Organization of the biosynthetic gene cluster for the macrolide concanamycin A in Streptomyces neyagawaensis ATCC 27449

Stephen F. Haydock, Anthony N. Appleyard, Tatiana Mironenko, John Lester, Natasha Scott and Peter F. Leadlay

Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK

The macrolide antibiotic concanamycin A has been identified as an exceptionally potent inhibitor of the vacuolar (V-type) ATPase. Such compounds have been mooted as the basis of a potential drug treatment for osteoporosis, since the V-ATPase is involved in the osteoclast-mediated bone resorption that underlies this common condition. To enable combinatorial engineering of altered concanamycins, the biosynthetic gene cluster governing the biosynthesis of concanamycin A has been cloned from Streptomyces neyagawaensis and shown to span a region of over 100 kbp of contiguous DNA. An efficient transformation system has been developed for S. neyagawaensis and used to demonstrate the role of the cloned locus in the formation of concanamycin A. Sequence analysis of the 28 ORFs in the region has revealed key features of the biosynthetic pathway, in particular the biosynthetic origin of portions of the backbone, which arise from the unusual polyketide building blocks ethylmalonyl-CoA and methoxymalonyl-ACP, and the origin of the pendant deoxysugar moiety 4-O-carbamoyl-2-deoxyrhamnose, as well as the presence of a modular polyketide synthase (PKS) encoded by six giant ORFs. Examination of the methoxymalonyl-specific acyltransferase (AT) domains has led to recognition of an amino acid sequence motif which can be used to distinguish methylmalonyl-CoA- from methoxymalonyl-ACP-specific AT domains in natural PKSs.

INTRODUCTION

Polyketide macrolides constitute a large family of bioactive natural products, including many compounds with important clinical applications in both human and veterinary medicine. Concanamycin A (folimycin) is an unusual macrolide, first isolated from Streptomyces diastatochromogenes S-45 (Kinashi et al., 1984), whose absolute configuration was established as shown in Fig. 1 (Westley et al., 1984). Concanamycin A and closely related compounds (concanamycins B–F) are characterized by an 18-membered tetraenic macrolide ring incorporating a methyl enol ether, and by a β-hydroxyhemiacetal side chain. They are closely related to the 16-membered macrolides called bafilomycins (Goetz et al., 1985) and also share structural features with the antifungal macrolide, elaiophylin A (Arcamone et al., 1959) whose biosynthetic pathway we recently elucidated (Haydock et al., 2004). These unusual macrolides have been dubbed the ‘plecomacroclides’ (from the Greek word plectos, describing their unusually folded side chain) (Bindseil & Zeeck, 1994).

The concanamycins exhibit a wide range of important biological activities: antiviral (Omura et al., 1983), antiprotozoal (Omura et al., 1983) and antineoplastic (Seki-Asano et al., 1994). These activities are thought to be due, at least in part, to their potent inhibition of the vacuolar (V-type) H⁺-ATPase (Muroi et al., 1993) mediated by their binding to the membrane-spanning Vₒ component of the ATPase (Drose et al., 2001). Interest in these compounds has increased recently with the recognition of the importance of the V-type ATPase in osteoclast-mediated resorption of bone (Laitala-Leinonen et al., 1999). Overactivity of osteoclasts underlies osteoporosis, a degenerative bone-thinning that affects over 200 million people worldwide. Synthetic mimics of the concanamycin molecule have been shown to inhibit the V-type ATPase in vitro (Drose et al., 2001), and such studies have led to the identification of the pharmacophoric portion of the concanamycin structure that is involved in V-ATPase inhibition. An intact closed macrocycle containing a diene functionality appears to be particularly important. Unfortunately, concanamycin and...
related compounds show significant toxicity due to inhibition of the V-ATPase in other tissues. The genetic engineering of the biosynthetic gene cluster would offer an independent and complementary route to total chemical synthesis, as a way of developing more selective inhibitors. Until now the biosynthetic pathway to concanamycin has remained completely obscure, although the involvement of a modular polyketide synthase (PKS), as for other macrolides, would be anticipated (Staunton & Weissman, 2001). Even the biosynthetic origin of certain portions of the polyketide chain, presumed to arise from use of the unusual building blocks ethylmalonyl-CoA and methoxymalonyl-ACP, could not be confirmed by feeding labelled precursors (Grond & Schuhmann, 2004). To elucidate the unusual features of concanamycin biosynthesis and pave the way to biosynthetic engineering of novel derivatives, we have therefore undertaken the isolation and characterization of the gene cluster for the biosynthesis of concanamycin A. An unanticipated discovery from this analysis was the identification of an amino acid sequence motif which clearly distinguishes acyltransferase (AT) domains of a modular PKS which recruit methylmalonate extender units from those which recruit exclusively methoxymalonate extender units, which may assist in the direct reengineering of AT active sites in PKSs (Del Vecchio et al., 2003). We also present here techniques for convenient genetic manipulation of \textit{S. neyagawaensis} and demonstrate that disruption of the cloned locus abolishes concanamycin biosynthesis.

**METHODS**

**Bacterial strains and cloning vectors.** \textit{S. neyagawaensis} ATCC 27449 was obtained from the American Type Culture Collection (Manassas, VA, USA). The following \textit{Escherichia coli} strains were used: DH10B (Gibco-BRL) for routine cloning; XL-1 Blue MR for cosmid library construction; and ET12567 (MacNeil et al., 1992) into which the helper plasmid pUZ8002 (Paget et al., 1999) was cloned, as a donor strain for conjugation into \textit{S. neyagawaensis}. Insertional inactivation of the concanamycin cluster was carried out using a plasmid based on the pOJ260 vector (Bierman et al., 1992). Routine subcloning was performed in pSHG397 (Hashimoto-Gotoh et al., 1995).

**Culture conditions and molecular genetic procedures.** \textit{E. coli} strains were grown on 2 × TY medium supplemented with appropriate antibiotics. \textit{S. neyagawaensis} strains were maintained on tap water medium (Kieser et al., 2000). Liquid \textit{S. neyagawaensis} cultures for isolation of genomic DNA were grown in tryptone soya broth (Difco). Cultures of \textit{S. neyagawaensis} to be used in the analysis of concanamycin A production were grown in an aqueous medium containing 1% glucose, 3% soluble starch, 0.5% peptone, 1% soybean flour, 0.5% yeast extract and 0.2% calcium carbonate. Standard genetic techniques for \textit{E. coli} and for \textit{in vitro} DNA manipulations were as described by Sambrook \textit{et al}. (1989). Total DNA was isolated from \textit{S. neyagawaensis} by ‘procedure B’ of Kieser \textit{et al}. (2000). For the generation of the cosmid library, total DNA was partially digested with Sau3A, dephosphorylated with shrimp alkaline phosphatase, ligated directly into pSuperCos and packaged with Gigapack Gold packaging extract (Stratagene) without size fractionation. All procedures were in accordance with the manufacturer’s recommendations. Cosmid DNA for complete sequencing was prepared using Qiagen midi-prep DNA purification kits from 50 ml liquid cultures in 2 × LB medium containing ampicillin and kanamycin. The complete sequence of each cosmid was obtained from overlapping Sau3A fragments obtained after a partial digestion of the parent cosmid. The 2–5 kbp fraction of this digest was eluted from an agarose gel using the Gene Clean kit (Bio101, Bio-Rad) and subcloned into pSG397 (Hashimoto-Gotoh et al., 1995). Colony hybridization was performed with a digoxigenin-labelled probe in accordance with the manufacturer’s recommendations (Boehringer Mannheim). For gene disruption by insertional activation, plasmid pOJconE was constructed by cloning a BamHI fragment of CosI6b internal to the \textit{conE} gene (part of the putative concanamycin modular PKS), into the unique BamHI site of pOJ260. This plasmid was transferred by conjugation from \textit{E. coli} into \textit{S. neyagawaensis} by a previously described protocol (Kieser \textit{et al}. , 2000). Conjugation took place on agar plates containing A medium (Kieser \textit{et al}. , 2000) incubated overnight at 30 °C and overlaid after 12 h with nalidixic acid and apramycin. Potential exconjugants were transferred to fresh apramycin and nalidixic acid plates after a further 5 days’ growth. Apramycin-resistant colonies were isolated and shown by Southern hybridization (data not shown) to contain the pOJconE plasmid integrated into the predicted genomic location by single crossover recombination within the \textit{conE} ORF.

**DNA sequencing analysis.** Automated DNA sequencing was carried out on double-stranded DNA templates by the dideoxynucleotide chain-termination method in the University of Cambridge Department of Biochemistry DNA sequencing facility. An Applied Biosystems 800 molecular biology CATALYST robot was used to apply Taq dideoxy terminator sequencing reactions (Big Dye Terminator kit; ABI) to an ABI 373A sequencer, according to the manufacturer’s protocols. Cosmid DNA for initial end sequencing was prepared from overnight growth of 1 ml cultures in 96-well plates in 2 × LB medium containing kanamycin and apramycin. The complete sequence of the cosmid was obtained from analysis of overlapping Sau3A fragments obtained via a partial digestion of the parent cosmid. The 2–5 kbp size fraction of this digest was eluted from an agarose gel using the Gene Clean kit (Bio101, Bio-Rad).
and subcloned into the BamHI site of dephosphorylated pSG397 (Hashimoto-Gotoh et al., 1995). These random subclones were sequenced using template DNA obtained from 1 ml cultures grown overnight at 37 °C in 96-well plates, using the universal forward and reverse sequencing primers. SeqEd v 1.03 was used for sequence editing. Database searches used the BLAST algorithm. Sequence assembly employed GAP (Genome Assembly Program) version 4.2. (Bonfield et al., 1995). To fill remaining contig gaps and resolve any remaining ambiguities, custom-designed oligonucleotide primers were used to sequence from selected subclones or from the parent cosmId. Both strands were completely sequenced with a mean eightfold coverage per base.

**Extraction, isolation and characterization of concanamycin A.** Cultures of wild-type S. neyagawaensis were grown in the aqueous medium described previously. A 100 ml seed culture of wild-type S. neyagawaensis was grown in a 500 ml Erlenmeyer flask for 4 days at 30 °C and 250 r.p.m. in an orbital shaker. The seed cultures were used to provide a 3% inoculum in 2 l Erlenmeyer flasks that contained 300 ml of the same medium. The cultures were grown for a further 4 days under the same conditions and the cells were then harvested by centrifugation.

Cells of S. neyagawaensis collected from 3-6 l of culture were extracted in acetone (2 × 500 ml) and the acetone extracts were concentrated by evaporation to a volume of 400 ml. The acetone solution was then extracted with an equal volume of dichloromethane. The organic extract was dried with ammonium sulphate and then evaporated to afford a brown oil. The crude extract was analysed by LC-MS and LC-MS/MS. Concanamycin A was detected in the extracts from wild-type S. neyagawaensis, a component with a retention time of 14 min and a molecular ion of mass 888.7 (M + Na+). These results were identical to those obtained for an authentic sample of concanamycin A (Sigma) analysed in the same manner. The fragmentation pattern obtained for the selected 888.7 ion in LC-MS/MS analysis was also identical in the crude extract and standard concanamycin A samples. A mutant of S. neyagawaensis obtained by chromosomal integration of a pOJ260-derived construct into the putative concanamycin A locus (described above) was grown under identical conditions and the culture was harvested. The harvested cells from 3-6 l of culture were extracted and analysed using conditions identical to those applied to the cells of wild-type S. neyagawaensis. Concanamycin A was not detected by LC-MS in extracts from cells of the S. neyagawaensis mutant.

**RESULTS AND DISCUSSION**

**Cloning and sequencing of the concanamycin biosynthetic cluster**

Examination of the structure of concanamycin A and comparison with polyketide biosynthetic gene clusters encoding macrolides of a similar size led us to predict that approximately 100 kbp of the S. neyagawaensis genome would be required to encode the biosynthesis of this antibiotic. Information from the sequencing of Streptomyces coelicolor (Bentley et al., 2002), Streptomyces avermitilis (Omura et al., 2001) and Saccharopolyspora erythraea (M. Oliynyk & M. Samborsky, personal communication) suggests a total genome size of 8-9 Mb, so that the concanamycin A cluster would account for over 1% of the genome of the producer organism.

A total of 480 cosmid-bearing colonies, from a cosmid library of S. neyagawaensis DNA, housed in the vector pSuperCos, were therefore grown up in 2 × LB medium in 96-well plates. Replica deeps were prepared in further 96-well plates by adding 100 μl of each culture to 100 μl 40 % glycerol. The remaining colonies were used to prepare DNA sequencing template which were each end-sequenced using T3 and T7 primers. BLAST searches were performed on the sequence data and nine cosmids were identified that encoded either modular PKS genes or genes previously identified as components of biosynthetic gene clusters for complex macrolides. Large-scale cosmid DNA preparations were made from the glycerol deeps of these clones and partial sequence data were obtained for each of the nine clones by end sequencing of BamHI restriction fragments. Three clones (1B6, 2C7 and 4A8) between them were found to span a large portion of the same PKS locus whose organization appeared consistent with a role in the biosynthesis of concanamycin A, and this was confirmed by their subsequent complete sequencing. The other six clones initially identified by random sequencing did not belong to this locus. S. neyagawaensis appears to harbour genes for at least two other complex polyketides (S. F. Haydock, unpublished data). This provides further evidence of the unexpected biosynthetic potential of the streptomycetes (Omura et al., 2001; Haydock et al., 2004).

Further cosmids spanning the remainder of the putative concanamycin cluster were obtained by hybridization of the cosmid library with digoxigenin-labelled BamHI fragments from the end of the cosmid clones 1B6, 2C7 and 4A8. The complete sequencing of these overlapping cosmids clones led to the determination of the sequence of the entire concanamycin biosynthetic gene cluster and the deduced gene organization was found to be as shown in Fig. 2. In addition to the six large genes required for the synthesis of the macrolactone core, the sequenced gene cluster encoded 22 other ORFs. The inferred protein products can be subdivided into (a) sugar biosynthesis and attachment; (b) regulatory genes and genes of unknown function; (c) provision of extender units or modification of the PKS; and (d) indeterminate (Table 1).

**Genes for deoxysugar biosynthesis and attachment (ORFs 5*, 6*, 7*, 8*, 9*, 10* and 11*)**

The concanamycins are modified by attachments at the hydroxyl group of C23 of the hemiketal ring moiety. The substituent varies according to the specific concanamycin subtype A–F. In the case of concanamycin A this is the modified deoxysugar derivative 4′-O-carbamoyl-2′-deoxyrihamnose. The seven genes encoding the biosynthesis of this sugar and its attachment are grouped together, all are transcribed in the same direction and they may constitute an operon. ORF5* and ORF6* encode proteins with very high sequence homology to TDP-glucose 4,6-dehydratase and TDP glucose synthase, respectively. The stop codon of ORF9 overlaps the start codon of ORF8 and the ribosome-binding site of ORF8 lies within the coding sequence of ORF9, suggesting that these two genes may be co-transcribed. The initial steps in biosynthesis of the sugar are similar to
Table 1. Proposed function of deduced peptides of the identified ORFs of the concanamycin gene cluster

The deduced peptide translations of the identified ORFs of the concanamycin gene cluster were used to perform a BLAST sequence search using the UNIPROT protein sequence database. The accession number, percentage identity and species of origin of the best identified match are given together with the inferred role of the gene product of that ORF in concanamycin biosynthesis.

<table>
<thead>
<tr>
<th>ORF</th>
<th>Best match</th>
<th>Identity (%)</th>
<th>Species</th>
<th>Proposed function in concanamycin biosynthesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Q93HJ7</td>
<td>38</td>
<td><em>Streptomyces avermitilis</em></td>
<td>LuxR class regulator</td>
</tr>
<tr>
<td>2</td>
<td>Q9ZN75</td>
<td>37</td>
<td><em>Streptomyces griseus</em></td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>1</td>
<td>Q6D3P9</td>
<td>22</td>
<td><em>Erwinia carotovora</em></td>
<td>Malonate-specific AT</td>
</tr>
<tr>
<td>A</td>
<td>Q8CJN6</td>
<td>48</td>
<td><em>Streptomyces coelicolor</em></td>
<td>Type I PKS</td>
</tr>
<tr>
<td>B</td>
<td>Q6W5P6</td>
<td>42</td>
<td><em>Streptomyces sp. FR-008</em></td>
<td>Type I PKS</td>
</tr>
<tr>
<td>C</td>
<td>Q9EWA1</td>
<td>51</td>
<td><em>‘Streptomyces natalensis’</em></td>
<td>Type I PKS</td>
</tr>
<tr>
<td>D</td>
<td>Q5VKQ3</td>
<td>53</td>
<td><em>Saccharopolyspora erythraea</em></td>
<td>Type I PKS</td>
</tr>
<tr>
<td>E</td>
<td>Q6W5P6</td>
<td>47</td>
<td><em>Streptomyces sp. FR-008</em></td>
<td>Type I PKS</td>
</tr>
<tr>
<td>F</td>
<td>Q76KZ3</td>
<td>42</td>
<td><em>Streptomyces halstedii</em></td>
<td>Type I PKS</td>
</tr>
<tr>
<td>1*</td>
<td>Q84G15</td>
<td>61</td>
<td><em>Streptomyces hygroscopicus</em></td>
<td>Methoxymalonate biosynthesis (putative dehydrogenase)</td>
</tr>
<tr>
<td>2*</td>
<td>Q8KUG2</td>
<td>63</td>
<td><em>Actinosynnema pretiosum</em></td>
<td>Methoxymalonate biosynthesis (acyl carrier protein)</td>
</tr>
<tr>
<td>3*</td>
<td>Q9KIE5</td>
<td>60</td>
<td><em>Streptomyces hygroscopicus</em></td>
<td>Methoxymalonate biosynthesis (putative dehydrogenase)</td>
</tr>
<tr>
<td>4*</td>
<td>Q9KIE6</td>
<td>70</td>
<td><em>Streptomyces hygroscopicus</em></td>
<td>Methoxymalonate biosynthesis</td>
</tr>
<tr>
<td>5*</td>
<td>Q7WTE3</td>
<td>73</td>
<td><em>‘Streptomyces nanchangensis’</em></td>
<td>dTDP-glucose 4,6-dehydratase</td>
</tr>
<tr>
<td>6*</td>
<td>Q7WTE4</td>
<td>72</td>
<td><em>‘Streptomyces nanchangensis’</em></td>
<td>dTDP-glucose synthase</td>
</tr>
<tr>
<td>7*</td>
<td>Q84G19</td>
<td>60</td>
<td><em>Streptomyces hygroscopicus</em></td>
<td>Carbamoyltransferase</td>
</tr>
<tr>
<td>8*</td>
<td>Q9ZGC5</td>
<td>40</td>
<td><em>Saccharopolyspora spinosa</em></td>
<td>NDP-hexose 4-ketoreductase</td>
</tr>
<tr>
<td>9*</td>
<td>Q9ALN5</td>
<td>52</td>
<td><em>Saccharopolyspora spinosa</em></td>
<td>NDP-hexose 3-ketoreductase</td>
</tr>
<tr>
<td>10*</td>
<td>Q76KZ0</td>
<td>58</td>
<td><em>Streptomyces halstedii</em></td>
<td>NDP-hexose 2,3-dehydratase</td>
</tr>
<tr>
<td>11*</td>
<td>Q5SFA4</td>
<td>41</td>
<td><em>Streptomyces bikieniensis</em></td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>12*</td>
<td>Q9KIE7</td>
<td>53</td>
<td><em>Streptomyces hygroscopicus</em></td>
<td>Methoxymalonate biosynthesis (O-methyltransferase)</td>
</tr>
<tr>
<td>13*</td>
<td>Q73TL8</td>
<td>36</td>
<td><em>Mycobacterium paratuberculosis</em></td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>14*</td>
<td>Q82PG7</td>
<td>53</td>
<td><em>Streptomyces avermitilis</em></td>
<td>Type II thioesterase</td>
</tr>
<tr>
<td>15*</td>
<td>Q82MB1</td>
<td>69</td>
<td><em>Streptomyces avermitilis</em></td>
<td>Phosphoesterase</td>
</tr>
<tr>
<td>16*</td>
<td>Q6T709</td>
<td>52</td>
<td><em>Saccharopolyspora erythraea</em></td>
<td>Phosphopantetheinyl transferase</td>
</tr>
<tr>
<td>17*</td>
<td>Q9SOY6</td>
<td>33</td>
<td><em>Streptomyces griseus</em></td>
<td>AfsR family regulator</td>
</tr>
<tr>
<td>18*</td>
<td>Q9HI6</td>
<td>64</td>
<td><em>Streptomyces avermitilis</em></td>
<td>Ethylmalonate biosynthesis (crotonyl CoA reductase)</td>
</tr>
<tr>
<td>19*</td>
<td>Q5YP10</td>
<td>54</td>
<td><em>Nocardia farcinica</em></td>
<td>Ethylmalonate biosynthesis (hydroxybutyryl CoA reductase)</td>
</tr>
</tbody>
</table>
that described for other deoxysugars with the activation of glucose (ORF6*) and 4,6-dehydration of TDP glucose (ORF5*) to generate TDP-4-keto-6-deoxyglucose (Decker et al., 1996). ORFs 8*, 9* and 10* encode gene products with very high identity to a dTDP-hexose 4-ketoreductase, d-TDP-sugar-3-ketoreductase and dTDP-sugar-2,3-dehydratase respectively.

Formation of the sugar moiety of concanamycin A requires the elimination of the 2'-hydroxyl from the glucose ring. Such an elimination is seen in the formation of L-mycarose in erythromycin A biosynthesis (Summers et al., 1997). This process requires the action of two enzymes, namely a 2,3-dehydratase followed by a 2,3-enoyl reductase. A homologue of the gene for the latter enzyme is not seen in the concanamycin cluster. An alternative possibility is suggested by the presence of the two putative ketoreductases. We propose that the 2,3-dehydratase catalyses the formation of the 2-deoxy derivative as in erythromycin A. This intermediate is in equilibrium with the 2,4-diketo derivative. The 4- and 3-ketoreductases then directly reduce each keto group at C3 and C4 to give the 2-deoxy derivative with appropriate stereochemistry at the C3 and C4 positions. ORFs 8*, 9* and 10* again appear closely coupled together and may form part of the same transcription unit.

The pendant sugar of concanamycin A is unusual in that it is modified by the addition of a carbamoyl unit. Carbamoylation has been observed in other secondary metabolite natural products such as novobiocin (Steffensky et al., 2000) and cephamycin (Liras & Rodriguez-Garcia, 2000). ORF7* shows high end-to-end similarity to numerous authentic carbamoyltransferases. The best matches are to the carbamoyltransferase gene required for sugar O-carbamoylation in the synthesis of nodulation factors in many bacterial species. This homology might reflect the recruitment of a gene from such a nodulation cluster.

The final gene of this group, ORF11*, appears closely coupled to the downstream gene ORF10*. This gene product shows high end-to-end similarity to authentic glycosyltransferases. The best match is to the glycosyltransferase required for the addition of deoxyallose to tylosin, reflecting the close structural similarity in the sugar substrates of these two enzymes. The proposed pathway for the synthesis and attachment of the sugar moiety of concanamycin A is shown in Fig. 3.

**Regulatory genes and genes of unknown function**

The concanamycin biosynthetic gene cluster contains three genes with a putative regulatory role, ORFs 2, 3 and 17*. ORF3 shows high end-to-end homology to members of the LuxR regulatory family and has been described in a wide range of secondary metabolite biosynthetic gene clusters in Gram-positive organisms. Such molecules bind to DNA through a helix–turn–helix motif when activated by the binding of an autoinducer molecule such as N-(3-oxohexanoyl)-L-homoserine lactone. A rare TTA leucine codon encodes the 13th amino acid of the gene product. Such TTA codons have an important regulatory function in that they are encoded by the bldA gene product and identify a gene coupled to sporulation and secondary metabolism (Leskiw et al., 1991). ORFs 2 and 17* have also been assigned

---

**Fig. 3.** Proposed biosynthetic pathway for synthesis and addition of the sugar moiety 4'-O-carbamoyl-2'-deoxyrhamnose as inferred from the sugar biosynthetic genes identified within the concanamycin A gene cluster.
putative regulatory roles on the basis of their similarity to genes in other PKS gene clusters.

ORF1 shows end-to-end homology with discrete malonyltransferase enzymes acting on malonyl-CoA. The function of this enzyme in the biosynthesis of concanamycin A is uncertain. One other potential small ORF can be identified within the gene cluster. A potential 114 aa ORF, designated ORF4*, lies between ORF4 and ORF5, transcribed in the same direction as ORF5*. This ORF has no match in any published databases and its identity remains speculative.

The concanamycin gene cluster contains two ‘pseudogenic remnants’ within the gene cluster. These remnants flank the carbamoyltransferase activity. Downstream of this gene and transcribed in the same direction is the start of a gene that encodes the first 33 aa of deoxyallosyltransferase (the complete gene is located upstream and separate from it) and upstream of the carbamoyltransferase is a stretch of DNA that encodes 30 aa from an insertion sequence. We speculate that these remnants show evidence of recent ‘evolution’ of the cluster characterized by the incorporation of a carbamoylation activity that modifies the attached sugar. Such gene remnants are not an unusual finding in macrolide biosynthetic clusters but are rarely commented upon.

The margins of the biosynthetic gene cluster can only be accurately determined by systematic gene disruption and analysis of the resulting phenotype. However, the upstream region of the cluster has been extended for a further 15 kbp and has been shown to encode a second cluster encoding the synthesis of a member of the angucyclinone group of antibiotics. Downstream of the last two genes are the β-hydroxybutyryl-CoA dehydrogenase- and crotonyl-CoA reductase-encoding genes as found flanking the oligomycin cluster (Omura et al., 2001). Transcribed towards these ORFs is a gene with very high end-to-end similarity to ribonuclease and these two ORFs are therefore likely to mark the margin of the cluster on this side.

**Provision of extender units and modification of the PKS**

The aglycone core of complex polyketides is derived from the repeated condensation of acyl-CoA building blocks, most commonly acetate units derived from malonyl-CoA and propionate units derived from methylmalonyl-CoA. In addition, less commonly, butyrate extender units are found, derived from ethylmalonyl-CoA and methoxacetate units, derived from methoxymalonyl-ACP. Condensation 10 of the biosynthesis of the concanamycin aglycone would appear to require ethylmalonyl-CoA. The examination of the AT domain for this extension cycle is consistent with a preference for ethylmalonyl-CoA over methylmalonyl-CoA (Reeves et al., 2001) (see also below). ORFs 18 and 19 appear to encode a crotonyl-CoA reductase and β-hydroxybutyryl-CoA reductase, respectively, which could both play a role in furnishing butyrate extender units. Crotonyl-CoA reductase is commonly found in polyketide biosynthetic gene clusters where an ethylmalonyl extender unit is required, as in tylosin (Butler et al., 2001), elaio-phylin (Haydock et al., 2004) and oligomycin (Omura et al., 2001) biosynthesis. β-Hydroxybutyryl-CoA reductase is less common but is also found in the oligomycin biosynthetic cluster (Omura et al., 2001). ORFs 18 and 19 are therefore likely to be involved in the synthesis of the butyrate extender unit for module 10. The extension units for modules 6 and 14 could be derived from malonyl-CoA, followed by subsequent hydroxylation and methylation. Alternatively, this unit could be derived from direct incorporation of methoxymalonyl-ACP as is seen in the biosynthetic clusters for ascomycin (Wu et al., 2000; Reeves et al., 2002), ansamitocin (Yu et al., 2002) and geldanamycin (Rascher et al., 2003). The sequenced gene cluster contains no P450 hydroxylase genes and only one O-methyltransferase. This latter gene, ORF12, is most similar to FkbG of the ascomycin biosynthetic cluster and is one of several genes suggested to be involved in biosynthesis of methoxymalonyl-ACP. Homologues of the remaining genes required for this pathway are also present and clustered together immediately downstream of ORFF in the concanamycin cluster, but separated from the FkbG homologue. ORF1* is homologous to FkbB, ORF2* to FkbJ, ORF3* to FkbI and ORF4* to FkbH. The products of these ORFs are thought to convert a primary metabolite to methoxymalonyl-ACP by a series of reactions that have not been well characterized. ORFs 1*–4* again may be transcriptionally coupled to each other and to the upstream PKS genes. It is therefore reasonable to conclude that modules 6 and 13 of the concanamycin PKS incorporate methoxymalonate extender units.

**Organization of the concanamycin PKS**

The biosynthesis of the polyketide backbone of concanamycin A is encoded by six large ORFs designated conA, conB, conC, conD, conE and conF which are adjacent and transcribed in the same direction and which encode, respectively, the multienzyme polypeptides CON1–6. Examination of the encoded AT domains of CON polypeptides for distinctive amino acid sequence motifs allows the chemical nature of the extender unit recruited by each AT domain (especially malonate versus methylmalonate or ethylmalonate) to be inferred (Haydock et al., 1995; Lau et al., 1999; Reeves et al., 2001; Del Vecchio et al., 2003). A signature motif for AT domains that incorporate methoxymalonate has, because of the rarity of this extension unit, hitherto eluded definition (but see below). Examination of the ‘reductive loop’ (Staunton & Weissman, 2001) domains (KR, DH, ER) in each extension module allows the level of reduction to be deduced for that extension. In the case of hydroxacyl extension units, even the stereochemical configuration of the hydroxy group can be predicted (Caffrey, 2003). The predicted polyketide product of the concanamycin PKS locus from our analysis is entirely consistent with assembly of the polyketide chain of concanamycin A, the modules being encoded along the chromosome in the order they are used from conA to conF (Fig. 2).
CON1 (2841 aa, approx. 300 kDa) encodes the loading domain and module 1 of the polyketide chain extension; CON2 (4825 aa, approx. 500 kDa) encodes modules for extensions 2–4; CON3 (4970 aa, approx. 520 kDa) encodes modules for extensions 5–7; CON4 (4490 aa, approx. 430 kDa) encodes modules 8–9; CON5 (5599 aa, approx. 530 kDa) encodes modules 10–13; CON6 (2056 aa, approx. 215 kDa) encodes the final module for condensation 14 and has a C-terminal integral thioesterase domain required for lactonization and release of the polyketide chain (Fig. 4).

Identification of an AT sequence motif predicting the incorporation of extension units from methoxymalonyl-ACP

It has previously been found possible to distinguish the specificity of AT domains for malonyl, methylmalonyl and ethylmalonyl extender units by analysis of short sequence motifs found within these domains (Haydock et al., 1995; Lau et al., 1999; Reeves et al., 2001; Del Vecchio et al., 2003) and this has been invaluable in helping predict the products of newly cloned PKS loci, as well as providing the basis for attempts to engineer changes in active-site specificity (Reeves et al., 2001; Del Vecchio et al., 2003). However, it has not been possible so far to identify those AT domains that specifically incorporate methoxymalonate extensions.

We therefore sought to exploit the high mutual sequence homology between AT domains within the concanamycin PKS to identify residues that might hallmark AT domains specific for methoxymalonyl-ACP and differentiate them from AT domains that recruit methylmalonyl-CoA. We identified a short span of residues that distinguished these two AT groups and which lie immediately N-terminal of the previously described YASHS motif that distinguishes methylmalonyl- from malonyl-incorporating AT domains (Lau et al., 1999). All the concanamycin AT domains that are proposed to incorporate methylmalonate have the sequence RXRR(X)3YASHS (where X is any amino acid) whilst those proposed to incorporate methoxymalonate all have the sequence MXWR(X)3YASHS.

To confirm the significance of this motif we used the program BLAST to search the UNIPROT protein sequence databases with a 40 residue span of the peptide sequence of module 13 of the concanamycin PKS centred on the methionine and tryptophan residues identified above. Interestingly, we found very few streptomycete modular PKS AT domains with a tryptophan in the position found in the concanamycin AT domains. This position is highly conserved in methylmalonyl-specific AT domains and is usually occupied by an arginine (as in the methylmalonate-specific domains of the concanamycin PKS), or by a lysine or less commonly by serine or threonine. In addition, the position corresponding to the methionine in the motif is usually occupied by a positively charged residue such as arginine. We were able to identify only six streptomycete modular AT peptide sequences in the database in which the tryptophan was present at the specified position with a non-polar residue in the position corresponding to methionine. The alignment of these sequences is shown in Table 2.

This search correctly identified four modular AT domains that are proposed to encode methoxymalonate extensions, from the geldanamycin, ansamytocin and niddamycin gene clusters. In addition, it identified two domains from a cluster of genes of unknown function that had been previously sequenced from a rapamycin-producing strain of *Streptomyces hygroscopicus* (Ruan et al., 1997). It is not known if this cluster does encode the use of methoxymalonyl-ACP-derived extender units. However, we have recently sequenced a cluster of genes in its entirety from *Streptomyces* sp. DSM 4137 that is almost identical to this published sequence (S. F. Haydock & T. Mironenko, unpublished data). It also contains the conserved ‘methoxymalonate-specific’ motif in the corresponding domains. Immediately upstream of the PKS genes in this cluster is a complete set of genes required for the biosynthesis of the methoxymalonyl-ACP precursor of methoxymalonyl-CoA. This provides further (albeit indirect) evidence that this short sequence could be involved in the recognition and specificity of these domains for methoxymalonyl-ACP.
Microbiology

S. F. Haydock and others

Table 2. Short sequence motif that identifies AT domains that accept methoxymalonyl-ACP extender units

<table>
<thead>
<tr>
<th>Gene cluster</th>
<th>Module</th>
<th>UNIPROT accession no.</th>
<th>Alignment</th>
<th>Proposed module AT specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concanamycin</td>
<td>Module 6</td>
<td>–</td>
<td>RLSEDGVTVTIPGVTFAHSPQVDE</td>
<td>Methoxymalonyl-ACP</td>
</tr>
<tr>
<td>Concanamycin</td>
<td>Module 13</td>
<td>–</td>
<td>ELSAEVGWMRRIP–VDYASHSDVST</td>
<td>Methoxymalonyl-ACP</td>
</tr>
<tr>
<td>Geldanamycin</td>
<td>Module 2</td>
<td>Q84G24</td>
<td>TLSSAKVYMNQLLEDFAHSGHVDT</td>
<td>Methoxymalonyl-ACP</td>
</tr>
<tr>
<td>Geldanamycin</td>
<td>Module 5</td>
<td>Q84G23</td>
<td>TLSSAKVYMNQLPGDFAHSGHVDT</td>
<td>Methoxymalonyl-ACP</td>
</tr>
<tr>
<td>Unknown</td>
<td>Module 1</td>
<td>O30479</td>
<td>QLSHRHIMNQLPVDFAHSGHIDQ</td>
<td>Unknown</td>
</tr>
<tr>
<td>Unknown</td>
<td>Module 3</td>
<td>O30481</td>
<td>QLSHRHIMNQLPVDFAHSGHIDQ</td>
<td>Unknown</td>
</tr>
<tr>
<td>Ansamytocin</td>
<td>Module 3</td>
<td>Q8KUH4</td>
<td>VLSARRVLLKIPGDFAHSPQVDA</td>
<td>Methoxymalonyl-ACP</td>
</tr>
<tr>
<td>Niddamycin</td>
<td>Module 6</td>
<td>O30767</td>
<td>RLSAEGVWRPLPGDFAHSPQQEE</td>
<td>Methoxymalonyl-ACP</td>
</tr>
</tbody>
</table>

Confirmation of the identity of the concanamycin biosynthetic locus by PKS gene disruption

Although the organization of the gene cluster isolated and sequenced from S. neyagawaensis is entirely and exactly consistent with its encoding concanamycin A biosynthesis, confirmation of the identity of this locus was sought by gene disruption. To do this a 9 kbp BamHI fragment internal to the putative PKS gene conE was subcloned into the E. coli Streptomyces shuttle vector pOJ260 (Bierman et al., 1992). The resulting plasmid was used to transform E. coli ET12567 (MacNeil et al., 1992) containing the helper plasmid pUZ8002 (Paget et al., 1999) and this recombinant strain was used for conjugation with S. neyagawaensis. Stable apramycin-resistant integrants were obtained via homologous recombination between the 9 kbp insert cloned into the suicide vector and the host genome. Both wild-type and recombinant strains were grown under conditions optimal for production of concanamycin A. Using liquid chromatography linked to mass spectroscopy, a peak was observed for the wild-type strain, whose exact mass corresponded to that predicted for concanamycin A. No compound of this mass was detected in the recombinant strain grown under the same conditions (see Methods), confirming the proposed role for this locus in concanamycin A biosynthesis.

Concluding remarks

The analysis of the concanamycin A biosynthetic locus from S. neyagawaensis has revealed a set of modular PKS genes, clustered with an extensive set of genes involved in provision of unusual precursor units. The uncovering of a specificity motif in certain of the AT domains of the PKS which reliably distinguishes those that are specific for methoxymalonate will not only aid in the identification of newly cloned PKS genes, but paves the way for AT engineering by mutagenesis of this motif. Finally, the provision of the sequence for this key plecomacroide, together with methods for genetic manipulation of the producing strain, should allow the genetic modification of the cluster to derive novel structural variants of concanamycin with potential utility in a wide range of medical conditions.

ACKNOWLEDGEMENTS

This work was supported by the Wellcome Trust through the award of an Advanced Clinical Research Fellowship to S. F. H.

REFERENCES


