-leucine regions. This could be because the region recognises *R*-leucine rather than *S*-leucine, as proposed above.

The genes for the synthesis of the 'non-standard' amino acids are not found on the two cosmids. The finding that ORF1, a prephenate dehydrogenase, is at the start of pCZA363, however, might indicate that these genes are clustered to the left of the 5′ end of cosmid pCZA363. We propose that the introduction of the β-hydroxyl groups on the tyrosine residues, and chlorination and cross-linking of the different amino acid sidechains occurs after assembly of the peptide backbone.

Oxidative reactions

To convert the linear heptapeptide derived as above to the unglycosylated chloroeremomycin nucleus requires several oxidative processes. These include the introduction of β-hydroxyl groups onto tyrosine residues 2 and 6 and coupling of rings 5 and 7, rings 4 and 6, and rings 4 and 2. Furthermore, it is likely that the chlorine atoms on rings 2 and 6 are introduced via oxidative processes, that is using a haloperoxidase. Together, these processes should require at least seven enzymes.

Analysis of the enzymes found on the two cosmids shows that there are eight enzymes (ORFs 7, 8, 9, 10, 14, 18, 20 and 21) that might be involved in these oxidations. Four of these, ORF7, ORF8, ORF9 and ORF20, are P450-related enzymes, and contain an iron-heme binding motif (FGHGxHxCLG). They show identity scores ranging from 26% to 34% compared to similar enzymes found, for example, in *Streptomyces hygroscopicus* [19] or *Bacillus subtilis* [20]. Two ORFs (10 and 18) show homology to non-heme haloperoxidases. ORFs 9 and 14 were previously identified as putative hydroxylation enzymes [9]. Thus, the identified enzymes seem to be sufficient to carry out the expected oxidations, although there appears to be one enzyme present that is in addition to the requirements

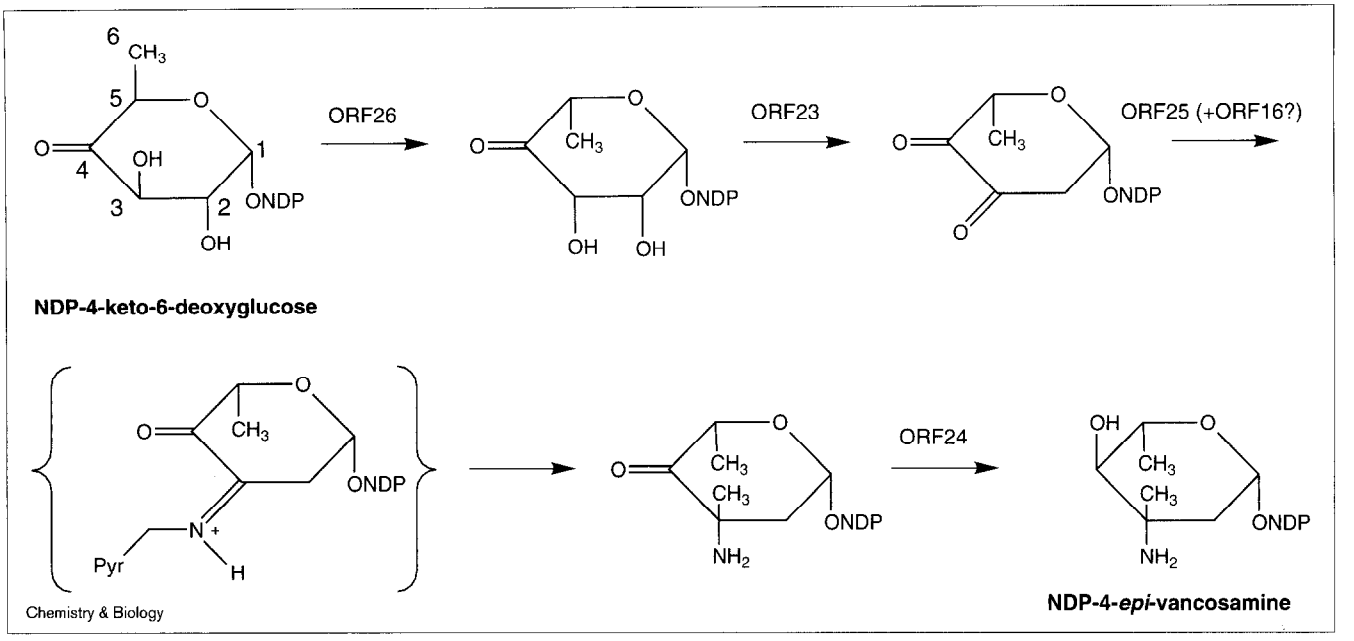
Table 1

Significant homologies for the seven amino-acid-recognition domains present in the three peptide synthetases.

Amino acid	Significant homologies (% similarity,% identity)
1 : N-Me-*R*-Leu	Amino acid 5 recognition (69%, 61%)* Pristinamycin PG recognition (58%, 50%) Pristinamycin Abu recognition (54%, 49%)
2 : *R*-CH-Tyr	Amino acid 6 recognition (63%, 57%) Virginiamycin Phe recognition (52%, 48%) Pristinamycin DMPA recognition (48%, 43%)
3 : *S*-Asn	Surfactin Asp recognition (49%,37%)
4 : *R*-HPG	Amino acid 5 recognition (73%, 69%) Amino acid 7 recognition (64%, 58%) Pristinamycin PG recognition (53%, 47%)
5 : *R*-HPG	Amino acid 4 recognition (73%, 69%) Amino acid 7 recognition (66%, 61%) Pristinamycin PG recognition (51%, 43%)
6 : *S*-CH-Tyr	Amino acid 2 recognition (63%, 57%) Virginiamycin Phe recognition (50%, 45%) Pristinamycin DMPA recognition (46%, 40%)
7 : *S*-DHPG	Amino acid 5 recognition (66%, 61%) Amino acid 4 recognition (64%, 58%) Pristinamycin PG recognition (51%, 43%)

% Similarities and % identities are over the whole region between the motifs C and E of Kleinkauf and Von Döhren [10] in all cases and were obtained using the GAP program of the GCG package. Where homologies are described in terms of 'Amino acid' and a number, these refer to the constituent amino acids of chloroeremomycin. *Amino acid 1 recognition domain also shows significant homologies (~60% similarity) with domains recognising amino acids 4 and 7. Abu, aminobutyric acid; CH-Tyr, *m*-chloro-β-hydroxytyrosine, DMPA, α–*N*-methyl-4-dimethylaminophenylalanine, PG, phenylglycine; HPG, 4-hydroxyphenylglycine; DHPG, 3,5-dihydroxyphenylglycine.

Figure 4



Postulated route for the biosynthesis of NDP-4-*epi*-vancosamine starting from NDP-4-keto-6-deoxyglucose, with the corresponding putative enzymes involved. The putative involvement of the pyridoxamine cofactor is shown in brackets.

identified above. The exact function of each individual enzyme remains to be elucidated.

Chlorination

Chloroeremomycin contains chlorine atoms on residues 2 and 6. It is most likely that there are two enzymes responsible for the chlorination, because eremomycin (also known as MM45289) contains a chlorine atom on residue 2 only [21]. Indeed, two enzymes located on cosmid pCZA361 show homology to non-heme haloperoxidases. ORF18 shows identity scores ranging from 24% to 29% compared to known haloperoxidases found in *Streptomyces lividans* [22,23]. Only the first 150 amino acids of ORF10 show homology with a known haloperoxidase found in *Pseudomonas fluorescens* (identity score 20%) [24]. The relatively low homologies may reflect the special character of these enzymes.

Glycosyl transferases

Once the cyclised and oxidised peptide core is available, two molecules of 4-*epi*-vancosamine and one of glucose must be added to it at specific sites. This process should require three enzymes and indeed three glycosyl transferases are located on pCZA361 (available under accession number U84349), annotated as GtfA (ORF11), GtfB (ORF12), and GtfC (ORF13) [9]. Furthermore, homology was found between these three glycosyl transferases and two glycosyl transferases involved in vancomycin biosynthesis [9]. Expression of one these genes (*gtfB*) showed that GtfB is capable of introducing glucose onto a heptapeptide core [9].

*Biosynthesis of 4-*epi*-vancosamine*

Inspection of the gene clusters of other secondary metabolites such as daunorubicin and erythromycin [25,26] suggested that the genes that code for the biosynthesis of the unusual 4-*epi*-vancosamine sugar (4-e-V) should be located in the same region as other biosynthetic genes, and this did indeed prove to be the case. On the basis of analogy to the biosynthesis of other 6-deoxysugars [26–28], the biosynthesis should proceed through an NDP-4-keto-6-deoxyglucose intermediate. Enzymes for the production of this key intermediate were not identified in the cluster, but it is possible that general cellular pools of this precursor are used for the biosynthesis of 4-e-V, an idea recently postulated by Summers *et al.* [26] for the biosynthesis of the mycarose and desosamine sugars of erythromycin.

Given NDP-4-keto-6-deoxyglucose as a precursor, homologies to enzymes identified in the biosynthetic pathways for daunosamine and mycarose [26,28,29] allowed a probable route to 4-e-V to be deduced (Figure 4). Four putative enzymes were identified. ORF26 was assigned as a 3,5-epimerase on the basis of its strong homology to DnmU (60% identity) of the daunosamine pathway of *Streptomyces peucetius* [28] and EryBVII (55% identity) of the mycarose pathway of *Saccharopolyspora erythraea* [26]. ORF23 is closely related to DnmT (49% identity) from the daunosamine pathway of *S. peucetius* [29] and EryBVI (45% identity) from the mycarose pathway of *S. erythraea* [26], both of which have been linked with C2 deoxygenation.

Table 2

Summary of genes identified on cosmids PCZA361 and PCZA363 obtained from *Amycolatopsis orientalis*.

ORF (name)	Position in base pairs (cosmid / start..stop)	Size (amino acids)	Proposed function /homology	Accession numbers of selected homologous proteins	Typical homology (% similarity, % identity)
1	363/255..1358	367	Prephenate dehydrogenase	P20692, D90910	(36%, 24%)
2	363/1562..3514	650	Mdr/ABC transporter	P45861, U78609	(45%, 32%)
3 (CepA)	363/3511..12987	3158	Peptide synthetase†	X98690, S46968	(41%, 32%)
4 (Cep B)	363/13162..25395	4077	Peptide synthetase†	X98690, I40485	(41%, 32%)
5 (Cep C)	363/25415..30997	1860	Peptide synthetase†	X98690, Y11547	(43%, 36%)
6	363/31033..31203	56	Hypothetical protein in *M. tuberculosis*	Z95208	(86%,74%)
7	363/31325..32500	391	P450-related oxidase‡	P46373, P23296	(39%, 30%)
8	363/32673..33740	355	P450-related oxidase‡	P46373, S15809	(47%, 39%)
9	363/33791..35143	450	P450-related oxidase‡	P46373, P18327	(47%, 38%)
10	361/1076..2551	491	Non-heme halogenase	U74493	(30%, 21%)
11 (GtfA)	361/2596..3786	396	Glycosyl transferase	U84350, U77459	(75%, 63%)
12 (GtfB)	361/3821..5044	407	Glycosyl transferase	U84350, X81885	(88%, 82%)
13 (GtfC)	361/5193..6422	409	Glycosyl transferase	U84350, U77459	(75%, 69%)
14	361/6495..7721	408	Hydroxylase	S39963, S18530	(56%, 48%)
15	361/7732..8556	275	?		
16	361/8589..9431	281	Methyl transferase	U77549, U21300	(31%, 27%)
17*	361/9400..10935	511	Amino transferase	Y08257	(44%, 32%)
18	361/10934..11764	276	Non-heme halogenase	P29715, AF031242	(32%, 26%)
19	361/11761..13503	580	Peptide synthetase†	X98690, Y11547	(52%,47%)
20	361/13519..14712	397	P450-related oxidase‡	P53554, S49051	(35%, 27%)
21	361/14957..16030	357	Hydroxyphenyl Pyruvate dioxygenase	U11864, S32821	(43%,34%)
22	361/16027..17100	357	Glycolate oxidase	D14044, U80071	(50%, 40%)
23	361/17100..18515	471	2,3 Dehydratase	U77891, U77459	(58%, 49%)
24	361/18517..19494	325	4-Ketoreductase	U77459	(56%, 49%)
25	361/19491..20600	369	Sugar biosynthesis	P25048	(77%, 72%)
26	361/20624..21241	205	3,5-Epimerase	AF006631, U77459	(69%, 60%)
27	361/21434..22552	372	Chalcone synthase	P54157, X91340	(37%, 28%)
28	361/22552..23352	217	?		
29	361/23511..24647	378	Carnitine racemase	P31551	(39%, 30%)
30	361/24644..25447	267	Carnitine racemase	P31551	(52%, 41%)
31	361/25569..26657	362	Phospho-2-dehydro-3-Deoxyheptonate aldolase	P44303, Q09755	(57%,48%)
32*	361/26664..28190	508	?		
33	361/28418..32428	1336	?		
34	361/32572..32874	100	?		
35	361/32901..33746	281	?		
36*	361/33785..34333	183	?		
37*	361/34355..35743	462	?		
38*	361/35740..37227	495	Endo/exo gluconase precusor	P54583, P50899	(24%, 20%)
39*	361/37530..37928	132	Sugar transporter	D90910, D90769	(50%, 29%)

*Gene on opposite strand. †Contains these motifs from the PROSITE dictionary of protein sites and patterns: YTSGSTGXPK (AMP binding) ELGGDSIXMXXXARA (4′-phosphopantetheine attachment site).

‡Contains this motif from the PROSITE dictionary of protein sites and patterns: FGHGXHXCLG (cytochrome P450 cysteine heme-iron ligand signature).

In particular, EryBVI was assigned as a 2,3-dehydratase and a similar role is proposed for ORF23. ORF25 has impressive homology (72% identity) to DnrJ, an enzyme found in daunosamine biosynthesis and implicated in the introduction of the amine group [25,27], probably with the aid of pyridoxamine as a cofactor. At this stage, it remains unclear whether the methyl group is introduced using ORF25 or if another enzyme is necessary. ORF16 shows some homology with known methyl transferases, and might therefore be involved in this step. The proposed biosynthesis is completed by C-4 ketone reduction with ORF24, which has strong homology only to EryBIV (49% identity), the 4-ketoreductase of the mycarose pathway of *S. erythraea* [26].

The remaining proteins

There is no clearly defined role for the remaining putative proteins in the biosynthesis of chloroeremomycin. The Supplementary material includes comments on the relevant ORFs.

The results from the analysis of the two cosmids pCZA361 and pCZA363 are summarized in Table 2. The enzymes identified on cosmids pCZA361 and pCZA363 seem sufficient to account for all, or most, of the steps involved in the biosynthesis of chloroeremomycin, given the provision of activated glucose and the required amino acids.

Significance

We have sequenced two overlapping cosmids spanning a total of 72 kb of genomic DNA originating from *Amycolatopsis orientalis*, the organism responsible for the production of the glycopeptide antibiotic chloroeremomycin. Analysis of the sequence led to the identification of 39 putative genes, including those coding for enzymes responsible for the biosynthesis of the antibiotic. The enzymes responsible for the biosynthesis of the aryl ethers, the biaryl linkage, β-hydroxylation of tyrosine residues, and introduction of the chlorine substituents are probably coded for within the 72 kb that have been sequenced. These enzymes are of special interest because their mechanisms are poorly understood. Expression of selected genes should give enzymes that could be used for the modification of glycopeptides, leading to new drugs to combat the increasing antibiotic resistance in bacteria.

Materials and methods

Preparation and isolation of the two cosmids pCZA361 and pCZA363 (both in *Escherichia coli* S17-1) was described by Solenberg *et al.* [9]. An overlap of 5 kb between the two cosmids was confirmed by the sequence data, showing that the two cosmids cover 72 kb of contiguous DNA.

Sequencing

Individual colonies from the two clones (pCZA361 and pCZA363) were inoculated into 2 × TY medium with apramycin (50 μg/ml) and grown to saturation. DNA was prepared from these cultures by alkaline lysis [30], followed by CsCl/ethidium bromide density gradient ultracentrifugation. Approximately 15 μg of DNA were sonicated and fragments were blunt-ended with mung bean nuclease. Fragments in the size range 1.4 to 2 kb were isolated from agarose gels using Prep-A-Gene (Biorad). An aliquot of the isolated fragments was ligated into pUC18 cut with *Sma*1 and dephosphorylated. Fractions of the ligation were transformed into *E. coli* TG1 cells by electroporation. Recombinants were selected by plating onto medium containing ampicillin (100 μg/ml), IPTG and *Xgal*, and were prepared for sequencing.

The sequencing strategy involved two phases. Approximately 400 random pUC18 subclones were selected from each cosmid for sequencing. Double-stranded DNA was prepared using an alkaline lysis method (www.sanger.ac.uk.). Both ends of the DNA inserts were sequenced using dye-terminator chemistry, yielding on average 400 bases of high quality data from each end. Approximately 600 good sequencing reads were collected from each cosmid and the random reads assembled automatically with the program PHRAP (P. Green, unpublished observations). In the second phase of sequencing, each assembly was subjected to human review to produce continuity of data and high accuracy (> 99.99%) At this point, discrepancies between different sequence reads over the same region were resolved and additional reads were obtained in a directed manner to recover sequence from both strands throughout the cosmids.

Sequence analysis

DNA sequence analysis was carried out within the environment of the ACEDB database (R. Durbin and J. Thierry-Mieg, A *C. elegans* Database. Documentation, code and data available from anonymous ftp servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk, and ncbi.nlm.nih.gov.). Protein coding sequences were detected by protein similarities determined by BLASTX [31]. In the absence of protein similarities coding elements were detected using hexamer coding statistics [32] using the program HEXEXON (R. Durbin, unpublished observations). Further analysis and comparison, including similarity and identity scores (using GAP), were carried out using the Wisconsin GCG package. Databases used include SWISS protein and TREMBL.

Accession numbers

The cosmid sequences have been deposited in the EMBL nucleotide databank, and are available under accession numbers AJ223998 (pCZA361) and AJ223999 (pCZA363).

Acknowledgements

A.v.W. would like to thank Dutch Technology Foundation (STW), and P.K. would like to thank the EPSRC for financial support. We would like to thank Dr. Matt C. Cooper and Team 14 at the Sanger Centre for their useful help and discussions.

References

1. Foldes, M., Munro, R., Sorrell, T.C., Shankar, S. & Toohey, M. (1983). *In vitro* effects of vancomycin, rifampicin, and fusidic acid, alone and in combination, against methicillin-resistant *Staphylococcus aureus*. *J. Antimicrob. Chemother.* **11**, 21–26.
2. Courvalin, P. (1990). Resistance of enterococci to glycopeptides. *Antimicrob. Agents Chemother.* **34**, 2291–2296.
3. Neu, H.C. (1992). The crisis in antibiotic resistance. *Science* **257**, 1064–1073.
4. Walsh, C.T., Fisher, S.L., Park, I.-S., Prahalad, M. & Wu, Z. (1996). Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story. *Chem. Biol.* **3**, 21–28.
5. Williams, D.H. (1996). The glycopeptide story - how to kill deadly superbugs. *Natl Prod. Rep.* **13**, 469.
6. Allen, N.E., LeTourneau, D.E. & Hobbs, J.N. (1997). The role of hydrophobic side chains as determinants of the antibacterial activity of semi-synthetic glycopeptide antibiotics. *J. Antibiot.* **50**, 677–684.
7. Hammond, S.J., Williamson, M.P., Williams, D.H., Boeck, L.D. & Marconi, G.G. (1982). On the biosynthesis of the antibiotic vancomycin. *J. Chem. Soc. Chem. Comm.* 344–346.
8. Hammond, S.J., Williams, D.H. & Nielsen, R.J. (1983). The biosynthesis of ristocetin. *J. Chem. Soc. Chem. Comm.* 116–117.
9. Solenberg, P.J., Matsushima, P., Stack, D.R., Wilkie, S.C., Thompson, R.C. & Baltz, R.H. (1997). Production of hybrid glycopeptide antibiotics *in vitro* and in *Streptomyces toyocaensis*. *Chem. Biol.* **4**, 195–202.
10. Kleinkauf, H. & Von Döhren, H. (1996). A nonribosomal system of peptide biosynthesis. *Eur. J. Biochem.* **236**, 335–351.
11. Stachelhaus, T. & Marahiel, M.A. (1995). Modular structure of genes encoding multifunctional peptide synthetases required for nonribosomal peptide synthesis. *FEMS Lett.* **125**, 3–14.
12. Turgay, K., Krause, M. & Marahiel, M.A. (1992). 4 homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol. Microbiol.* **6**, 529–546.
13. Lipmann, F. (1973). Attempts to map a process evolution of peptide biosynthesis. *Science* **173**, 875–884.
14. Stein, T., *et al.*, & Morris, H.R. (1996). The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *J. Biol. Chem.* **271**, 15428–15435.
15. Schlumbohm, W., *et al.*, & Wittmanliebold, B. (1991). An active serine is involved in covalent substrate amino-acid binding at each reaction center of gramicidin-S synthetase. *J. Biol. Chem.* **266**, 23135–23141.
16. Cosmina, P., *et al.*, & Van Sinderen, D. (1993). Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol. Microbiol.* **8**, 821–831.

17. Kratzschmar, J., Krause, M. & Marahiel, M.A. (1989). Gramicidin-S biosynthesis operon containing the structural genes GrsA and GrsB Has an open reading frame encoding a protein homologous to fatty-acid thioesterases. *J. Bacteriol.* **171**, 5422–5429.
18. Konz, D., Klens, A., Schörgendorfer, K. & Marahiel, M.A. (1997). The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem. Biol.* **4**, 927–937.
19. Schwecke, T., *et al.,* & Leadlay, P.F. (1995). The biosynthetic gene-cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl Acad. Sci.* **92**, 7839–7844.
20. Bower, S., Perkins, J.B., Yom, R.R., Howitt, C.L., Rahaim, P. & Pero, J. (1996). Cloning, sequencing and characterization of the *Bacillus subtilis* biotin biosynthetic operon. *J. Bacteriol.* **78**, 4122–4130.
21. Good, V.M., Gwynn, M.N. & Knowles, D.J.C. (1990). MM 45289, a potent glycopeptide which interacts weakly with diacetyl-L-lysyl-D-alanyl-D-alanine. *J. Antibiot.* **43**, 550–555.
22. Pfeiffer, O., Pelletier, I., Altenbuchner, J. & van Pee, K.-H. (1992). Molecular cloning and sequencing of a non-heme bromoperoxidase gene from *Streptomyces lividans*. *J. Gen. Microbiol.* **138**, 1123–1131.
23. Bantleon, R., Altenbuchner, J. & van Pee, K.-H. (1994). Chloroperoxidase from *Streptomyces lividans* - isolation and characterization of the enzyme and the corresponding genes. *J. Bacteriol.* **179**, 2339–2347.
24. Hammer, P.E., Hill, D.S., Lam, F.S. & van Pee, K.-H. (1997). Four genes from *Pseudomonas fluorescens* that encode for the biosynthesis of pyrrolnitrin. *Appl. Environ. Microbiol.* **63**, 2147–2154.
25. Madduri, K. & Hutchinson, C.R. (1995). Functional characterization and transcriptional analysis of the *dnrR*1 locus, which controls daunorubicin biosynthesis in *Streptomyces peucetius*. *J. Bacteriol.* **177**, 1208–1215.
26. Summers, G.R., Donadio, S., Staver, M.J., Wendt-Pienkowski, E., Hutchinson, R. & Katz, L. (1997). Sequencing and mutagenesis of genes from the erythromycin biosynthetic cluster of *Saccharopolyspora erythrea* that are involved in L-mycarose and D-desosamine production. *Microbiology* **143**, 3251–3262.
27. Thorson, J.S., Lo, S.F. & Liu, H. (1993). Biosynthesis of 3,6-dideoxyhexoses: new mechanistic reflections upon 2,6-dideoxy, 4,6-dideoxy, and amino sugar construction. *J. Am. Chem. Soc.* **115**, 6993–6994.
28. Otten, S.L., Gallo, M.A., Madduri, K., Liu, X. & Hutchinson, R. (1997). Cloning and characterization of the *Streptomyces peucetius* dnmZUV genes encoding three enzymes required for the biosynthesis of daunorubicin precursor thymidine diphospho-L-daunosamine. *J. Bacteriol.* **179**, 4446.
29. Scotti, C. & Hutchinson, C.R. (1996). Enhanced antibiotic production by manipulation of the *Streptomyces peucetius* dnrH and dnrT genes involved in doxorubicin (adriamycin) biosynthesis). *J. Bacteriol.* **178**, 716–7321.
30. Birnboim, H.C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513–1523.
31. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment tool. *J. Mol. Biol.* **215**, 403.
32. Fickett, J.W. (1994). Inferring genes from open reading frames. *Comput. Chem.* **18**, 203–205.

Because *Chemistry & Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cmb – for further information, see the explanation on the contents pages.