

The enduracidin biosynthetic gene cluster from *Streptomyces fungicidicus*

Xihou Yin and T. Mark Zabriskie

Correspondence

Xihou Yin
yinx@onid.orst.edu

T. Mark Zabriskie
mark.zabriskie@oregonstate.edu

Department of Pharmaceutical Sciences, College of Pharmacy, Oregon State University, Corvallis, OR 97331-3507, USA

The biosynthetic gene cluster for the 17 aa peptide antibiotic enduracidin has been cloned and sequenced from *Streptomyces fungicidicus* ATCC 21013. The 84 kb gene cluster contains 25 ORFs and is located within a 116 kb genetic locus that was fully sequenced. Targeted disruption of non-ribosomal peptide synthetase (NRPS) genes in the cluster abolished enduracidin production and confirmed function. The cluster includes four genes, *endA–D*, encoding two-, seven-, eight- and one-module NRPSs, respectively, and includes unique modules for the incorporation of citrulline and enduracididine. The NRPS organization generally follows the collinearity principle, and starts with a condensation domain (C domain) similar to those found in other lipopeptide systems for the coupling of an acyl group to the starting amino acid. The sixth module of EndB, corresponding to Thr⁸, is missing an adenylation domain (A domain) and this module is presumed to be loaded *in trans* by the single module protein EndD. The most striking feature of the NRPS organization is the lack of epimerization domains (E domains) in light of the fact that the product has seven D-amino acid residues. Sequence analysis reveals that C domains following modules corresponding to D-amino acids belong to a unique subset of C domains able to catalyse both epimerization and condensation reactions. Other genes directing lipid modification and activation, and formation of the non-proteinogenic amino acids 4-hydroxyphenylglycine and enduracididine are readily identified, as are genes possibly involved in regulation of antibiotic biosynthesis and export. These findings provide the basis to further genetically manipulate and improve lipodepsipeptide antibiotics via combinatorial and chemical methods.

Received 6 April 2006
Revised 16 June 2006
Accepted 6 July 2006

INTRODUCTION

The global emergence of multidrug-resistant bacterial infections has resulted in enormous healthcare costs and has become a major threat to public health. In the USA alone, the total cost linked to antibiotic-resistant infections has been estimated at US\$5 billion annually (Zinner, 2005). Since its launch in 1958, vancomycin was the drug of last resort for treating Gram-positive pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA) (Barrett, 2005). However, vancomycin-resistant *S. aureus* strains have emerged and over 28% of enterococci responsible for intensive care unit infections in the USA are now vancomycin-resistant (Tenover & McDonald, 2005). In addition, resistance to the newest antibiotics to treat these infections, linezolid and daptomycin, has already

been described (Mangili *et al.*, 2005; Meka *et al.*, 2004; Tsiodras *et al.*, 2001). To stay ahead of the development of antibacterial drug resistances, there is a pressing necessity to identify new antibiotics, especially those with novel mechanisms of action.

Enduracidin is a 17 aa lipodepsipeptide produced by the soil bacterium *Streptomyces fungicidicus* ATCC 21013 (Fig. 1). The peptide is isolated from the fermentation broth and mycelia as a mixture of enduracidins A and B, which differ in the length of the attached lipid chain (Hori *et al.*, 1973; Iwasaki *et al.*, 1973). The only characterized analogue of enduracidin is ramoplanin (Fig. 1), which is currently in Phase III clinical trials for the oral treatment of intestinal vancomycin-resistant *Enterococcus faecium* (VRE) and in phase II trials for nasal MRSA (Fang *et al.*, 2006; Walker *et al.*, 2005). At present, there is no known form of developed or acquired resistance to ramoplanin or enduracidin.

Like ramoplanin, enduracidin is active towards a wide variety of Gram-positive bacteria, including VRE and MRSA (Goto *et al.*, 1968; Kawakami *et al.*, 1971; Komatsuzawa *et al.*, 1994; Peromet *et al.*, 1973; Tsuchiya *et al.*, 1968; Yourassowsky & Monsieur, 1972). These peptides disrupt

Abbreviations: Am, apramycin; Cit, L-citrulline; Dpg, 3,5-dichloro-L-4-hydroxyphenylglycine; End, D- and L-enduracididine; Hpg, D- and L-4-hydroxyphenylglycine; MRSA, methicillin-resistant *Staphylococcus aureus*; NRPS, non-ribosomal peptide synthetase; Orn, D-ornithine; PCP, peptidyl carrier protein; PDH, prephenate dehydrogenase; VRE, vancomycin-resistant *Enterococcus faecium*.

The GenBank/EMBL/DBJ accession number for the sequence reported in this paper is DQ403252.

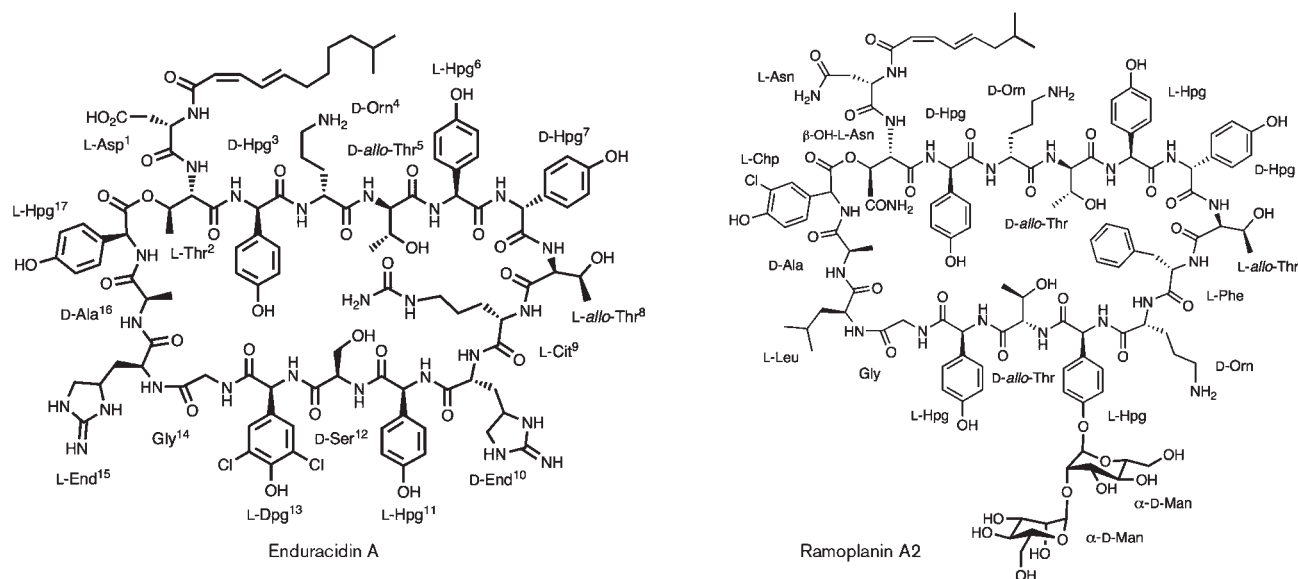


Fig. 1. Structures of enduracidin A and ramoplanin A2.

bacterial cell wall biosynthesis, but have a mechanism of action that is distinct from the β -lactams and vancomycin. Enduracidin and ramoplanin block the elongation step of peptidoglycan biosynthesis by binding to the transglycosylase substrate Lipid II (Cudic *et al.*, 2002; Fang *et al.*, 2006). This substrate-binding mechanism is analogous to the vancomycin mode of action, but these different peptide antibiotics recognize distinct regions of Lipid II. The promising activity and unique mode of action of enduracidin, coupled with novel structural features and intriguing questions in the biosynthesis, make this peptide an attractive target for further antibiotic development studies. As a first step to facilitate such studies, we have localized the enduracidin biosynthetic gene cluster to a 116 kb contiguous segment of the *S. fungicidicus* ATCC 21013 chromosome, cloned the cluster on a set of overlapping fosmid inserts and completed the DNA sequence of this region. Initial genetic characterization of the gene cluster through disruption of non-ribosomal peptide synthetase (NRPS) genes confirmed the involvement in enduracidin biosynthesis.

METHODS

Bacterial strains, plasmids, fosmids and culture conditions.

Streptomyces fungicidicus ATCC 21013 and *Escherichia coli* S17-1 (ATCC 47055) were purchased from ATCC. *E. coli* strains DH5 α (Life Technologies Inc.), EPI300 (Epicentre) and XL10-Gold (Stratagene) were routinely used as hosts for *E. coli* plasmids, fosmids and *E. coli*-*Streptomyces* shuttle vectors. Plasmid pSET152 (Bierman *et al.*, 1992) was obtained from Professor K. F. Chater (Norwich), the pGEM-T easy cloning vector was from Promega and the pCC1FOS system was purchased from Epicentre. Media and culture conditions for *S. fungicidicus* were described by Higashide *et al.*

(1968). All *E. coli* procedures were performed according to standard protocols (Sambrook & Russell, 2001). Standards of enduracidin A and B were purchased from Sigma-Aldrich.

DNA isolation and manipulations. Isolation of chromosomal DNA from *S. fungicidicus* and agarose gel electrophoresis were performed according to Kieser *et al.* (2000), and QIAprep Spin Miniprep kits (Qiagen) were used to prepare plasmids and fosmids from *E. coli* strains. Restriction endonucleases, DNA ligase, DNA polymerase and alkaline phosphatase were purchased from various sources and used according to the manufacturers' recommendations. DNA fragments were purified using QIAquick Gel Extraction kits (Qiagen).

Generation of NRPS and prephenate dehydrogenase (PDH) gene probes.

PCR primers were designed to amplify internal fragments of NRPS genes, corresponding to the region between conserved motifs A3 and T (Marahiel *et al.*, 1997). Primer design took into consideration *Streptomyces* codon bias (Wright & Bibb, 1992). The degenerate primers were PSA3f, 5'-ATCTACACSTCSGGC-ACSACSGGCAAGCCSAAGGG-3', and PSTr, 5'-AWIGAGKSIC-CICCSRRSIMGAGAA-3' (S=G+C; W=A+T; K=T+G; R=G+A; M=A+C; I=inosine). The PCR template was *S. fungicidicus* genomic DNA digested with *Hind*III, *Bam*HI or *Bgl*II. PCR mixtures (final volume of 100 μ l) contained 2.5 μ g digested chromosomal DNA, 100 pmol each primer, 0.25 mM dNTPs, *Taq* DNA polymerase reaction buffer without MgCl₂ (Promega), 1.5 mM MgCl₂, 5% DMSO and 1 μ l polymix (added at 80 °C) from the Expand Long Template PCR System (Roche). PCR was performed as follows: 1 cycle for 3 min at 95 °C and 1 min at 80 °C, 30 cycles for 1 min at 95 °C, 1 min at 55 °C and 2 min at 72 °C. The reaction was terminated with one extension cycle at 72 °C for 10 min. PCR products of the correct size (1.2 kb) were gel-purified and cloned into the pGEM-T easy vector. DNA sequence analysis of randomly selected clones revealed five unique plasmids: pGEMTE-sfPS5, pGEMTE-sfPS9, pGEMTE-sfPS11, pGEMTE-sfPS15 and pGEMTE-sfPS18.

To amplify a PDH gene probe, degenerate oligonucleotide primers (Pdhf, 5'-GGSACCGGCTSATCGGBACSTCS-3', and Pdhr, 5'-GTGSGAGACGAGSGCCACSGCSCGTGCTG-3'; S=G+C, B=G+C+T) were designed based on the alignment of PDH proteins

from *Actinoplanes* sp. ATCC 33076 (AX417445), *S. coelicolor* A3(2) (NP733544) and *S. avermitilis* MA-4680 (NP827697). The PCR conditions were the same as described above. An amplicon of the expected size (0.5 kb) was purified and cloned into the pGEM-T easy vector to obtain plasmid pGEMTE-sfPDh-Nt. DNA sequence analysis confirmed that the insert (538 bp) encodes the N-terminal portion of a putative PDH.

Inactivation of enduracidin NRPS genes. The NRPS substrate specificity sequences extracted from the insert of pGEMTE-sfPS11, pGEMTE-sfPS15 and pGEMTE-sfPS18 predicted that they activate Thr, Hpg and Asp, respectively, and are consistent with involvement in enduracidin formation (Challis *et al.*, 2000; Rausch *et al.*, 2005; Stachelhaus *et al.*, 1999). These plasmid inserts were used to construct gene disruption plasmids in vector pXY300, an *E. coli*/*Streptomyces* temperature-sensitive conjugal vector (Yin *et al.*, 2003). A 1 kb fragment from plasmid pSET152 carrying the *aac(3)IV* gene conferring apramycin resistance (Am^R) was amplified by PCR using primers that introduced *Bam*HI/*Hind*III restriction sites at both ends (apraRf, 5'-CACGGATCCAAGCTTGGTTCATGTGCA-3', and apraRr, 5'-ATCGGATCCAAGCTTCACGTGTTGC-3'; *Bam*HI/*Hind*III sites are in bold). The gene disruption plasmids were constructed as follows. The Am^R fragment was ligated with the unique *Bgl*II site in the insert of pGEMTE-sfPS11 to yield plasmid pGEMTE-sfPS11- Am^R . The insert of pGEMTE-sfPS11- Am^R was excised with *Eco*RI and ligated with similarly restricted pXY300 to obtain the final construct, pXY300-sfPS11- Am^R . Site-directed mutagenesis was used to introduce *Bgl*II restriction sites near the middle of the pGEMTE-sfPS15 and pGEMTE-sfPS18 inserts. Two sets of primers were synthesized (sfPS15*Bgl*II, 5'-TCTACGTCCTGGAC-AGATCTCTGAACCCGGTG-3', and sfPS15*Bgl*IIr, 5'-CACCGGGTTCAGAGATCTGTCCAGGACGTAGA-3'; sfPS18*Bgl*II, 5'-TCAACCCGATACCCGTCAGATCTCTGGGGGAGCT-3', and sfPS18*Bgl*IIr, 5'-AGTCCCCCAGAGATCTGACGGGTACGGGGTTGA-3'; *Bgl*II sites are in bold) and the desired *Bgl*II sites were created using the QuikChange Site-directed Mutagenesis kit (Stratagene), following the manufacturer's protocol. The Am^R fragment was ligated into the newly created *Bgl*II sites to produce plasmids pGEMTE-sfPS15- Am^R and pGEMTE-sfPS18- Am^R . The inserts of these two plasmids were excised by *Eco*RI digestion and ligated with the similarly restricted pXY300 to give the final constructs, pXY300-sfPS15- Am^R and pXY300-sfPS18- Am^R .

The gene disruption plasmids were individually introduced into *E. coli* S17-1 by transformation and then conjugation with *S. fungicidicus* was carried out according to the literature (Kieser *et al.*, 2000). Briefly, freshly harvested *S. fungicidicus* spores were pre-germinated and *E. coli* S17-1 cells were grown overnight at 37 °C in Terrific broth (Sambrook & Russell, 2001). Serial dilutions of the germinated spore suspension were made and 100 µl of each dilution was mixed with an equal volume of *E. coli* S17-1 harbouring the pXY300-based disruption plasmids. The solutions were plated onto ISP4 agar plates with addition of 10 mM MgCl_2 and incubated for 22 h at either 30 or 37 °C. Each plate was overlaid with 3 ml soft nutrient agar containing sodium nalidixate and Am (0.5 mg each ml^{-1}) and further incubated at 30 °C for about 1 week. Isolated exconjugants that survived antibiotic selection were purified by streaking onto ISP4 agar plates supplemented with sodium nalidixate and Am (50 µg each ml^{-1}).

To conduct the gene disruption experiments, exconjugants were first cultured in TSB liquid medium containing Am (5 µg ml^{-1}) at 30 °C for 24 h at which time the mycelia were harvested, homogenized and used to inoculate TSB and YEME liquid media supplemented with Am (5 µg ml^{-1}). After 3–6 days incubation at 40 °C, the mycelia were homogenized and plated onto ISP4 agar plates containing Am (50 µg ml^{-1}) and incubated at 30 °C for 1 week. Genomic DNA was

isolated from randomly selected individual surviving colonies and analysed by Southern blotting to confirm that single- or double-crossover disruption had occurred.

Construction of *S. fungicidicus* genomic libraries. Genomic DNA was prepared from *S. fungicidicus* ATCC 21013 wild-type and disruptant strains according to Kieser *et al.* (2000). Three genomic libraries were constructed using the CopyControl Fosmid Library Production kit and EPI300 *E. coli* plating strain (Epicentre), following the manufacturer's specifications. Briefly, chromosomal DNA was end-repaired and fractionated on a 1% low-melting-point agarose gel. The band at approximately 40 kb was excised and recovered from the agarose gel. After gelase digestion, the DNA was precipitated with sodium acetate and ethanol. Purified genomic DNA was ligated with linearized, dephosphorylated pCC1FOS^M vector at room temperature for 2 h. The ligation mixture was packaged using MaxPlax Lambda Packaging Extract (Epicentre) followed by transfection into *E. coli* EPI300 cells.

The first library was constructed using *S. fungicidicus* wild-type genomic DNA. A partial genome scan was performed by end-sequencing the inserts of 389 randomly selected clones using the universal T7 promoter primer. Eleven clones were identified as harbouring a portion of the enduracidin or other NRPS gene cluster. Both ends of these inserts were sequenced.

The second fosmid library was constructed using the pooled genomic DNA from *S. fungicidicus* strains with disrupted enduracidin NRPS genes. This library included approximately 60 000 colonies and was plated onto LB agar medium containing Am (100 mg ml^{-1}). Only 91 clones survived antibiotic selection. The fosmids were purified and the inserts were end-sequenced in order to find DNA segments that spanned gaps between positive fosmids in the first library.

A third library was constructed using wild-type genomic DNA and was divided into two portions of approximately 5500–6000 colonies. Each portion was screened by *in situ* hybridization. The first portion of the library was screened using a 5.8 kb *Bam*HI fragment from the left end of the pXYF24 insert and a fragment encoding a ThrA domain amplified by PCR. Forty positive colonies were obtained and 24 of these were randomly selected for end-sequencing. This allowed us to identify the extending inserts in fosmids pXYF103 and pXYF305. The second portion of this library was screened using a 2.6 kb internal *Bam*HI fragment from the pXYF305 insert. Thirty-one positive colonies were identified and 16 were end-sequenced, identifying fosmid pXYF607 which further extended the contiguous cloned region of DNA.

HPLC analysis of enduracidin production. Fermentation conditions for the production of enduracidin from wild-type *S. fungicidicus* ATCC 21013 and the disruptant strains SfPS18D17 and SfPS18D29 were as described by Higashide *et al.* (1968). For HPLC analysis, 60 g freshly harvested mycelia was washed with 120 ml deionized water, resuspended in 120 ml methanol and sonicated with a microprobe tip for 1 min at 15 W. The mixture was shaken at 230 r.p.m. at 18 °C for 3 h and then centrifuged at 2000 g for 20 min. The supernatant was collected and evaporated to near dryness at 35 °C under reduced pressure and resuspended in 10 ml 90% methanol. The solution was adjusted to pH 4.3 with 1 M HCl and centrifuged at 2000 g for 20 min. The supernatant was filtered through a 0.45 µm syringe filter prior to analysis. HPLC was performed using a Gemini C₁₈ column (4.6 × 150 mm, 5 µm; Phenomenex) with isocratic elution in 30% acetonitrile and 70% 50 mM NaH_2PO_4 , pH 4.5, at a flow rate of 1.0 ml min^{-1} . The UV region from 200–300 nm was scanned with a photodiode array detector; 267 nm was monitored with a variable wavelength detector.

Southern hybridization. *S. fungicidicus* genomic DNA was cleaved with restriction endonucleases, electrophoresed in 0.8% agarose gels

and transferred onto Hybond-N nylon membranes (Roche). The manufacturer's protocol for colony lifts for *in situ* hybridization was followed. DNA probes were prepared using a digoxigenin-labelled system and hybridization was revealed using a digoxigenin-DNA detection kit (Roche).

DNA sequencing and analysis. Routine DNA sequencing of plasmids, PCR products, primer walking sequencing and fosmid insert end-sequencing was performed at the Oregon State University Center for Genome Research and Biocomputing (CGRB) using the AmpliTaq Dye-terminator Sequencing System (Perkin Elmer) and Applied Biosystems automated DNA sequencers (models 373 and 377). Fosmid DNA used for sequencing was prepared from 5 ml induced LB culture with a QIAprep Spin Miniprep kit (Qiagen). Sequencing of the entire fosmid inserts at 10- to 20-fold coverage was performed by Macrogen (Seoul, South Korea). The nucleotide sequences were determined for both strands. Sequence analysis was carried out using the VectorNTI (Invitrogen) software package. Nucleotide and amino acid sequence similarity comparisons were carried out in public databases using the BLAST program (Altschul *et al.*, 1990).

RESULTS AND DISCUSSION

Cloning enduracidin biosynthesis gene probes

NRPSs have highly conserved core motifs located in the A and peptidyl carrier protein (PCP) domains that permit the use of degenerate PCR primers to specifically amplify fragments of the NRPS genes (Marahiel *et al.*, 1997). Using *S. fungicidicus* ATCC 21013 genomic DNA as template and a set of degenerate oligonucleotide primers targeted for regions encoding the A3 and PCP motifs, PCR yielded amplicons of the expected size (approx. 1.2 kb) which were cloned and sequenced. Five unique peptide synthetase (PS) probes were identified and designated sfPS5, sfPS9, sfPS11, sfPS15 and sfPS18. The standard *in silico* methods to predict A domain substrate specificity indicated that each of the cloned PS probes was expected to activate an amino acid found in enduracidin (Challis *et al.*, 2000; Rausch *et al.*, 2005; Stachelhaus *et al.*, 1999).

Enduracidin contains 10 non-proteinogenic amino acid residues, including D- and L-enduracididine (End), D- and L-4-hydroxyphenylglycine (Hpg), 3,5-dichloro-L-4-hydroxyphenylglycine (Dpg), L-citrulline (Cit) and D-ornithine (Orn) (Fig. 1). The genes directing Hpg biosynthesis in the chloroeremomycin pathway have been identified and their products characterized (Choroba *et al.*, 2000; Hubbard *et al.*, 2000; van Wageningen *et al.*, 1998). Homologues of one of the genes, encoding PDH, are present in several other NRPS biosynthesis gene clusters (Chiu *et al.*, 2001; Hojati *et al.*, 2002; Li *et al.*, 2004; Pootoolal *et al.*, 2002; Sosio *et al.*, 2003). To aid rapid identification of the enduracidin gene cluster, we also sought to amplify a specific PDH gene probe. Degenerate primers were designed from conserved regions identified in PDHs from other NRPS gene clusters and used to amplify a 0.5 kb PCR product from *S. fungicidicus* genomic DNA. The PCR product was cloned into pGEM-T easy vector to yield pGEMTE-sfPhd-Nt.

Sequence analysis confirmed that it encoded the N-terminal portion of a putative PDH.

Inactivation of the sfPS11-, sfPS15- and sfPS18-containing NRPS genes

To provide direct evidence that the genes corresponding to the PCR-amplified probes are involved in enduracidin biosynthesis, probes correlating to Thr- (sfPS11), Hpg- (sfPS15) and Asp-activating (sfPS18) A domains were used for gene disruption by single- or double-crossover homologous recombination. Gene disruption constructs were made by introducing the Am^R marker into sfPS11, sfPS15 and sfPS18, carried in the pGEM-T vector. These *aac(3)IV*-containing inserts were then transferred into the *Streptomyces-E. coli* temperature-sensitive conjugal shuttle vector pXY300 (Yin *et al.*, 2003). The three gene disruption constructs, pXY300-sfPS11-Am^R, pXY300-sfPS15-Am^R and pXY300-sfPS18-Am^R, were separately introduced into *S. fungicidicus* by intergeneric conjugation (Kieser *et al.*, 2000). Independent double-crossover disruptants obtained with pXY300-sfPS18-Am^R (Fig. 2a) were confirmed by Southern blot analysis of the chromosomal DNA (Fig. 2b) and later by sequencing the insert of disrupted fosmid pXYFD16. The loss of enduracidin production from the double-crossover mutant strains SfPS18D17 and SfPS18D29 was confirmed by HPLC (Fig. 2c).

Southern blot analysis revealed that constructs pXY300-sfPS11-Am^R and pXY300-sfPS15-Am^R produced untargeted disruptants (SfPS11D9, SfPS15D12 and SfPS15D31) through partial homologous recombination (data not shown). These results are not surprising because sfPS11 exhibits 96 % identity over 879 nt with the DNA sequence encoding the Ser¹²-activating A domain from the *end* cluster. Likewise, six Hpg/Dpg-activating A domains in the *end* cluster, including module 17 corresponding to sfPS15, share 78–90 % identity over 879 nt. Inactivation of the enduracidin Ser¹²-activating A domain via single-crossover partial homologous recombination was also confirmed by sequencing the insert of the disrupted fosmid pXYFD18. HPLC analysis indicated that these disruptants also lost the ability to produce enduracidin (data not shown).

Cloning and sequencing the enduracidin biosynthetic gene cluster

Construction of a wild-type fosmid library. The first fosmid genomic library was constructed from *S. fungicidicus* wild-type chromosomal DNA. The structural similarity of enduracidin and ramoplanin and the availability of the sequenced ramoplanin gene cluster allowed us to conduct a partial genome scan to identify fosmid inserts with ends homologous to genes expected for enduracidin biosynthesis. Eleven clones were identified with NRPS genes or sequence similar to genes in the ramoplanin cluster. Three fosmids, pXYF24, pXYF148 and pXYF200, were selected for further analysis. Restriction and Southern blot analysis indicated that pXYF24 overlaps pXYF148, but

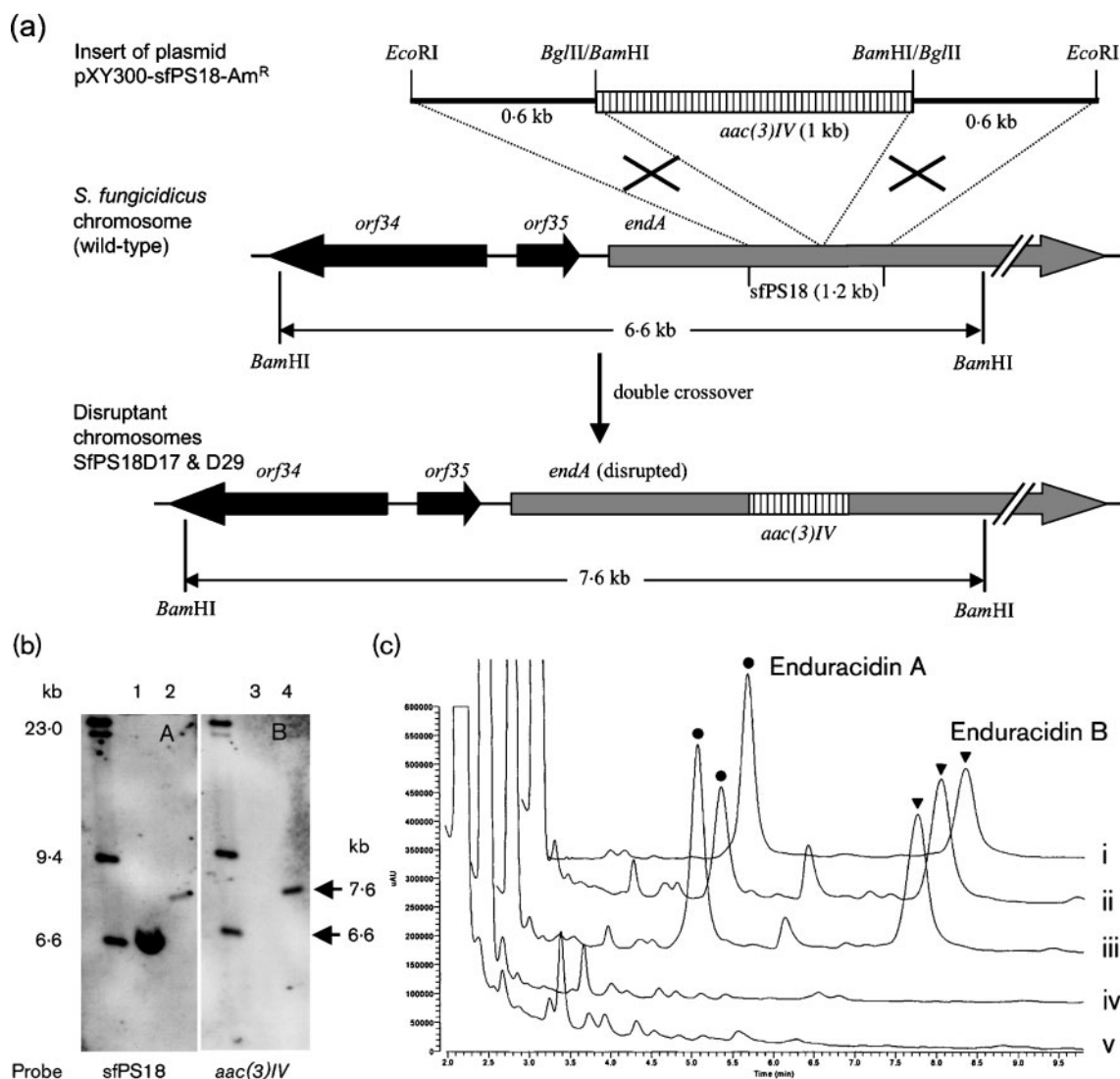


Fig. 2. Inactivation of the enduracidin biosynthetic gene cluster and metabolite analysis. (a) Insertion of the Am^R marker *aac(3)IV* into the sfPS18 A domain region of *endA* via double-crossover homologous recombination. (b) Southern blot analysis of wild-type *S. fungicidicus* (lanes 1 and 3) and disruptant (SfPS18D29) genomic DNA digested with *Bam*HI. Blot A was probed with DIG-labelled sfPS18 and blot B was probed with DIG-labelled *aac(3)IV*. (c) HPLC analysis of enduracidin production by *S. fungicidicus* wild-type and mutant strains. (i) Enduracidin A and B standards; (ii) methanolic extract of wild-type mycelia; (iii) co-injection of wild-type extract and standards; (iv and v) methanolic extract of *S. fungicidicus* disruptant strains SfPS18D17 and SfPS18D29.

neither of them overlaps pXYF200. One end of pXYF200 encodes an Hpg-activating A domain and an incomplete thioesterase domain, suggesting this segment should be near the end of the clustered NRPS genes.

Construction of a disruptant fosmid library. A second library was constructed using mixed genomic DNA prepared from the five strains with inactivated enduracidin NRPS genes. The aim was to use a positive selection strategy to find fosmids that would connect pXYF148 to

pXYF200 and also extend the region located downstream of the thioesterase end of pXYF200. Because we showed the *aac(3)IV* gene was inserted into the enduracidin cluster in these strains, clones in this library that survive antibiotic selection must harbour a fragment of the targeted cluster. A total of 91 Am^R clones were identified. Restriction fragment length analysis and additional sequence from primer walking and subcloning identified two disrupted fosmids, pXYFD16 and pXYFD18, that connected pXYF148 with pXYF200. With these fosmids, we could construct an

overlapping fosmid insert map covering approximately 100 kb (Fig. 3a).

Construction and screening of a second wild-type genomic library. Fosmids extending the contiguous segment of the chromosome beyond the ends of pXYF24 and pXYF200 were not identified in either the original wild-type or the disruptant library. Therefore, we constructed and screened a second *S. fungicidicus* wild-type genomic library. The new library was divided into two portions and each was screened by *in situ* hybridization. The first portion was screened using a fragment from the left end of the pXYF24 insert and a fragment encoding an NRPS A domain predicted to activate Thr. This allowed us to identify fosmid pXYF103, which overlaps with and extends 15 kb beyond the left end of pXYF24, and fosmid pXYF305, which overlaps with and extends 14 kb beyond the right end of pXYF200 (Fig. 3a).

The second portion of this library was screened using an internal *Bam*HI fragment from the pXYF305 insert. Fosmid pXYF607 was identified which overlaps with and extends 15 kb beyond the right end of pXYF305. In all, eight overlapping fosmid inserts were identified which span approximately 148 kb of contiguous DNA on the *S. fungicidicus* chromosome. The inserts from fosmids pXYF24, pXYF148, pXYFD16, pXYFD18 and pXYF200 were completely sequenced and that of pXYF305 was partially sequenced, covering a 116 kb region that harbours the entire enduracidin (*end*) biosynthetic gene cluster (Fig. 3b).

Overall analysis of the enduracidin (*end*) biosynthetic gene cluster

The *end* gene cluster resides on an 84 kb segment of the *S. fungicidicus* chromosome and includes 25 ORFs (*orf22* to *orf46*). The predicted function for each ORF was assigned by comparing the translated product with known proteins in public databases (Table 1). Genes identified include those required for the formation of the lipid and non-proteinogenic amino acid precursors, assembly of the peptide backbone, export and regulation of enduracidin biosynthesis. Twenty-one ORFs exhibited significant similarity with counterparts in the ramoplanin cluster (Table 1). The probable boundaries of the *end* cluster were established by comparison with the ramoplanin gene cluster and from the deduced functions of the *end* and flanking gene products.

Analysis of the enduracidin NRPSs

Four peptide synthetases, designated EndA, EndB, EndC and EndD, assemble the 17-residue enduracidin peptide backbone. The organization of the modules and domains in these four proteins is shown in Fig. 4 and, in general, follows the NRPS collinearity principle (Marahiel *et al.*, 1997). The most immediately notable features are a missing A domain from module 8 in EndB that should incorporate L-*allo*-Thr⁸, and EndD which is a single-module NRPS with a unique N-terminal region and an A domain predicted to activate Thr. The other modules and domains exhibit typical NRPS features, and the predicted substrate for the A domain of each module correlated with the residue expected based on

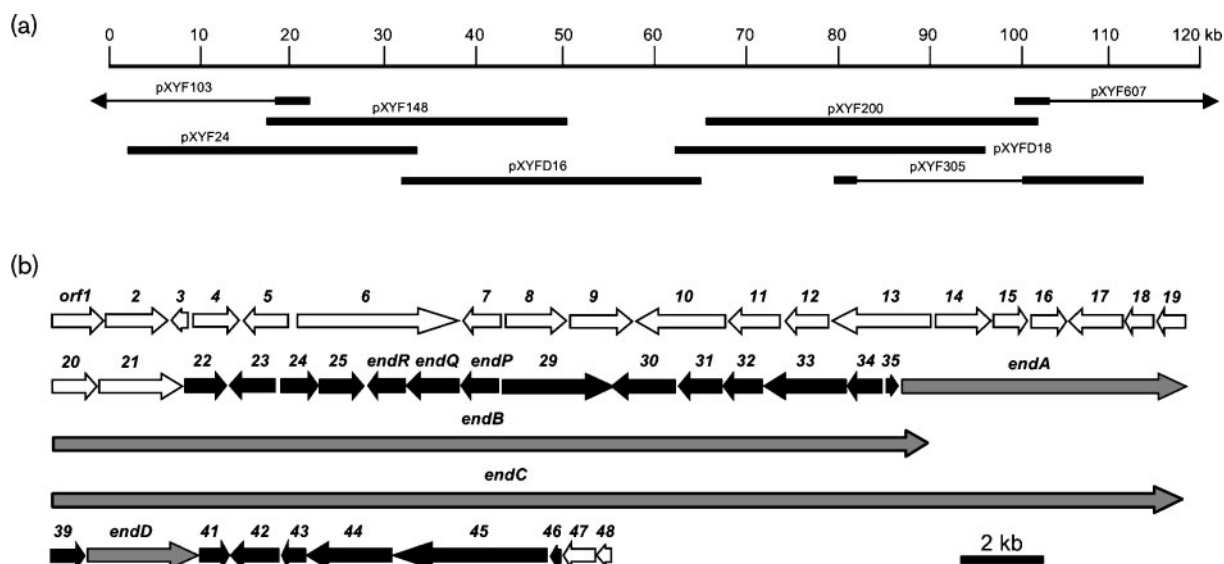


Fig. 3. (a) Overlapping fosmid inserts cloned from the *S. fungicidicus* genome that harbours the enduracidin biosynthetic gene cluster. Thick lines represent the 116 kb region that was fully sequenced. Arrows indicate where the insert extends beyond the 120 kb scale shown. (b) Organization of the sequenced *end* cluster (filled arrows) and flanking regions (white arrows). Grey arrows represent the NRPS genes.

Table 1. Summary of ORFs identified in the enduracidin gene cluster and flanking regions

The organization of the *end* cluster is shown in Fig. 3(b). Size indicates the number of amino acids in the translated product. Homologues found in the related ramoplanin biosynthetic gene cluster are presented separately from the best matching protein identified by BLAST analysis.

ORF	Gene	Size (aa)	Ramoplanin homologue (% identity)	Best match accession no.	Proposed function
1		409		BAC69955	DXP synthase 2 (incomplete, C-terminal portion)
2		498		CAA19226	Cationic amino acid transporter
3		133		CAA19227	Unknown
4		373		CAA19231	Unknown
5		360		CAA19232	Transcriptional regulator
6		1290		CAD55196	Glycosyltransferase
7		305		CAC16706	Unknown
8		482			Unknown
9		496			Unknown
10		709		CAA19238	Fatty acid oxidation complex α -subunit
11		405		CAA19239	Acetyl-CoA acetyltransferase (thiolase)
12		345	ORF33 (29)	CAA20609	Transcriptional regulator
13		793		CAA20608	β -Mannosidase
14		440		CAA20607	Probable sugar transport system lipoprotein
15		258		CAA20606	Sugar transport system permease protein
16		276		CAA20605	Sugar transport system permease protein
17		430		BAC69942	Ribonuclease D
18		220	ORF21 (26)	BAC69941	Two-component response regulator
19		224		CAA19242	Unknown
20		364		BAC69939	Uroporphyrinogen decarboxylase
21		664		AAZ54144	PAS protein phosphatase 2C-like
22		328	ORF5 (43)	BAC76461	StrR-like regulatory protein
23		362	ORF4 (51)	CAD55177	PDH
24		311	ORF5 (49)	AAM80553	Unknown (StaQ homologue)
25		356	ORF30 (48)	EAM81892	4-Hydroxyphenylpyruvate dioxygenase (HmaS homologue)
26	<i>endR</i>	279	–	AAU34211	Unknown (MppR homologue)
27	<i>endQ</i>	419	–	AAU34210	PLP-dependent aminotransferase (MppQ homologue)
28	<i>endP</i>	293	–	AAU34209	PLP-dependent aminotransferase (MppP homologue)
29		808	ORF7 (66) and ORF6 (69)	EAM81890 and EAM81900	FMN-dependent α -hydroxyacid dehydrogenase (HmaO homologue) and PLP-dependent class I and II aminotransferase (HpgT homologue)
30		504	ORF20 (62)	CAG15020	Halogenase
31		341	ORF1 (50)	EAP97899	Transmembrane transport protein
32		307	ORF2 (69) and ORF23 (55)	AAP03101	ABC transporter
33		651	ORF8 (72)	EAM81899	ABC transporter
34		275	ORF9 (76)	AAZ23080	Type II thioesterase
35		90	ORF11 (67)	BAC71361	Carrier protein (ACP/PCP)
36	<i>endA</i>	2101	RamoA		NRPS
37	<i>endB</i>	6943	RamoB		NRPS
38	<i>endC</i>	8986	RamoC		NRPS
39		274	ORF16 (79)	BAC70911	Acyl-CoA dehydrogenase/reductase
40	<i>endD</i>	859	ORF17		NRPS
41		223	ORF21 (70)	AAP03103	Two-component response regulator
42		370	ORF22 (54)	AAP03102	Two-component system sensor kinase
43		181	ORF22 (43)	AAP03102	Two-component system sensor kinase
44		625	ORF24 (62)	BAC71362	Acyl-CoA dehydrogenase
45		1177	ORF26 (53)	BAC71363	Acyl-CoA ligase/dehydrogenase fusion protein
46		71		AAX31560	MbtH-like protein
47		245		CAA19250	Integral membrane protein
48		118		CAA19250	Integral membrane protein

the collinearity rule (Table 2). The NRPS substrate-binding pocket databases do not contain signature sequences for citrulline (L-Cit, module 9) or enduracidine (D- and L-End, modules 10 and 15, respectively). Interestingly, the same A domain substrate recognition sequence, DFWSVGMV, is found in modules incorporating L-Thr² (EndA-m2), D-*allo*-Thr⁵ (EndB-m3) and L-*allo*-Thr⁸ (EndD). EndA-m2 and EndB-m3 are both predicted to recognize L-Thr as substrate and, accordingly, the regions between core motifs A4 and A5 that define much of the substrate-binding pockets of the A domains are identical at the amino acid level. The A domain of EndD is expected to recognize the rare L-*allo* diastereomer of Thr, so it is surprising that the same 8 substrate-specifying residues are shared with EndA-m2 and EndB-m3. However, the A4–A5 region of EndD shares only 59 % identity with that of EndA-m2 and EndB-m3 and is 3 residues longer, allowing for differences that could discriminate between L-Thr and L-*allo*-Thr. Also, the location of EndC-m3 corresponds to the incorporation of D-Ser¹², but the substrate recognition sequence, DVWSVAMV, most closely resembles Thr-incorporating A domains and shows moderate similarity to Ser A domains (DVWHVSLI) from the *Pseudomonas fluorescens* Pf-5 pyoverdine synthetase (Paulsen *et al.*, 2005). Interestingly, the corresponding amino acid in ramoplanin is D-*allo*-Thr and the deduced substrate specificity sequence for the ramoplanin module 12 A domain is DFWSVGMV. The presence of a Phe in the second position of the substrate specificity sequence has been suggested to be a key factor for Thr recognition and Phe is found at this position in all three enduracidin synthetase Thr A domains (Table 2) (Challis *et al.*, 2000). Inspection of the *endC* and *ramoC* sequence encoding this residue reveals a GTC in *endC* and a TTC in *ramoC*. The mutation changes the Val codon GUC to the Phe codon UUC (or vice versa) and accounts for the switch in substrate specificity from Ser to Thr by the respective peptide synthetases. Finally, the A domain in the 3,5-dichloro-L-4-hydroxyphenylglycine (Dpg) module shares the same substrate specificity sequence, DAYHLGLL, as four of the Hpg A domains, suggesting that chlorination of Hpg to yield the Dpg residue occurs on an NRPS-bound species or the nascent peptide, rather than at the free amino acid stage.

Enduracidin assembly presumably begins with EndA, predicted to be a 2101 aa (224 kDa) two-module NRPS with an N-terminal condensation domain (EndA-C1) similar to those found in other lipopeptide-loading modules proposed to accept a lipid substrate from an acyl carrier protein (Duitman *et al.*, 1999). These have been referred to as C^{III} domains and examples are found in the ramoplanin, CDA and daptomycin systems (Miao *et al.*, 2005, 2006). EndA-C2 then couples N-lipo-L-Asp¹ and L-Thr² to form a dipeptide that is transferred to EndB (6943 aa, 741 kDa). EndB is composed of seven modules, but as noted above, it is missing an A domain from module 8 (EndB-m6). EndD is a single-module NRPS (859 aa, 91 kDa) with an A domain predicted to activate Thr and a PCP domain. Preceding the A domain is an N-terminal region of approximately 280 aa that shares weak homology with NRPS C domains. Presumably, EndD activates and tethers L-*allo*-Thr to the PCP domain and the N-terminal domain directs the protein–protein interaction for the transfer of L-*allo*-Thr to module 8 (EndB-m6). Similar proposals are suggested for syringomycin and ramoplanin biosynthesis (Guenzi *et al.*, 1998; McCafferty *et al.*, 2002). EndB, with the possible participation of EndD, then catalyses the peptidyl chain elongation from amino acid residue L-Thr² to L-Cit⁹. Experimental confirmation of the function of the N-terminal portion of EndD and proof of *in trans* loading of EndB remains to be determined. EndC is the final NRPS (8986 aa, 955 kDa) and consists of eight modules and catalyses the peptidyl chain extension from amino acid residues D-End¹⁰ to L-Hpg¹⁷, and terminates with a thioesterase (TE) domain that promotes the cyclization and release of the peptide.

Analysis of the PCP domains

There are 18 thiolation (T) or PCP domains in the four enduracidin synthetases. Conceptually, 17 PCP domains are sufficient for enduracidin assembly. The extra PCP domain occurs because of the proposed transfer of L-*allo*-Thr from EndD to EndB-m6 (module 8). The exact mechanism of the substrate transfer between these NRPSs is unknown at present. One possibility is that the EndB-m6 PCP is redundant or non-functional and the C domain of EndB-m7 directly condenses the L-*allo*-Thr on EndD with the L-Cit residue bound to the EndB-m7 PCP. The C domain of EndB-m6 could couple the

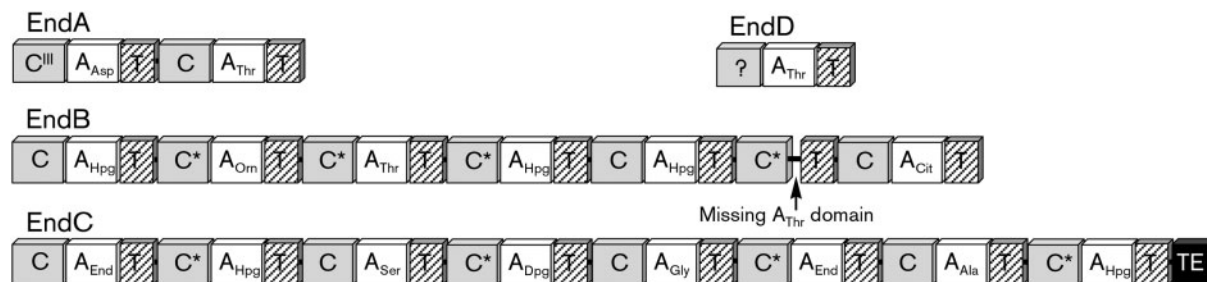


Fig. 4. Module and domain organization of the enduracidin NRPSs. Proposed dual function condensation/epimerization domains are indicated with C*.

Table 2. Derived substrate signature sequences for NRPS adenylation domains identified in the enduracidin cluster

Module	Substrate recognition sequence	Corresponding amino acid in enduracidin	Predicted amino acid
EndA-m1	DLTKVGHV	Lipo-L-Asp	Asp
EndA-m2	DFWSVGMV	L-Thr	Thr
EndB-m1	DAYHLGLL	D-Hpg	Hpg
EndB-m2	DMETDGSV	D-Orn	Orn
EndB-m3	DFWSVGMV	D- <i>allo</i> -Thr	Thr
EndB-m4	DAYHLGLL	L-Hpg	Hpg
EndB-m5	DAYHLGLL	D-Hpg	Hpg
EndB-m6	No A-domain	L- <i>allo</i> -Thr	–
EndB-m7	DMEADGAV	L-Cit	Orn/Lys
EndC-m1	DAETDGSV	D-End	Lys
EndC-m2	DAYHLGML	L-Hpg	Hpg
EndC-m3	DVWSVAMV	D-Ser	Thr
EndC-m4	DAYHLGLL	L-Dpg	Hpg
EndC-m5	DILQLGLV	Gly	Gly
EndC-m6	DAETDGSV	L-End	Lys
EndC-m7	DIFQLALV	D-Ala	Ala
EndC-m8	DAYHLGLL	L-Hpg	Hpg
EndD	DFWSVGMV	L- <i>allo</i> -Thr	Thr

resulting dipeptide on EndB-m7 with the peptidyl species attached to EndB-m5, allowing normal peptide assembly to continue. Alternatively, EndD could transfer the tethered L-*allo*-Thr directly to the PCP of EndB-m6, but a PCP to PCP transfer may require the action of a separate acyltransferase (Vaillancourt *et al.*, 2005). Interestingly, the PCP domain core motif found in both EndD and EndB-m6 is LGGNSL. The occurrence of an Asn residue preceding the Ser to which the 4'-phosphopantetheine arm is attached is rarely observed. For PCP domains that accept L-amino acids, T(L) domains, the core motif is LGGDSI, whereas for PCP domains that are associated with epimerization (E) domains and accept D-amino acids, T(D) domains, the consensus core motif is XGGHSL (Linne *et al.*, 2001). The corresponding PCP domains in the ramoplanin synthetases also have an Asn residue in the core motif. Table 3 correlates the PCP domains in the enduracidin synthetases with the respective modules and corresponding substrate amino acids. With the exception of EndB-m6 and EndD discussed above, all enduracidin synthetase PCP domains are of the T(L) type. This is consistent with the lack of E domains in the enduracidin synthetases, but does not correlate with occurrence of 7 D-amino acids in the peptide product. Interestingly, analysis of the translated products of NRPS genes from the actinomycete *Nocardia farcinica* IFM10152 reveals the core motifs of the PCP domains primarily contain Asn instead of Asp or His (Ishikawa *et al.*, 2004). The role of the rare Asn residue in the PCP domains remains to be elucidated.

Origination of the D-amino acid residues in enduracidin

Enduracidin contains 7 D-amino acids but the megasynthetase contains no epimerization (E) domains and there were

no racemase or epimerase genes detected in the sequenced region of the *S. fungicidicus* chromosome. This unusual situation has previously been observed in *Pseudomonas* peptide synthetases assembling the lipopeptides syringomycin (Guenzi *et al.*, 1998), syringopeptin (Scholz-Schroeder *et al.*, 2003) and arthrofactin (Roongsawang *et al.*, 2003). Ramoplanin synthetase is the only other actinomycete NPRS we are aware of that shares this feature (McCafferty *et al.*, 2002). It was suggested that the epimerization function required for the biosynthesis of syringomycin may be provided *in trans* by external amino acid racemase(s) (Guenzi *et al.*, 1998). Biochemical analysis of recombinant A domains from these systems demonstrated that they preferentially activate L-amino acids even when the excised A domain correlates with the position of a D-amino acid in the peptide product (Guenzi *et al.*, 1998; Recktenwald *et al.*, 2002; Roongsawang *et al.*, 2003; Trauger & Walsh, 2000). This discrepancy was recently resolved when Balibar *et al.* (2005) demonstrated that the D-amino acids in arthrofactin are generated on the surface of the NRPS by dual function C/E domains.

The formation of D-amino acid residues in enduracidin possibly involves the same mechanism demonstrated for arthrofactin. The C domains in both peptide assembly lines that are immediately downstream of modules corresponding to D-amino acids share 50–60 % overall amino acid identities and are clearly distinct from the more typical C domains following L-amino acid modules (Fig. 5). The studies by Balibar *et al.* (2005) also revealed that epimerization of the donor peptidyl-S-enzyme species only occurred when the adjacent downstream acceptor aminoacyl-S-enzyme species was present. This is relevant to the *in trans* loading of EndB-m6 discussed above because this

Table 3. Comparison of conserved core motifs in the PCP domains of enduracidin synthetases

Module	PCP domain core motif region	Corresponding amino acid
EndA-m1	DDDDFFALGGHSLPATR	Lipo-L-Asp
EndA-m2	DDDDFFALGGHSLLATR	L-Thr
EndB-m1	EDDDFFALGGHSLLA VS	D-Hpg
EndB-m2	DDDDFFALGGHSLLVVS	D-Orn
EndB-m3	DDDDFFALGGHSLLA VS	D- <i>allo</i> -Thr
EndB-m4	DDDDFFLLGGHSLLAMR	L-Hpg
EndB-m5	EDDDFFALGGHSLLA VS	D-Hpg
EndB-m6	DDDDFFALGGNSLVATR	L- <i>allo</i> -Thr
EndB-m7	DDHFFALGGHSLLATR	L-Cit
EndC-m1	EDDDFFALGGHSLLA VS	D-End
EndC-m2	DDDDFFTLLGGHSLLVTR	L-Hpg
EndC-m3	DDDDFFALGGHSLLA VS	D-Ser
EndC-m4	DDDDFFAKGGHSLLATV	L-Dpg
EndC-m5	DDDDFFALGGHSLLA VS	Gly
EndC-m6	DDDDFFALGGHSLLATR	L-End
EndC-m7	DDDDFFALGGHSLLA VS	D-Ala
EndC-m8	DDDDFFALGGHSLLA VR	L-Hpg
EndD	SDSFWE LGGNSLLAVR	L- <i>allo</i> -Thr

module possesses a C/E domain that should epimerize the Hpg residue on EndB-m5. The necessity of having the downstream PCP module loaded for epimerization to occur may support a loading mechanism wherein the L-*allo*-Thr tethered to EndD is directly transferred to the PCP of EndB-m6.

Biosynthesis of non-proteinogenic amino acids

The biosynthesis of Hpg has been elucidated by biochemical analysis of recombinant enzymes from the chloroeremomycin pathway (Choroba *et al.*, 2000; Hubbard *et al.*, 2000; van Wageningen *et al.*, 1998). The process requires four gene products: PDH, 4-hydroxymandelic acid synthase (HmaS), 4-hydroxymandelic acid oxidase (HmaO) and 4-hydroxyphenylglycine aminotransferase (HpgT). Orf23 and Orf25 are homologues of PDH and HmaS, respectively, and

orf29 apparently encodes an HpgT/HmaO fusion protein. In addition to *orf23*, *S. fungicidicus* possesses another PDH gene located outside the *end* cluster that is presumably involved in normal aromatic amino acid biosynthesis. A portion of this second PDH gene was amplified by PCR and comparison of the available translated sequence (168 aa) with Orf23 showed that this region shared 61 % identity. Inactivation of the second PDH gene had no effect on enduracidin production (data not shown).

The unusual organization of the *hmaO* and *hpgT* homologues in a single ORF in the *end* cluster was reconfirmed by double-strand sequencing of multiple fosmid templates. HmaO and HpgT catalyse sequential reactions in the generation of Hpg and the fusion protein may lead to more efficient production of Hpg inasmuch as it contributes to 6 of the 17 amino acid residues in enduracidin. In the

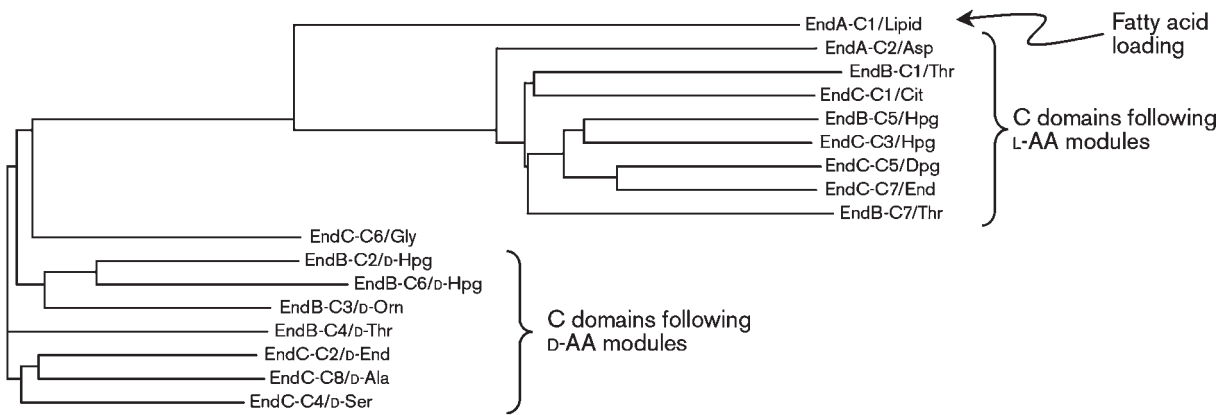


Fig. 5. Phylogram of enduracidin synthetase condensation domains.

chloroeremomycin (van Wageningen *et al.*, 1998), balhimycin (Recktenwald *et al.*, 2002) and teicoplanin (Li *et al.*, 2004) clusters, *hmaS* and *hmaO* are organized in a two-gene operon and, in the CDA cluster, *hmaO* and *hpgT* form an operon (Hojati *et al.*, 2002). Although the genes encoding HpgT and HmaO are adjacent in the ramoplanin cluster, there is no operon organization observed among the HpgT-, HmaO- and HmaS-encoding genes (McCafferty *et al.*, 2002).

Little is known about the formation of the rare D- and L-End residues other than they originate from L-Arg (Hatano *et al.*, 1984). The mannopeptimycins, isolated from a strain of *S. hygroscopicus*, contain D- and L- β -hydroxyenduracididine (β -OHEnd) residues and are the only other peptide antibiotics we are aware of that contain this unusual amino acid (He *et al.*, 2002). Both the mannopeptimycin (*mpp*, accession no. AY735112) and *end* clusters contain a three-gene operon, *mppPQR* and *endPQR*, respectively, that share very high identity (Fig. 6). Given that End and β -OHEnd are the only non-proteinogenic amino acids common to mannopeptimycin and enduracidin, the products of these three genes are proposed to direct End biosynthesis. The gene immediately upstream of the *mppPQR* operon, *mppO*, has a translated product that is homologous to VioC, a non-haem iron, 2-oxoglutarate-dependent oxygenase that catalyses the formation of 2S-hydroxy-L-Arg from L-Arg (Yin & Zabriskie, 2004). We and others have shown that VioC and the pyridoxal phosphate-dependent enzyme VioD act in tandem to form the 2S,3R-capreomycin residue found in the peptide antibiotic viomycin (Ju *et al.*, 2004; Yin *et al.*, 2004). Through gene

disruption experiments and biochemical characterization of recombinant MppO, we have shown that this enzyme hydroxylates the β -position of L-End and does not generate a precursor for cyclization of the Arg side chain (Haltli *et al.*, 2005). The fact that MppO is not involved in L-End formation is consistent with absence of an *mppO* homologue in the *end* cluster.

Questions remain regarding the formation of L-End in both the mannopeptimycin and enduracidin pathways. Routes to enduracididine can be proposed that are analogous to the formation of capreomycin by VioD, and rely on the putative PLP-dependent enzymes EndP and/or EndQ to catalyse elimination/replacement reactions beginning with either β -hydroxy or γ -hydroxyarginine. However, we identified no genes in the sequenced regions of the *S. fungicidicus* chromosome predicted to encode a hydroxylase or similar enzyme. Similarly, the only hydroxylase gene located in the *mpp* cluster is *mppO*. The respective function of EndP and EndQ, and the role of EndR, which exhibits some similarity to acetoacetate decarboxylases, will probably require *in vitro* analysis of the recombinant proteins in combination with *in vivo* analysis of disruption and complementation mutants. Attempts to date to express these genes in *E. coli* have only yielded insoluble protein.

The other two non-proteinogenic amino acids in enduracidin are ornithine and citrulline. Genes directing the synthesis of these compounds were not expected in the cluster inasmuch as these amino acids are normally found in the bacterial cell as intermediates in arginine metabolism.

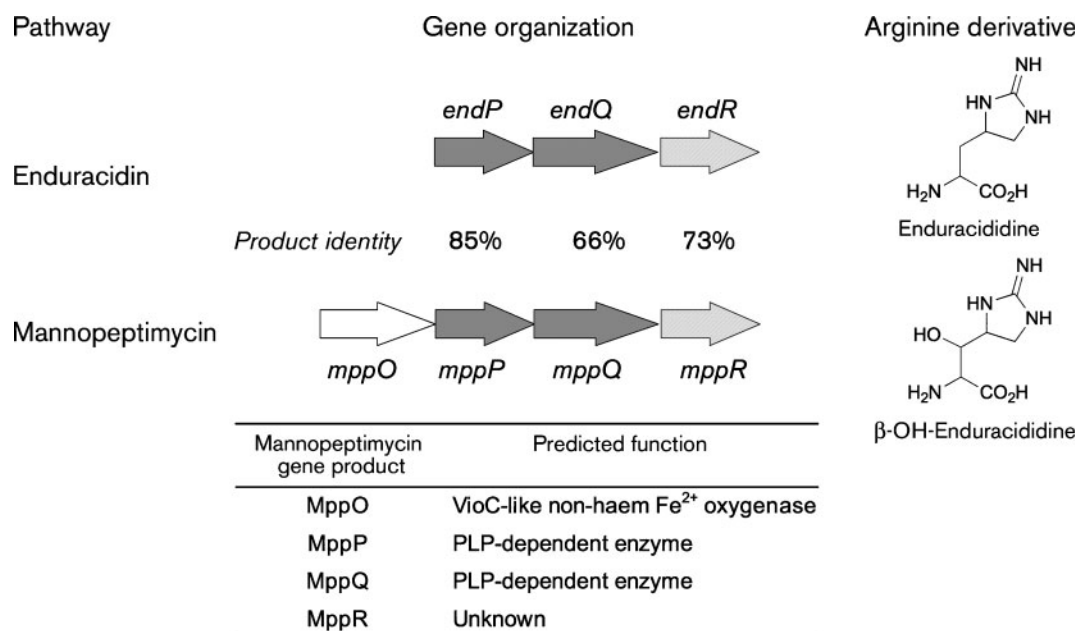


Fig. 6. Relationship and organization of the *mppPQR* and *endPQR* operons.

Formation and attachment of the fatty acid tail

Enduracidins A and B differ only in the fatty acid side chain attached to the starter Asp unit (Iwasaki *et al.*, 1973). Four gene products, Orf35, Orf39, Orf44 and Orf45, are predicted to activate and modify a precursor fatty acid and transfer it to the amine of Asp¹ on EndA (Fig. 7). Counterparts of these proteins are found in the ramoplanin pathway, which incorporates a shorter, but similar, 2*Z*,4*E* unsaturated lipid (cf. Fig. 1 and Table 1). However, Orf45 from the *end* pathway appears to be a fusion protein consisting of an acyl-CoA ligase and an acyl-CoA dehydrogenase. In the ramoplanin cluster *orf26* encodes an acyl-CoA ligase and *orf24* and *orf25* encode homologous acyl-CoA dehydrogenases (McCafferty *et al.*, 2002).

We expect the dehydrogenase function of Orf45 introduces the first double bond in the fatty acid and then Orf39 and/or Orf44 participate in the second dehydrogenation and double-bond isomerization. Recently, an acyl-CoA dehydrogenase in the friulimicin pathway was shown to be involved in the formation of the $\Delta cis3$ double bond in the lipid side chain of that lipopeptide (Heinzelmann *et al.*, 2005). Transfer of the modified fatty acid to the carrier protein Orf35 facilitates acylation of the Asp residue on EndA. It is also possible that Orf35 accepts an intermediate species that is further modified as the carrier protein-bound substrate rather than as a coenzyme A adduct.

Chlorination of 4-hydroxyphenylglycine

The deduced product of *orf30* shows at least 69 % amino acid identity to several halogenases associated with various glycopeptide biosynthesis pathways and is predicted to carry out the chlorination of Hpg¹³ to form 3,5-dichloro-L-4-hydroxyphenylglycine (Dpg) (Li *et al.*, 2004; Pootoolal *et al.*, 2002; Recktenwald *et al.*, 2002; van Wageningen *et al.*, 1998). *orf30* encodes the only halogenase in the *end* cluster, suggesting that both chlorine atoms in the Dpg residue are introduced by a single enzyme. Similarly, complestatin contains two Dpg residues and the gene cluster only contains a single halogenase gene, *comH* (Chiu *et al.*, 2001). In balhimycin biosynthesis, one halogenase introduces two chlorine atoms at different positions of the glycopeptide antibiotic (Puk *et al.*, 2002). PltA, an FADH₂-dependent halogenase involved in pyoluteorin biosynthesis, has been shown to catalyse the dichlorination of a carrier-protein-bound pyrrole species (Dorrestein *et al.*, 2005). These FADH₂-dependent halogenases required an NADH-dependent flavin reductase for cofactor recycling and two-component systems have been characterized that generate 7-chlorotryptophan during rebeccamycin and pyrrolnitrin biosynthesis (Keller *et al.*, 2000; Yeh *et al.*, 2005). A flavin reductase gene has not been identified near the enduracidin biosynthetic genes.

The timing of chlorination has not been determined but the substrate specificity sequence for the A domain of module 13 (EndC-m4), which corresponds to Dpg, is identical to the

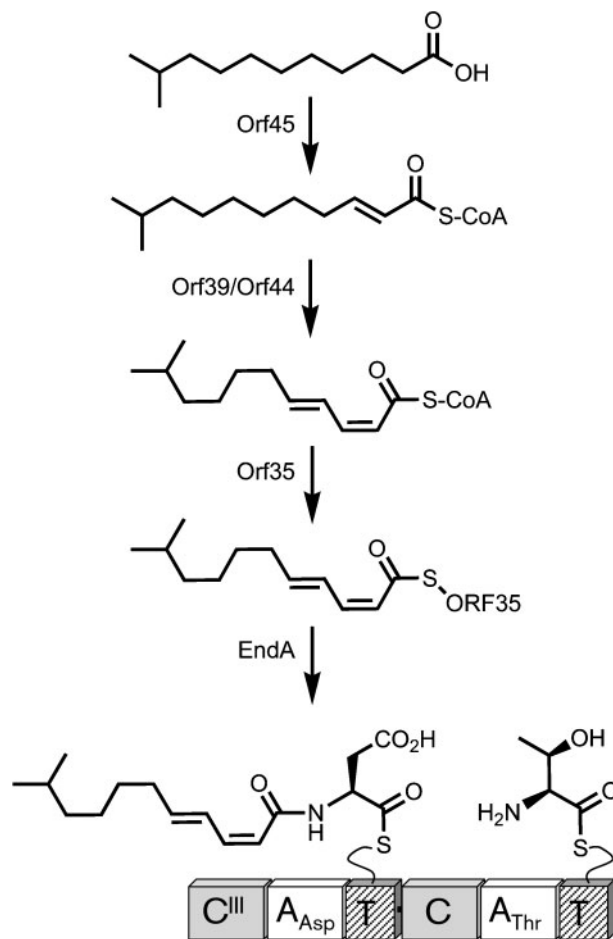


Fig. 7. Proposed activation, modification and attachment of the lipid tail of enduracidin A.

substrate specificity sequences deduced for four A domains predicted to activate Hpg (Table 2). Therefore, it would seem unlikely that free L-Hpg is chlorinated by Orf30 and the resulting Dpg is directly incorporated by the NRPS. Given the lack of additional A domain/carrier protein genes and the fact that EndC-m4 appears to be a functional module, we expect halogenation to occur on an NRPS-bound Hpg or on the nascent peptide.

Genes for self-resistance, regulation and export

The mechanism(s) employed by the enduracidin and ramoplanin producers to protect intracellular Lipid I and Lipid II from being complexed by these peptide antibiotics is still unknown. Self-resistance mechanisms for cell-wall-active glycopeptide antibiotics like vancomycin typically involve alteration of the terminal residue of the pentapeptide moiety of Lipid II (Marshall *et al.*, 1998; Li *et al.*, 2004). Our partial scan of the *S. fungicidicus* genome identified vancomycin-type resistance genes (data not shown), but this mechanism is unlikely to be related to enduracidin

self-resistance because the peptides recognize different regions of Lipid II and enduracidin is active against vancomycin-resistant bacteria. In other lipopeptide biosynthetic gene clusters, such as those for daptomycin, CDA and ramoplanin, the identity of the genes conferring antibiotic immunity also remains unknown. Export of the peptide from the cell is likely to involve Orf31, Orf32 and Orf33, predicted to function as components of ABC transporters similar to those in other antibiotic biosynthetic gene clusters. Regulation of antibiotic production, and possibly self-resistance determinant expression, is believed to involve *orf22*, *orf24*, *orf41*, *orf42* and *orf43*. Orf41 is similar to the LuxR family of DNA-binding proteins that are activated by bacterial autoinducer molecules (Demain, 1998). *orf42* and *orf43* encode elements of two-component sensor kinase systems such as those activating expression of vancomycin resistance genes and *orf22* and *orf24* also exhibit similarity to known regulatory genes in other secondary metabolite clusters. All of these putative regulatory genes have orthologues in the ramoplanin cluster (Table 1).

Boundaries and remaining genes identified in the *end* cluster

The probable boundaries of the *end* cluster were assigned based on the deduced functions of the *end* and flanking gene products and by comparison with the ramoplanin gene cluster. With the possible exception of *orf10* being involved in lipid tail formation, there are no ORFs in the sequenced region upstream of *orf22* with predicted functions necessary for enduracidin formation. Likewise, the region located downstream of *orf46* has no ramoplanin counterparts or genes with a possible function in enduracidin biosynthesis. Therefore, we propose the boundaries of the enduracidin gene cluster are defined by *orf22* and *orf46*. Whether any of the genes in the flanking regions that encode products of unknown function are involved in antibiotic regulation or resistance remains to be determined.

Conclusion

Enduracidin and the closely related ramoplanin are members of a novel class of peptide antibiotics with promising activity against serious infections caused by Gram-positive pathogens. The unique mechanism of action shared by these peptides results in a lack of cross-resistance with other known classes of bacterial cell-wall inhibitors and makes the development of improved analogues an attractive pursuit. The sequenced *end* cluster offers opportunities to study the formation of the rare amino acid enduracididine, will allow studies on the unknown mechanism of enduracidin self-resistance and provides the means to explore the scope and limitations of using dual action C/E domains in actinomycete NRPS systems. Importantly, knowledge of the sequence and organization of the *end* cluster and the demonstrated ability to genetically manipulate *S. fungicidicus* should accelerate the development of novel analogues through a combined genetic and chemical approach.

ACKNOWLEDGEMENTS

We are grateful to Dr Ling Zhang for excellent technical support and to Professor Keith F. Chater for providing plasmid pSET152. This work was supported in part by National Institutes of Health grant GM69320 (M.Z.) and by funds from the Oregon State University College of Pharmacy Research and Scholarship Fund, the OSU General Research Fund and the Medical Research Foundation of Oregon (X.Y.).

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
- Balibar, C. J., Vaillancourt, F. H. & Walsh, C. T. (2005). Generation of D-amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem Biol* **12**, 1189–1200.
- Barrett, J. (2005). Recent developments in glycopeptide antibacterials. *Curr Opin Invest Drugs* **6**, 781–790.
- Bierman, M. R., Logan, R., O'Brien, K., Seno, E. T., Rao, R. N. & Shoner, B. E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **116**, 43–49.
- Challis, G. L., Ravel, J. & Townsend, C. A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* **7**, 211–224.
- Chiu, H. T., Hubbard, B. K., Shah, A. N., Eide, J., Fredenburg, R. A., Walsh, C. T. & Khosla, C. (2001). Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc Natl Acad Sci U S A* **98**, 8548–8553.
- Choroba, O. W., Williams, D. H. & Spencer, J. B. (2000). Biosynthesis of the vancomycin group of antibiotics: involvement of an unusual dioxygenase in the pathway to (S)-4-hydroxyphenylglycine. *J Am Chem Soc* **122**, 5389–5390.
- Cudic, P., Kranz, J. K., Behenna, D. C., Kruger, R. G., Tadesse, H., Wand, A. J., Veklich, Y. I., Weisel, J. W. & McCafferty, D. G. (2002). Complexation of peptidoglycan intermediates by the lipoglycopeptide antibiotic ramoplanin: minimal structural requirements for intermolecular complexation and fibril formation. *Proc Natl Acad Sci U S A* **99**, 7384–7389.
- Demain, A. L. (1998). Induction of microbial secondary metabolism. *Int Microbiol* **1**, 259–264.
- Dorrestein, P. C., Yeh, E., Garneau-Tsodikova, S., Kelleher, N. L. & Walsh, C. T. (2005). Dichlorination of a pyrrolyl-S-carrier protein by FADH₂-dependent halogenase PltA during pyoluteorin biosynthesis. *Proc Natl Acad Sci U S A* **102**, 13843–13848.
- Duitman, E. H., Hamoen, L. W., Rembold, M. & 10 other authors (1999). The mycosubtilin synthetase of *Bacillus subtilis* ATCC 6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc Natl Acad Sci U S A* **96**, 13294–13299.
- Fang, X., Tiyanont, K., Zhang, Y., Vanner, J., Boger, D. & Walker, S. (2006). The mechanism of action of ramoplanin and enduracidin. *Mol Biosyst* **2**, 69–76.
- Goto, S., Kuwahara, S., Okubo, N. & Zenyogi, H. (1968). *In vitro* and *in vivo* evaluation of enduracidin, a new peptide antibiotic substance. *J Antibiot* **21**, 119–125.
- Guenzi, E., Galli, G., Grgurina, I., Gross, D. C. & Grandi, G. (1998). Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J Biol Chem* **273**, 32857–32863.
- Haltli, B., Tan, Y., Magarvey, N. A., Wagenaar, M., Yin, X., Greenstein, M., Hucul, J. A. & Zabriskie, T. M. (2005).

Investigating beta-hydroxyenduracididine formation in the biosynthesis of the mannopeptimycins. *Chem Biol* **12**, 1163–1168.

Hatano, K., Nogami, I., Higashide, E. & Kishi, T. (1984). Biosynthesis of enduracidin: origin of enduracididine and other amino acids. *Agric Biol Chem* **48**, 1503–1508.

He, H., Williamson, R. T., Shen, B., Graziani, E. I., Yang, H. Y., Sakya, S. M., Petersen, P. J. & Carter, G. T. (2002). Mannopeptimycins, novel antibacterial glycopeptides from *Streptomyces hygroscopicus*, LL-AC98. *J Am Chem Soc* **124**, 9729–9736.

Heinzelmann, E., Berger, S., Muller, C., Hartner, T., Poralla, K., Wohlleben, W. & Schwartz, D. (2005). An acyl-CoA dehydrogenase is involved in the formation of the $\Delta cis3$ double bond in the acyl residue of the lipopeptide antibiotic friulimycin in *Actinoplanes friuliensis*. *Microbiology* **151**, 1963–1974.

Higashide, E., Hatano, K., Shibata, M. & Nakazawa, K. (1968). Enduracidin, a new antibiotic. I. *Streptomyces fungicidicus* No. B5477, an enduracidin producing organism. *J Antibiot* **21**, 126–137.

Hojati, Z., Milne, C., Harvey, B. & 9 other authors (2002). Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem Biol* **9**, 1175–1187.

Hori, M., Iwasaki, H., Horii, S., Yoshida, I. & Hongo, T. (1973). Enduracidin, a new antibiotic. VII. Primary structure of the peptide moiety. *Chem Pharm Bull* **21**, 1175–1183.

Hubbard, B. K., Thomas, M. G. & Walsh, C. T. (2000). Biosynthesis of L-p-hydroxyphenylglycine, a non-proteinogenic amino acid constituent of peptide antibiotics. *Chem Biol* **7**, 931–942.

Ishikawa, J., Yamashita, A., Mikami, Y., Hoshino, Y., Kurita, H., Hotta, K., Shiba, T. & Hattori, M. (2004). The complete genomic sequence of *Nocardia farcinica* IFM 10152. *Proc Natl Acad Sci U S A* **101**, 14925–14930.

Iwasaki, H., Horii, S., Asai, M., Mizuno, K., Ueyanagi, J. & Miyake, A. (1973). Enduracidin, a new antibiotic. VIII. Structures of enduracidin A and B. *Chem Pharm Bull* **21**, 1184–1191.

Ju, J., Ozanick, S. G., Shen, B. & Thomas, M. G. (2004). Conversion of (2S)-arginine to (2S,3R)-capreomycinide by VioC and VioD from the viomycin biosynthetic pathway of *Streptomyces* sp. strain ATCC 11861. *Chembiochem* **5**, 1281–1285.

Kawakami, M., Nagai, Y., Fujii, T. & Mitsunashi, S. (1971). Antimicrobial activities of enduracidin (enramycin) *in vitro* and *in vivo*. *J Antibiot* **24**, 583–586.

Keller, S., Wage, T., Hohaus, K., Holzer, M., Eichhorn, E. & van Pee, K. H. (2000). Purification and partial characterization of tryptophan 7-halogenase (PrnA) from *Pseudomonas fluorescens*. *Angew Chem Int Ed Engl* **39**, 2300–2302.

Kieser, T., Bibb, M. J., Buttner, M. J., Charter, K. F. & Hopwood, D. A. (2000). *Practical Streptomyces Genetics*. Norwich: John Innes Centre.

Komatsuzawa, H., Suzuki, J., Sugai, M., Miyake, Y. & Suganaka, H. (1994). Effect of combination of oxacillin and non-beta-lactam antibiotics on methicillin-resistant *Staphylococcus aureus*. *J Antimicrob Chemother* **33**, 1155–1163.

Li, T. L., Huang, F., Haydock, S. F., Mironenko, T., Leadlay, P. F. & Spencer, J. B. (2004). Biosynthetic gene cluster of the glycopeptide antibiotic teicoplanin: characterization of two glycosyltransferases and the key acyltransferase. *Chem Biol* **11**, 107–119.

Linne, U., Doekel, S. & Marahiel, M. A. (2001). Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases. *Biochemistry* **40**, 15824–15834.

Mangili, A., Bica, I., Snyderman, D. R. & Hamer, D. H. (2005). Daptomycin-resistant, methicillin-resistant *Staphylococcus aureus* bacteremia. *Clin Infect Dis* **40**, 1058–1060.

Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem Rev* **97**, 2651–2674.

Marshall, C. G., Lessard, I. A., Park, I. & Wright, G. D. (1998). Glycopeptide antibiotic resistance genes in glycopeptide-producing organisms. *Antimicrob Agents Chemother* **42**, 2215–2220.

McCafferty, D. G., Cudic, P., Frankel, B. A., Barkallah, S., Kruger, R. G. & Li, W. (2002). Chemistry and biology of the ramoplanin family of peptide antibiotics. *Biopolymers* **66**, 261–284.

Meka, V. G., Pillai, S. K., Sakoulas, G. & 7 other authors (2004). Linezolid resistance in sequential *Staphylococcus aureus* isolates associated with a T2500A mutation in the 23S rRNA gene and loss of a single copy of rRNA. *J Infect Dis* **190**, 311–317.

Miao, V., Coeffet-LeGal, M. F., Brian, P. & 10 other authors (2005). Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* **151**, 1507–1523.

Miao, V., Brost, R., Chapple, J., She, K., Gal, M. F. & Baltz, R. H. (2006). The lipopeptide antibiotic A54145 biosynthetic gene cluster from *Streptomyces fradiae*. *J Ind Microbiol Biotechnol* **33**, 129–140.

Paulsen, I. T., Press, C. M., Ravel, J. & 26 other authors (2005). Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat Biotechnol* **23**, 873–878.

Peromet, M., Schoutens, E. & Yourassowsky, E. (1973). Clinical and microbiological study of enduracidin in infections due to methicillin-resistant strains of *Staphylococcus aureus*. *Chemotherapy* **19**, 53–61.

Pootoolal, J., Thomas, M. G., Marshall, C. G., Neu, J. M., Hubbard, B. K., Walsh, C. T. & Wright, G. D. (2002). Assembling the glycopeptide antibiotic scaffold: the biosynthesis of A47934 from *Streptomyces toyocaensis* NRRL15009. *Proc Natl Acad Sci U S A* **99**, 8962–8967.

Puk, O., Huber, P., Bischoff, D., Recktenwald, J., Jung, G., Sussmuth, R. D., van Pee, K. H., Wohlleben, W. & Pelzer, S. (2002). Glycopeptide biosynthesis in *Amycolatopsis mediterranei* DSM5908: function of a halogenase and a haloperoxidase/perhydrolase. *Chem Biol* **9**, 225–235.

Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D. H. (2005). Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res* **33**, 5799–5808.

Recktenwald, J., Shawky, R., Puk, O., Pfennig, F., Keller, U., Wohlleben, W. & Pelzer, S. (2002). Nonribosomal biosynthesis of vancomycin-type antibiotics: a heptapeptide backbone and eight peptide synthetase modules. *Microbiology* **148**, 1105–1118.

Roongsawang, N., Hase, K., Haruki, M., Imanaka, T., Morikawa, M. & Kanaya, S. (2003). Cloning and characterization of the gene cluster encoding arthrofactin synthetase from *Pseudomonas* sp. MIS38. *Chem Biol* **10**, 869–880.

Sambrook, J. & Russell, D. V. (2001). *Molecular Cloning: a Laboratory Manual*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Scholz-Schroeder, B. K., Soule, J. D. & Gross, D. C. (2003). The *sypA*, *sypS*, and *sypC* synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptin by *Pseudomonas syringae* pv. *syringae* B301D. *Mol Plant Microbe Interact* **16**, 271–280.

Sosio, M., Stinch, S., Beltrametti, F., Lazzarini, A. & Donadio, S. (2003). The gene cluster for the biosynthesis of the glycopeptide antibiotic A40926 by *Nonomuraea* species. *Chem Biol* **10**, 541–549.

- Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6**, 493–505.
- Tenover, F. C. & McDonald, L. C. (2005). Vancomycin-resistant staphylococci and enterococci: epidemiology and control. *Curr Opin Infect Dis* **18**, 300–305.
- Trauger, J. W. & Walsh, C. T. (2000). Heterologous expression in *Escherichia coli* of the first module of the nonribosomal peptide synthetase for chloroeremomycin, a vancomycin-type glycopeptide antibiotic. *Proc Natl Acad Sci U S A* **97**, 3112–3117.
- Tsiodras, S., Gold, H. S., Sakoulas, G., Eliopoulos, G. M., Wennersten, C., Venkataraman, L., Moellering, R. C. & Ferraro, M. J. (2001). Linezolid resistance in a clinical isolate of *Staphylococcus aureus*. *Lancet* **358**, 207–208.
- Tsuchiya, K., Kondo, M., Oishi, T. & Yamazaki, I. (1968). Enduracidin, a new antibiotic. 3. *In vitro* and *in vivo* antimicrobial activity. *J Antibiot* **21**, 147–153.
- Vaillancourt, F. H., Yeh, E., Vosburg, D. A., O'Connor, S. E. & Walsh, C. T. (2005). Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **436**, 1191–1194.
- van Wageningen, A. M., Kirkpatrick, P. N., Williams, D. H., Harris, B. R., Kershaw, J. K., Lennard, N. J., Jones, M., Jones, S. J. & Solenberg, P. J. (1998). Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem Biol* **5**, 155–162.
- Walker, S., Chen, L., Hu, Y., Rew, Y., Shin, D. & Boger, D. L. (2005). Chemistry and biology of ramoplanin: a lipoglycopeptide with potent antibiotic activity. *Chem Rev* **105**, 449–476.
- Wright, F. & Bibb, M. J. (1992). Codon usage in the G+C-rich *Streptomyces* genome. *Gene* **113**, 55–65.
- Yeh, E., Garneau, S. & Walsh, C. T. (2005). Robust *in vitro* activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc Natl Acad Sci U S A* **102**, 3960–3965.
- Yin, X. & Zabriskie, T. M. (2004). VioC is a non-heme iron, α -ketoglutarate dependent oxygenase that catalyzes the formation of 3S-hydroxy-L-arginine during viomycin biosynthesis. *Chembiochem* **5**, 1274–1277.
- Yin, X., O'Hare, T., Gould, S. J. & Zabriskie, T. M. (2003). Identification and cloning of genes encoding viomycin biosynthesis from *Streptomyces vinaceus* and evidence for involvement of a rare oxygenase. *Gene* **312**, 215–224.
- Yin, X., McPhail, K. L., Kim, K. J. & Zabriskie, T. M. (2004). Formation of the nonproteinogenic amino acid 2S,3R-capreomycinidine by VioD from the viomycin biosynthesis pathway. *Chembiochem* **5**, 1278–1281.
- Yourassowsky, E. & Monsieur, R. (1972). *In vitro* and *in vivo* activity of enduracidin on *Staphylococcus aureus*. *Chemotherapy* **17**, 182–187.
- Zinner, S. H. (2005). The search for new antimicrobials: why we need new options. *Expert Rev Anti Infect Ther* **3**, 907–913.