A complex multienzyme system encoded by five polyketide synthase genes is involved in the biosynthesis of the 26-membered polyene macrolide pimaricin in *Streptomyces natalensis*

Jesús F Aparicio [1,2], Roberto Fouces [1], Marta V Mendes [1], Noemí Olivera [1] and Juan F Martín [1,2]

Background: Polyene macrolides are a class of large macrocyclic polyketides that interact with membrane sterols, having antibiotic activity against fungi but not bacteria. Their rings include a chromophore of 3–7 conjugated double bonds which constitute the distinct polyene structure. Pimaricin is an archetype polyene, important in the food industry as a preservative to prevent mould contamination of foods, produced by *Streptomyces natalensis*. We set out to clone, sequence and analyse the gene cluster responsible for the biosynthesis of this tetraene.

Results: A large cluster of 16 open reading frames spanning 84 985 bp of the *S. natalensis* genome has been sequenced and found to encode 13 homologous sets of enzyme activities (modules) of a polyketide synthase (PKS) distributed within five giant multienzyme proteins (PIMS0–PIMS4). The total of 60 constituent active sites, 25 of them on a single enzyme (PIMS2), make this an exceptional multienzyme system. Eleven additional genes appear to govern modification of the polyketide-derived framework and export. Disruption of the genes encoding the PKS abolished pimaricin production.

Conclusions: The overall architecture of the PKS gene cluster responsible for the biosynthesis of the 26-membered polyene macrolide pimaricin has been determined. Eleven additional tailoring genes have been cloned and analysed. The availability of the PKS cluster will facilitate the generation of designer pimaricins by combinatorial biosynthesis approaches. This work represents the extensive description of a second polyene macrolide biosynthetic gene cluster after the one for the antifungal nystatin.

[1]Institute of Biotechnology INBIOTEC, Parque Científico de León, Avda. del Real, n° 1, 24006 León, Spain
[2]Area of Microbiology, Faculty of Biology, University of León, 24071 León, Spain

Correspondence: Jesús F Aparicio
E-mail: degjfa@unileon.es

Keywords: Macrolide; Pimaricin; Polyene; Polyketide synthase; *Streptomyces natalensis*

Received: **6 July 2000**
Revisions requested: **11 August 2000**
Revisions received: **11 September 2000**
Accepted: **21 September 2000**

Published: **6 October 2000**

Chemistry & Biology 2000, **7**:895–905

1074-5521/00/$ – see front matter

PII: S1074-5521(00)00038-7

Introduction

Polyenes antibiotics constitute a large group of antifungal agents produced mainly by *Streptomyces* and closely related bacteria. Structurally they are characterised by a hydroxylated macrocycle containing one or more sugars, but their distinct characteristic is the presence of a chromophore formed by a system of 3–7 conjugated double bonds in the macrolactone ring (see [1] for a review). They are synthesised by the pathway common to all polyketides, in which units derived from acetate, propionate or butyrate are condensed onto the growing chain by a polyketide synthase (PKS), in a manner that is conceptually similar to the biosynthesis of the long-chain fatty acids found in all organisms [2–4], except that the β-keto function introduced at each elongation step may undergo all, part or none of a reductive cycle comprising β-ketoreduction, dehydration and enoylreduction [5]. As macrocyclic polyketides they are produced through the action of so-called type I modular PKSs [6,7]. These enzymes usually consist of several extremely large polypeptides in which different modules (sets) of enzymatic activities catalyse each successive round of elongation [8,9].

The structural variation between naturally occurring polyenes arises, as in other polyketide macrolides, largely from the way in which the PKS controls the number and type of starter and extender units used [10–12], and from the extent and stereochemistry of reduction at each cycle [4,13]. Yet further diversity is produced by functionalisation of the polyketide chain by the action of glycosylases, methyltransferases and oxidative enzymes.

Pimaricin represents an archetype molecule of glycosylated polyenes [14] important for antifungal therapy and promising for antiviral activity, stimulation of the immune response, and action in synergy with other antifungal or antitumour compounds [15]. It is produced by *Streptomyces natalensis* and widely used in the food industry in order to prevent mould contamination of cheese and other non-sterile foods (i.e. cured meats). Initial studies showed

that its synthesis requires a complex PKS, which is encoded by two subclusters [7]. Here, we report the complete sequence and analysis of the PKS genes responsible for the biosynthesis of this 26-membered macrolide tetraene from *S. natalensis*. The sequenced region of 84 985 bp encodes 13 PKS modules within five multifunctional enzymes, and 11 additional proteins that presumably govern post-PKS modification of the polyketide skeleton, export and regulation of gene expression. This is the second example of an extensive DNA sequence analysis of a polyene biosynthetic gene cluster that shows a complex organisation related to that of nystatin [16].

Results

Identification and sequencing of the pimaricin synthase cluster

The pimaricin biosynthetic gene cluster was identified by genomic walking using a *S. natalensis* ATCC 27448 cosmid library and DNA segments from *pimS0* [7]. A total of 85 kb of contiguous DNA was sequenced, and the deduced gene organisation within this region is shown in Figure 1. The G+C content of the nucleotide sequence is 72.8%, well within the range of the reference values for *Streptomyces* DNA [17].

Structural organisation of the pimaricin biosynthetic gene cluster

Computer-assisted analysis of the sequenced region revealed 16 potential open reading frames (ORFs), including five PKS genes (namely *pimS0–pimS4*) (Figure 1). The putative functions of the deduced gene products of the sequenced ORFs are listed in Table 1.

The 16 genes putatively involved in pimaricin biosynthesis appear to be organised into a number of operons, although this will require confirmation by transcriptional analysis. In the centre of the pimaricin biosynthetic cluster, proximate to the previously characterised *pimS0* and *pimS1* genes [7], and separating these two PKS genes from a second set of PKS genes (*pimS2–pimS4*), lie eight 'small' genes spanning 10.5 kb (*pimABC*, *pimEFG*, *pimIJ*). All of them have a plausible role in pimaricin biosynthesis (Table 1), and are distributed in at least two transcriptional units as deduced from their divergent orientation (Figure 1).

pimA and *pimB* may be translationally coupled as judged by their overlapping coding sequences (*pimB* start codon is located 23 bp upstream from the *pimA* stop codon) and the presence of the ribosome-binding site (GAGG) 7 bp upstream (also internal to *pimA* coding sequence). No apparent transcriptional terminator could be predicted in the short intergenic regions between *pimA*, *pimB*, *pimE*, *pimC*, *pimG*, *pimF* and *pimS0* (sizes of the intergenic regions are 23 bp, 14 bp, 88 bp, 28 bp, 52 bp and 79 bp, respectively), suggesting that these genes are likely to form an operon resulting in a transcript over 13 600 bp. The lack of pimaricin production in mutants disrupted by phage-mediated integration through homologous recombination of an internal fragment of this region (M.V.M., J.F.M., J.F.A., unpublished observations) supports this idea.

Divergently transcribed to these genes are *pimJ*, *pimI*, the giant *pimS2* (see below), *pimS3* and *pimS4*, which together are possibly responsible for the existence of a third polycistronic mRNA as deduced by the lack of obvious terminators between the predicted 3′ end of the upstream gene and the 5′ end of the downstream gene. The gaps between genes are of 217 bp (*pimJ* and *pimI*), 190 bp (*pimI* and *pimS2*), 11 bp (*pimS2* and *pimS3*) or 35 bp (*pimS3* and *pimS4*). It therefore seems that an operon of exceptional

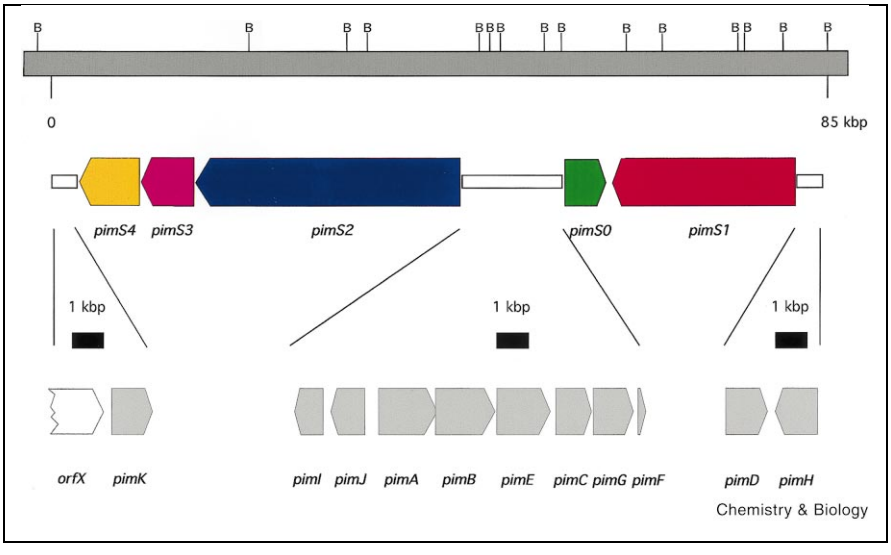

Figure 1. Organisation of the gene cluster for pimaricin biosynthesis. The transcriptional direction and the relative sizes of the predicted ORFs are indicated by pointed boxes. ORFs corresponding to the PKS are indicated in colour. Additional ORFs of putative pimaricin tailoring, regulation and resistance functions are indicated in grey. The 3′ end of an unidentified ORF (*orfX*) is indicated in white. Only *Bam*HI restriction sites (B) are indicated.

Table 1
Deduced functions of ORFs in the pimaricin biosynthetic gene cluster.

Polypeptide	Amino acids	Proposed function or sequence similarities detected					
PIMS0	1847	PKS					
Loading		CoL				ACP	
Module 0		KS*	ATa			ACP	
PIMS1	6797	PKS					
Module 1		KS	ATa		KR	ACP	
Module 2		KS	ATa	DH	KR	ACP	
Module 3		KS	ATa	DH	KR	ACP	
Module 4		KS	ATa	DH	KR	ACP	
PIMS2	9507	PKS					
Module 5		KS	ATa	DH	KR	ACP	
Module 6		KS	ATa		KR	ACP	
Module 7		KS	ATp		KR	ACP	
Module 8		KS	ATa		KR	ACP	
Module 9		KS	ATa		KR*	ACP	
Module 10		KS	ATa		KR	ACP	
PIMS3	1808	PKS					
Module 11		KS	ATa	DH	KR	ACP	
PIMS4	2024	PKS					
Module 12		KS	ATa	DH	KR	ACP	TE
PimA	602	ABC transporter					
PimB	626	ABC transporter					
PimC	352	Aminotransferase					
PimD	397	Cytochrome P-450 monooxygenase					
PimE	549	Cholesterol oxidase					
PimF	63	Ferredoxin					
PimG	398	Cytochrome P-450 monooxygenase					
PimH	432	Efflux pump					
PimI	255	TE					
PimJ	343	Sugar DH					
PimK	458	Glycosyl transferase					
OrfX	>532[a]	None					

Potential PKS catalytic activities are abbreviated as in Figure 2. ATa, AT incorporating an acetate extender unit; ATp, AT incorporating a propionate extender unit. Activities listed for PIMS proteins are colinear with the sequence of each ORF as listed from left to right and from top to bottom. Asterisks indicate that the enzyme activity is predicted to be non-functional (see text).
[a]The start codon of *orfX* encoding OrfX lies outside the sequenced region.

size (more than 42 kb) could be derived from this region.

The cluster is completed with a gene on the left end (*pimK*), and three on the right end (*pimS1*, *pimD* and *pimH*), all plausibly involved in pimaricin biosynthesis. The predicted start of *pimK* is separated by only 95 bp from the 3′ end of an unidentified ORF (*orfX)*, so these genes may also be co-transcribed, in a convergent manner to *pimS4*. The three remaining genes to complete the cluster *pimS1*, *pimD* and *pimH* seem to be transcribed as single transcripts (Figure 1).

Pimaricin ring formation: the PKS

10 567 bp upstream from *pimS0*, and divergently transcribed, we found a second set of PKS genes starting with a giant ORF named *pimS2*, encoding a polypeptide of 9507 amino acids (aa), including the initiating methionine. The molecular mass of the protein encoded by *pimS2* (hereafter named PIMS2) was calculated to be 994 390 Da as a monomer. Comparison of PIMS2 with sequences of the SWISSPROT protein sequence database revealed significant similarity to known fatty acid and PKSs. When the sequence of the *pimS2* gene product was compared with itself, clear evidence was obtained for internal reiterations, indicating that PIMS2 consists of six modules for chain extension (not shown). The first module of PIMS2 (module 5 of the pimaricin PKS) contains β-ketoacyl ACP synthase (KS), acyltransferase (AT), dehydratase (DH), β-ketoreductase (KR), and acyl carrier protein (ACP) domains that account for the last double bond of the chromophore in the final pimaricin molecule (Figure 2). Modules 6–10, however, lack the DH domain. Therefore they can only process the β-carbon up to a hydroxyl group in the polyketide chain (Figure 2). Strikingly, module 9, despite the presence of a KR domain, seems to be inactive (see below) thus the β-keto group would be left unreduced. This is in agreement with the initial presence at C9 of a carbonyl residue to form the hemiketal ring at C9–C13 by internal ketalisation between the C9 carbonyl and the C13 hydroxyl generated by the KR domain in module 7 during chain elongation by acyl condensation.

Eleven bp downstream from the TGA stop codon of *pimS2* lies the ATG start codon of *pimS3*, an ORF encoding a

protein of 1808 aa. PIMS3 consists of a single PKS module with KS, AT, DH, KR and ACP domains (Figure 1) responsible for the incorporation of carbons 3 and 4 and the introduction of a double bond that would undergo later oxidation to form the epoxy function of the aglycone (Figure 2).

Located 35 bp downstream the stop codon of *pimS3* is the last ORF needed to complete the PKS genes required for the biosynthesis of the polyketide backbone of pimaricin, *pimS4*. It encodes a protein of 2024 aa, with a molecular mass estimated to be 212 174 Da as a monomer. PIMS4 showed extensive homology to PIMS3 (it houses the same domains); however, in its C-terminal end it displays a thioesterase (TE) domain, a fashion already observed in other type I systems such as the erythromycin [18], the avermectin [19], or the epothilone [20,21] PKSs.

The formation of pimaricin is predicted to start on PIMS0, a single module PKS with an N-terminal ATP-dependent carboxylic acid:CoA ligase domain (CoL; Figure 2), and continue on PIMS1 (four modules) with the formation of

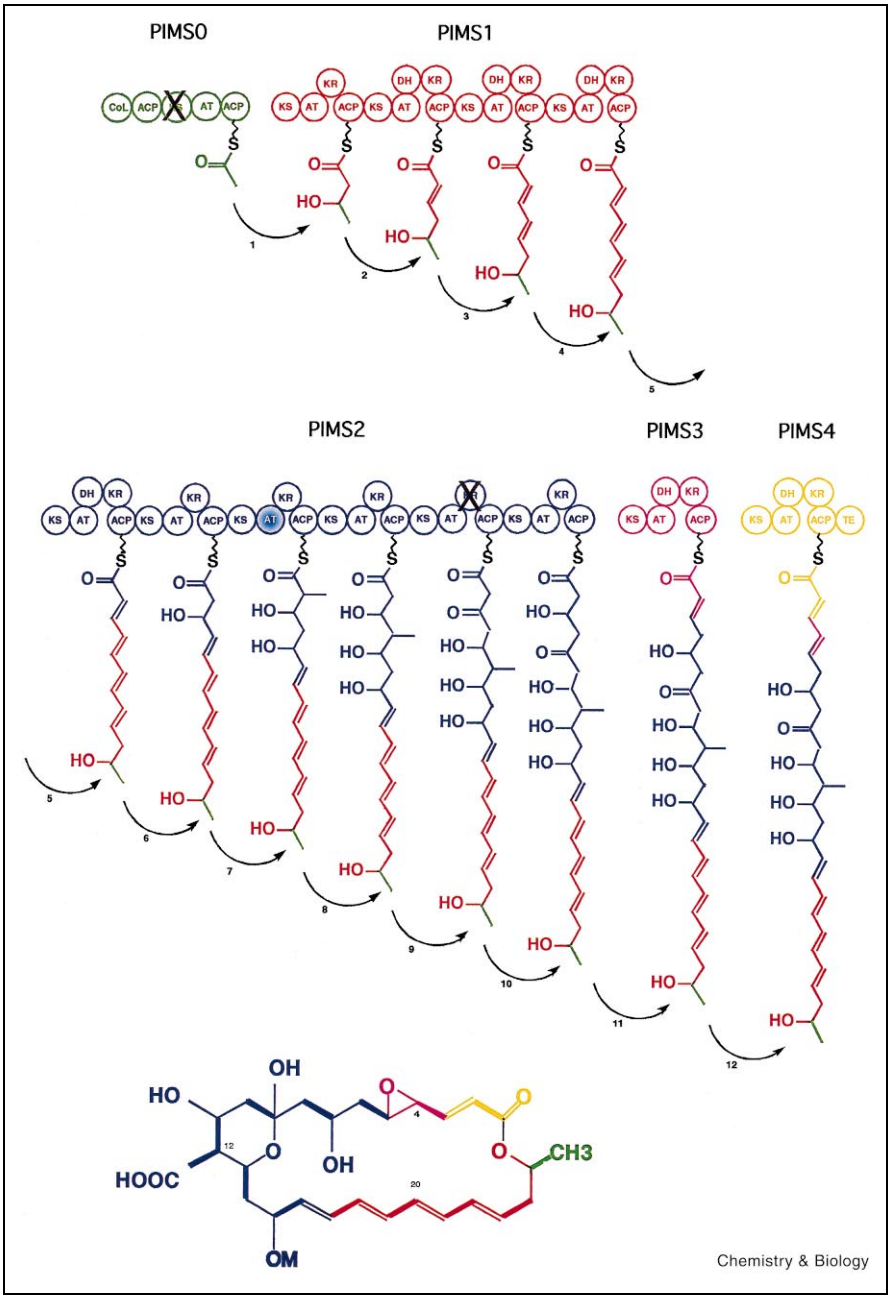

Figure 2. Predicted domain organisation and biosynthetic intermediates of the pimaricin synthase. Each circle represents an enzymatic domain. ACP, acyl carrier protein; AT, acyltransferase; CoL, carboxylic acid:CoA ligase; DH, β-hydroxyacyl-thioester dehydratase; KR, β-ketoacyl ACP reductase; KS, β-ketoacyl ACP synthase; TE, thioesterase. Crossed out domains are predicted to be non-functional (see text). The AT in module 7 (shaded) is predicted to incorporate a propionate extender unit (see text). Colour code is as follows: PIMS0 and product in green, PIMS1 and product in red, PIMS2 and product in blue, PIMS3 and product in pink, and PIMS4 and product in orange. Lactonisation of the acyl chain between C1 and C25 results in the formation of the pimaricinolide ring. Mycosamine (M) is attached to the hydroxyl at C15 in the pimaricin molecule (bottom). Bold lines indicate the building units.

most of the chromophore (three out of four conjugated double bonds [7]). After the first four elongation steps carried out by PIMS1, the chain is transferred to PIMS2 for subsequent elongation. PIMS2 (six modules), and subsequently PIMS3 (one module) and PIMS4 (one module) catalyse the incorporation of eight further carboxylic acid building blocks and their appropriate modifications to generate the skeleton of pimaricin (Figure 2). Finally, and by analogy with the role of the TE domain in the last module of the erythromycin PKS [22], PIMS4 TE would be involved in the release of the completed polyketide from the enzyme as a macrocyclic lactone.

The involvement of *pimS0* and *pimS1* on pimaricin biosynthesis was established in a previous work [7]. Phage-mediated gene replacement was used for the functional inactivation of *pimS2* on the left-hand PKS cluster. Mutants generated in this way showed no production of pimaricin (not shown), indicating that *pimS2* is strictly required for pimaricin biosynthesis. Conversely to the results obtained from the disruption of the rifamycin PKS genes [23], no intermediates of the polyketide assembly could be detected from the mutants with disruption in the pimaricin PKS genes.

Pimaricin ring formation: minimal PKS domains

The minimum of activities required for a PKS to be functional includes the KS, AT and ACP functionalities [24], thus allowing polyketide chain elongation, while the rest of activities that can be found on modular PKSs (i.e. DH, enoylreductase and KR domains) only contribute to increase molecular diversity [25]. KS domains of type I PKSs are the most highly conserved (above 40% identity) of all the constituent domains in these multifunctional enzymes [26–28]. Thus, PIMS KS domains exhibit 39–83% identity to other type I domains from the databanks, being most similar to the rapamycin [5], oleandomycin [29] and erythromycin [26] KSs. The identities amongst PIMS KSs are extraordinarily high, in the range 61–88%, and all contain the two invariant histidine residues located around 135 and 173 aa carboxy-terminal of the active-site cysteine [28]. The highly conserved His340 of KAS II from *Escherichia coli*, corresponding to the last histidine in the KSs of PKSs, has been implicated as a catalytic basic residue [30].

Strikingly, KS from module 0, the loading module of the pimaricin synthase, lacks the active-site cysteine residue, showing a serine instead. This KS has been suggested to be inactive [7], however, it maintains the rest of features typical of type I KS domains. Modified KS domains have been observed in several loading modules of other modular PKSs. 'KSQ domains', in which the active-site cysteine is replaced by a glutamine [31,32] resembling the so-called 'chain length factors' of type II PKSs, have been implicated in polyketide chain initiation by decarboxylation of malonyl-ACP in both systems [33,34]. Whether the serine-containing KSs in PIMS0 and the presumed loading module of the nystatin synthase [16], or the tyrosine-containing loading module KS of the epothilone synthase [20,21] serve a similar function remains to be determined.

AT domains of type I modular PKSs fall into two distinct categories depending upon whether their substrate is malonyl-CoA (or in the case of some ATs in 'loading modules', acetyl-CoA), or the longer-chain 2S-methylmalonyl-CoA [11], as discussed in detail elsewhere [12]. The AT domains of the pimaricin PKS can also be classified into two groups using the above criteria. All PIMS ATs belong to the class of acetate extenders except for module 7 AT that incorporates propionate (Table 1; Figure 2) and is thought to be responsible for the incorporation of carbons 11, 12 and the exocyclic methyl group that would undergo later oxidation to form the free carboxyl function of the aglycone [7].

ACP domains in PIMS are also highly conserved. Remarkably, several ACPs from the pimaricin PKS show deviations from the signature LGXDS, the active-site sequence of prokaryotic ACPs. These include the C-terminal ACP of PIMS0 (MGINS), and the ACPs from module 1 (QGFDS), module 6 (AGFDS), and module 7 (VGFDS).

Pimaricin ring formation: reductive and DH activities and chain termination

Apart from module 0, which is unusual in having a direct linkage between AT and ACP domains, all of the pimaricin PKS modules contain KR or DH activities predicted to modify the growing polyketide chain. All PIMS KR domains contain, at the N-terminal end of the domain, sequences that fit the consensus motif GxGxxGxxxA for NADP(H)-binding [35], showing in some cases conservative substitutions like in module 5 (AtGglGrvlA) or in

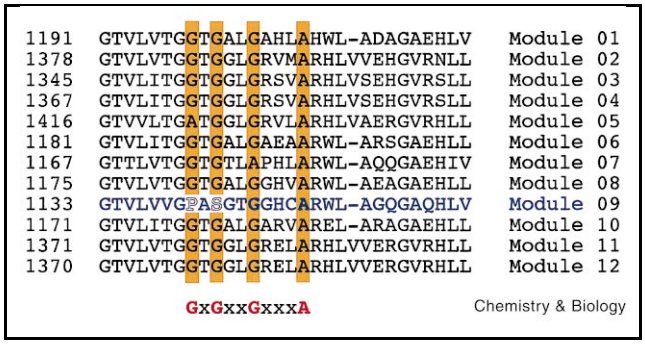

Figure 3. Comparison of the NADP(H)-binding sites of KR domains from the 12 pimaricin PKS modules with such domains. The consensus sequence for NADP(H)-binding [36] is shown beneath the module 12 sequence. Module 9 KR (in blue) is possibly inactive. The two residues likely to be responsible for the inactivity of KR9 are outlined (see text for details).

module 7 (GtGtlAphlA). The exception is the KR domain of module 9 which shows non-conservative replacements in the two highly conserved N-terminal glycines, displaying the sequence pasgtGghcA instead (Figure 3).

The fingerprint region GxGxxG constitutes a tight turn at the end of the first strand of a β-sheet and marks the beginning of the succeeding α-helix in the βαβ-fold (Rossmmann fold) common to the vast majority of NADP(H)-binding domains [36]. In most NADP-binding domains, the same fingerprint region can be identified, although in some cases the first or the third Gly of the conserved trio are replaced by Ala [35,36]; however, an alteration of the kind showed by module 9 KR should preclude the formation of the turn, and hence make such a KR inactive. Such inactive KR should leave C9 as a carbonyl group, which would be required for the formation of the pyrane ring of the final pimaricin molecule (Figure 2).

Modules 2, 3, 4, 5, 11 and 12 each contain a DH domain with the conserved active-site motif LxxHxxxGxxxxP [26,27]. DHs from modules 2, 3, 4 and 5 presumably contribute to the formation of the chromophore (Figure 2), while the presence of an active DH in module 11 would imply that the epoxy function between C4 and C5 in the final molecule of pimaricin derives from an original double bond. This arrangement in the nascent polyketide chain is different from the situation of other characterised polyketides with an epoxy ring in their molecule, the microtubule-stabilising agents epothilones A and B, where the epoxy group derives from a β-keto group that is reduced to a hydroxyl [20,21]. The formation of the epoxy ring would, in both cases, require post-PKS oxidations. Either PimD or PimG (see below) could carry out such oxidation of the pimaricin polyketide skeleton.

Chain termination on the pimaricin PKS may require a TE, which releases the finished polyketide chain. A TE domain has been identified adjacent to module 12, at the carboxy end of PIMS4, displaying the sequences GHSSA and ADH, good matches to the consensus motifs GxSxG and GdH [27].

Pimaricin-modifying genes

The biosynthesis of the final molecule of pimaricin requires a series of post-PKS modifications. The cluster contains a number of genes that appear to be involved in this further elaboration of the polyketide product.

By analogy with similar oxidations in the biosynthesis of erythromycin [37–39], oleandomycin [40] and rapamycin [41], the production of the hydroxyl group at C4 (that is converted later into the epoxide), and some of the steps in the formation of the exocyclic carboxyl function in pimaricin (Figure 2) could be catalysed by cytochrome P-450

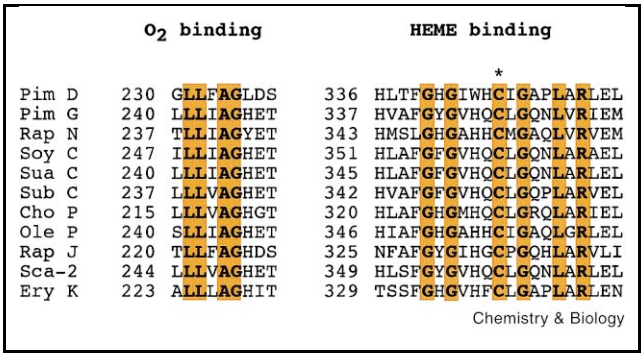

Figure 4. Comparison of oxygen-binding pockets and haem-binding sequences from *Streptomyces* P-450 monooxygenases. The asterisk shows the haem-binding Cys. Numbers indicate aa residues from the N-terminus of the protein. Conserved aa are shown in bold and highlighted. Rap N and Rap J, *Streptomyces hygroscopicus* P-450s involved in rapamycin biosynthesis (EMBL X86780); Soy C, soy flour-induced P-450 from *S. griseus* (EMBL X63601); Sua C and Sub C, sulfonylurea-induced P-450 SU1 and SU2, respectively, from *Streptomyces griseolus* (EMBL M32238, M32239); Cho P, P-450 from *Streptomyces* sp. cholesterol oxidase operon (EMBL M31939); Ole P, *S. antibioticus* P-450 involved in oleandomycin biosynthesis (EMBL L37200); Sca-2, *S. carbophilus* P-450 involved in provastatin biosynthesis (EMBL D30815).

monooxygenases. Two genes in the cluster, *pimD* and *pimG*, appear to encode P-450-like monooxygenases that might fulfil such roles. Thus, the sequences of PimD (397 aa) and of PimG (398 aa) show over 30% sequence identity along its full length with the *Streptomyces antibioticus* P-450 involved in oleandomycin biosynthesis (L37200), or the P-450 (*rapJ*) from the rapamycin biosynthetic gene cluster (X86780), among others. An alignment of the regions of bacterial P-450 proteins that show the greatest degree of sequence conservation is shown in Figure 4. These include part of the O_2-binding site and part of the haem-binding pocket, including the invariant Cys residue that co-ordinates the haem [42,43].

The action of P-450 monooxygenases requires electron transport from NADH, mediated by NADH:ferredoxin oxidoreductase and/or a ferredoxin [44]. Several prokaryotic P-450 monooxygenase-encoding genes have been found in an operon with such electron transport components [41,44] and that is the case of *pimG*. The *pimF* gene immediately downstream from *pimG* encodes a small acidic protein (63 aa) with convincing similarity to ferredoxins containing [3Fe–4S] clusters (Figure 4). Either this ferredoxin also partners PimD or some general-purpose ferredoxin may be recruited from elsewhere in cellular metabolism and serve as the electron carrier for this hydroxylase [45].

The pimaricin cluster also contains a gene, *pimE*, whose product showed a very high end-to-end sequence identity

with cholesterol oxidases. This value ranged from 58.9% for ChoB, the cholesterol oxidase of *Brevibacterium sterolicum* [46], to 79.9% or 82.6% for ChoM or ChoA, the enzymes from *Streptomyces* sp. strains A19249 and SA-COO, respectively [47,48]. Conceivably, PimE could participate in some of the oxidative modification steps of the polyketide product, although its precise function remains to be determined.

Curiously, located at the start of the transcriptional unit that controls the second set of PKS genes (PimS2–PimS4) is one gene, *pimJ*, that seems to be involved in the formation of the sugar moiety of pimaricin mycosamine (3-amino-3,6-dideoxy-D-mannose) which is attached to C15. PimJ (343 aa) showed significant sequence similarity over its entire length (up to 54% identity) to enzymes involved in sugar biosynthesis, including the GDP-mannose 4,6-DH from *Pseudomonas aeruginosa* (Q51366), or the dTDP-glucose 4,6-DH from *E. coli* (P37759). It is, therefore, predicted that PimJ catalyses the reductive deoxygenation of a nucleotide-mannose-like sugar.

The deduced gene product of *pimK* (458 aa) showed weak similarity with various galactosyl and glucuronosyl transferases from mammals (up to 26% identity), suggesting that it might be involved in the transfer of the sugar moiety to the aglycone of pimaricin. However, this will require experimental confirmation.

Downstream from *pimJ* lies *pimI*, whose deduced product shows strong homology to various TE gene products, such as the TE from the candicidin gene cluster of *Streptomyces griseus* [49], and Orf5 (TylO) from the tylosin gene cluster of *Streptomyces fradiae* [50]. Remarkably, PimI showed very weak homology to the TE domain that forms part of module 12, suggesting that both activities have different origins. The function of a second TE activity as a single separate protein is not obvious, but TE genes have also been found separate and distal from PKS genes in other antibiotic gene clusters, such as the rifamycin [23], the nystatin [16], or the tylosin [50] biosynthetic gene clusters. A putative editing activity has been suggested to be associated with TylO to promote correct tylosin accumulation [50], and a similar role could be suggested for PimI.

Export and regulatory genes

One of the mechanisms of self-resistance in macrolide-producing bacteria is the removal of the secondary metabolite from the cytoplasm by transporting it out of the cell. The products of three genes in the cluster might be involved in pimaricin export in *S. natalensis*. PimA (602 aa) and PimB (626 aa) both have several possible transmembrane segments and show convincing matches to proteins belonging to the ATP-dependent ABC transporter superfamily. PimA has 36–43% identity to the multidrug resistance pump from the fruit fly (Q00449), or to the carboxy-terminal end of permeases to haemolysin from *E. coli* (P08717), or cyclolysin from *Bordetella pertussis* (P19770). PimB seems to be even more closely related to the same kind of permeases (identities rise up to 55%). Both gene products are mutually homologous over their full length (28% identity).

PimH (432 aa), the third gene product putatively involved in pimaricin export, is 36% identical to an integral membrane protein, Orf4, from the *Mycobacterium smegmatis* ethambutol resistance region (U46844). It is also more distantly related to antibiotic efflux systems (identities range between 24 and 27%), like the tetracenomycin C resistance and export protein from *Streptomyces glaucescens*, the methylenomycin resistance protein (Q00538) from *Bacillus subtilis*, or the actinorhodin efflux pump (P46105) from *Streptomyces coelicolor*.

The remaining gene of the cluster has a potential regulatory function, but there is as yet no evidence for a link to pimaricin biosynthesis. PimC (352 aa) shows 30–38% sequence identity over its first 250 aa residues with some formerly considered sensor proteins, including EryC1 from the erythromycin-producing strain *Saccharopolyspora erythraea* (P14290), DnrJ from the daunorubicin-producing *Streptomyces peucetius* (P25048), or the pleiotropic regulatory protein DegT from *Bacillus stearothermophilus* (P15263). The members of this protein family seemed to be membrane proteins which functioned as sensors, transferring the signal of the environmental stimuli to the regulatory region of target genes, activating or repressing the transcription of certain genes [51,52], however, recent reports indicate that they are aminotransferase enzymes involved in the biosynthesis of aminosugars [53]. Therefore, PimC could conceivably operate similarly.

Divergent codon usage of the PKS genes suggests different origins for *pimS2* and the rest of PKS genes

An unusual feature of the pimaricin gene cluster is that there is a consistent divergence in codon usage between *pimS2* and the rest of PKS genes. While the overall G+C content remains unchanged, at about 71–75%, the percentage of G+C in the third position of the codon is slightly increased in *pimS2*, from 88.6% to 93.5%, almost wholly due to greater use of codons with C rather than T in those positions. This is true for all 15 aa specified by codons with C in the wobble position. Besides, there is a substantial increase on the use of codons ending with C rather than G, from 45–48% to 60.5%. This observation might just reflect the different origins of *pimS2* and the rest of the PKS genes, indicating that while *pimS0*, *pimS1*, *pimS3* or *pimS4* might have originated by duplication events from an ancestral PKS gene, *pimS2* could have been acquired from a different source. This, together with the arrangement of the PKS genes in separate subclusters, strengthens our previous hypothesis of pimaricin being assembled by

putting together genes encoding PKSs from different origins [7].

The presence of TTA codons in the cluster is of particular interest since their involvement in the regulation of secondary metabolism has been suggested [54]. There is one such rare Leu codon in the *pimS1* gene [7] which could allow such regulatory mechanism to control expression of this gene.

Discussion

Actinomycetes are well-known producers of antifungal agents, especially of polyketide-derived polyenes [1,14]. *S. natalensis* produces an archetype of glycosylated polyenes, the 26-membered macrolide pimaricin (polyene macrolides normally contain very large macrocyclic rings, up to 38-membered rings). In a previous work, we described the identification of the first two PKS genes of the cluster [7], here we complete the sequencing of the pimaricin PKS genes responsible for the synthesis of this relatively small polyene. The sequence includes the entire sequence of a type I PKS with the correct number of modules required for the assembly of the pimaricin polyketide backbone from one propionate and 12 acetate units. PKS genes are arranged in separate subclusters (Figure 1), however, the deduced activities of the PKS enzymes are in a remarkably good agreement with the final structure of pimaricin. Inactivation of PKS genes on either subcluster abolished pimaricin production, providing the final proof for the identity of the overall cluster.

The typical colinearity of the order of modules within each polypeptide with the order of their utilisation during biosynthesis is a general feature of type I PKSs that reflects the processivity of these enzymes [8,9,55], and is maintained here. This colinearity has been extended further to the order of the genes encoding the PKS in some cases, such as the erythromycin [9,27], pikromycin [32], rifamycin [23], or epothilone [20,21] clusters. However, in the case of the pimaricin synthase, there is no such gene colinearity, as also happens in other examples, such as the nystatin [16], the avermectin [19], the rapamycin [5] or the FK506 [56] clusters.

Five multifunctional enzymes compose the PKS, namely PIMS0–PIMS4, which collaborate in the formation of the polyketide backbone of pimaricin, in a fashion similar to the 6-deoxyerythronolide B synthase subunits [25]. PIMS1 contains the first four modules for polyketide chain extension, PIMS2 contains the next six modules containing the activities required to continue chain elongation up to C5 (Figure 2), then the chain is transferred to PIMS3 which contains a single module for chain extension, and finally PIMS4 contains the last module required to complete the polyketide structure of pimaricin. In total, the five proteins contribute 60 catalytic functions, making this multienzyme one of the most complex systems identified so far. Furthermore, 25 of these domains are housed in a single polypeptide, the 6-module PIMS2 protein, an extraordinarily large molecule only paralleled by the giant RAPS2 protein of the rapamycin synthase [5,28], and the recently described NysC and NysI proteins of the nystatin synthase [16]. Curiously, analysis of the codon usage of *pimS2* shows a divergence from the rest of PKS genes by a greater use of codons in which T or G in the 'wobble' position is substituted by C, suggesting a different origin for this gene as compared with the remaining PKS genes. This observation thus reinforces our previous assumption based on the distribution of the PKS genes among two different subclusters [7].

Based on the structure of pimaricin, it would appear that the pimaricin PKS initiates polyketide chain assembly from the starter unit acetate (Figure 2). The problem of chain initiation, in modular PKSs, is normally solved by having a loading domain included at the N-terminus of the first multienzyme [5,10,19,23,25]. However, in the pimaricin PKS there is not such an arrangement, and the loading of the starter unit acetate is carried out by a separate protein, PIMS0, harbouring an N-terminal CoL domain and a minimal PKS module tethered by an ACP domain (Figure 2) [7]. Pimaricin assembly thus begins with the loading of an acetate unit onto the C-terminal ACP domain of PIMS0. The first condensation step is then carried out by the module 1 KS domain of PIMS1, and chain extension subsequently resumes.

A separate polypeptide functioning as a loading module for polyketide biosynthesis has also been observed in NysA, the presumed loading module involved in the initiation of the nystatin aglycone biosynthesis[16]. Both the PIMS0 and NysA proteins are unusual in the sense that they contain KS domains in which the conserved active-site cysteine residue is replaced by a serine. The significance of the cysteine–serine replacement in the active sites of these separate loading polypeptides is unclear, and deserves further investigation.

Chain termination in macrolide formation is normally catalysed by a TE off-loading domain at the end of the last module of the PKS, and there is one such domain in module 12 of PIMS4. However, there is another, discrete, TE (PimI) embedded between the PKS genes, whose function could be similar to the one proposed for the separate additional TE gene (TylO) located within the tylosin gene cluster [50], which promotes tylosin accumulation by unblocking a PKS complex prone to blockage with aberrant polyketides.

The cluster includes at least four genes possibly involved in the oxidation of the polyketide skeleton derived from the PKS at C4 and C27. Their gene products include two

cytochrome P-450 monooxygenases (PimD and PimG) and a ferredoxin (PimF), activities which are commonly encountered in macrolide gene clusters, but also a cholesterol oxidase (PimE). The chromosomal location of *pimE* just in the middle of the pimaricin gene cluster could indicate a direct involvement of the *pimE* gene product in the formation of the final pimaricin molecule. It is tempting to speculate that the free carboxyl function at C27 of the pimaricin molecule (Figure 2) could be formed by the action of PimE, and maybe other oxidases on the exocyclic methyl residue of the six-membered hemiketal ring introduced by PIMS2. Bacterial cholesterol oxidases, in fact, have been proven to be multifunctional enzymes carrying 3β-hydroxysteroid oxidase and Δ^5-3-ketosteroid isomerase activities, but also a hydroxylase (monooxygenase) activity [57]. The presence of a cholesterol oxidase gene within a polyketide gene cluster is unprecedented. This, together with the presence of exocyclic carboxyl functions in many polyene macrolides [14], a characteristic feature absent in other type I PKS products, could indicate a specific involvement of such genes in polyene macrolide formation. However, future functional analysis of PimE will be required to elucidate its role in pimaricin biosynthesis. Similarly, future analysis of other biosynthetic gene clusters for polyene bearing carboxyl moieties will contribute to the validation of this hypothesis.

One of the mechanisms of self-resistance in macrolide-producing bacteria is the removal of the secondary metabolite from the cytoplasm by transporting it out of the cell. Polyene-producing bacteria, however, should not require such removal for self-resistance since polyenes are only toxic to cells containing sterols in their membranes. However, an efficient transport of pimaricin must conceivably play an important role in *S. natalensis* success as a soil dweller. The products of three genes of the cluster (PimA, PimB and PimH) seem to be involved in pimaricin export thus contributing to *S. natalensis* survival in its natural habitat.

The remaining genes of the cluster are plausibly involved in pimaricin biosynthesis through modification of the aglycone and regulatory functions. Among those, two genes possibly related with mycosamine biosynthesis and attachment have been located. Several genes presumably required for pimaricin biosynthesis remain now to be uncovered to complete the pimaricin pathway, at least a minimum of 2–3 to produce the mycosamine moiety, and maybe a transcriptional activator or other genes with regulatory functions.

Significance

The naturally occurring polyenes represent a major class of antifungal agents characterised by the presence of a series of conjugated double bonds in their polyketide macrolide ring. This work describes the gene cluster responsible for the biosynthesis of the archetype polyene pimaricin in *S. natalensis* from 12 acetate units and one propionate unit. The cluster is divided into two subclusters which encode a highly complex PKS distributed in five multifunctional polypeptides harbouring 12 homologous sets of enzyme activities (modules), each catalysing a specific round of chain elongation. The total of 60 enzymatic domains makes this one of the most complex multifunctional enzyme systems identified so far. The DNA region sequenced contains 11 additional ORFs which may govern other key steps in pimaricin biosynthesis. The modular architecture of the synthase lends itself to modifications through combinatorial biosynthesis and targeted mutagenesis to create improved pimaricins. However, a larger diversity of novel analogs can be obtained through directed alteration of the functionalisation genes.

Materials and methods

Bacterial strains, cloning vectors and cultivation

S. natalensis ATCC 27448 was routinely grown in YEME medium [58]. Sporulation was achieved in TBO medium (2% (w/v) tomato paste, 2% (w/v) oats flakes, 2.5% (w/v) agar) at 30°C. For pimaricin production, the strain was grown in phosphate-limited SPG medium [7] or YEME without sucrose. *E. coli* strain XL1-Blue MR was used for obtaining SuperCos 1 cosmid (Stratagene) recombinant derivatives, and also served as a host for plasmid subcloning in plasmids pBluescript (Stratagene), pUC18 and pUC19. *Candida utilis* (syn. *Pichia jadinii*) CECT 1061 was used for bioassay experiments. Phage KC515 (*c*$^+$ *attP::tsr::vph*), a ØC31-derived phage [59], was used for gene disruption experiments. *Streptomyces lividans* JII 1326 [60] served as a host for phage propagation and transfection. Standard conditions for culture of *Streptomyces* species and isolation of phages were as described by Hopwood et al. [58].

Genetic procedures

Standard genetic techniques with *E. coli* and in vitro DNA manipulations were as described by Sambrook et al. [61]. Recombinant DNA techniques in *Streptomyces* species and isolation of *Streptomyces* total and phage DNA were performed as previously described [58].

For the construction of the genomic library, *S. natalensis* genomic DNA was partially digested with *Sau*3AI and fragments in the 35–40 kb size range were cloned into SuperCos 1 digested with *Bam*HI and *Xba*I. The ligation mixture was packaged with Gigapack III XL (Stratagene) and used to transfect *E. coli* XL1. A 1.9 kb *Bam*HI–*Pst*I fragment internal to the *pimS0* gene [7] was used to probe the library of size-fractionated DNA as the starting point for chromosome walking. Sequencing templates were obtained by random subcloning of fragments generated by controlled partial *Hae*III digestions.

DNA sequencing and analysis

DNA sequencing was accomplished by the dideoxynucleotide chain termination method [62] using the Perkin Elmer Amplitaq Dye terminator sequencing system on double-stranded DNA templates with an Applied Biosystems model 310 sequencer (Foster City, CA, USA). Subclone junctions were verified by direct sequencing of cosmid clones using as primer sets of synthesised internal oligonucleotides. Each nucleotide was sequenced a minimum of three times on both strands. Alignment of sequence contigs was performed using the DNA Star program Seqman (Madison, WI, USA). DNA and protein sequences were analysed with the University of Wisconsin Genetics Computer Group software programs [63], and the NCBI World Wide Web blast server.

Construction of pim mutants by gene disruption

This was accomplished by inserting an internal fragment of a gene into

KC515 (all cloning steps were carried out with the *E. coli* plasmid vectors described above), and recombinant phage DNA was transfected into *S. lividans* protoplasts as described elsewhere [58]. Further infection of *S. natalensis* with the recombinant phage in R5 medium [58] allowed the selection for lysogen formation. Because the KC515 phage lacks *attP*, it can form lysogens only by homologous recombination into the chromosome. Typically, less than 10 lysogens were obtained by selection for thiostrepton resistance, and their pimaricin production was tested as described below. An example of the use of this method to produce the *pimS1* mutant is given in [7].

Assay of pimaricin production

To assay pimaricin in culture broths, 0.5 ml of culture was extracted with 2 ml of butanol, and the organic phase was diluted in water-saturated butanol to bring the optical density at 319 nm in the range of 0.1–0.4 absorbance units. Control solutions of pure pimaricin (Gist-Brocades, Delft, The Netherlands) were used as control. To confirm the identity of pimaricin, an UV-visible absorption spectrum (absorption peaks at 319, 304, 291 and 281 nm) was routinely determined in a Hitachi U-2001 spectrophotometer. For routine tests, the fungicidal activity of pimaricin was tested by bioassay using *C. utilis* CECT 1061 as test organism. Quantitative determination of pimaricin was performed with a Shimadzu VP HPLC with a diode array UV detector set at 303 nm, fitted with a μ-Bondapack RP-C18 column (10 μm; 3.9×300 mm). Elution was with a gradient (1.5 ml/min) of 100% methanol (methanol concentration: 50% 0–3 min, up to 90% 3–12 min, 90% 12–20 min, down to 50% 20–25 min, 50% 25–30 min). Retention time for pimaricin was 8.5 min.

Accession number

The sequence reported here has been deposited in the GenBank database under the accession number AJ278573.

Acknowledgements

This work was supported by a grant from the Comisión Interministerial de Ciencia y Tecnología (CICYT) and the European Commission to J.F.A. (1FD97-1419-C02-01). M.V.M. received a fellowship of the Fundação para a Ciência e a Tecnologia (PRAXIS XXI/BD/15850/98). N.O. was the recipient of a fellowship from Agencia de Desarrollo Económico de Castilla y León (12-2/98/LE/0355). We thank M. Driessen, R. Luiten, A.J. Colina and E. Recio for helpful discussions and the excellent technical assistance of M.E. García, J. Merino and B. Martín.

References

1. Gil, J.A. & Martín, J.F. (1997). Polyene antibiotics. In *Biotechnology of Antibiotics*. (Strohl, W.R., ed.), 2nd edn., pp. 551–576, Marcel Dekker, New York.
2. Hopwood, D.A. & Sherman, D.H. (1990). Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu. Rev. Genet.* **24**, 37–66.
3. Staunton, J. (1991). The extraordinary enzymes involved in erythromycin biosynthesis. *Angew. Chem. Int. Ed. Engl.* **30**, 1302–1306.
4. Hopwood, D.A. (1997). Genetic contributions to understanding polyketide synthases. *Chem. Rev.* **97**, 2465–2497.
5. Schwecke, T., Aparicio, J.F., Molnár, I., König, A., Khaw, L.E., Oliynyk, M., Haydock, S.F., Caffrey, P., Cortes, J., Böhm, G., Lester, J., Staunton, J. & Leadlay, P.F. (1995). The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. USA* **92**, 7839–7843.
6. Hu, Z., Bao, K., Zhou, X., Zhou, Q., Hopwood, D.A., Kieser, T. & Deng, Z. (1994). Repeated polyketide synthase modules involved in the biosynthesis of a heptaene macrolide by *Streptomyces* sp. FR-008. *Mol. Microbiol.* **14**, 163–172.
7. Aparicio, J.F., Colina, A.J., Ceballos, E. & Martín, J.F. (1999). The biosynthetic gene cluster for the 26-membered ring polyene macrolide pimaricin. *J. Biol. Chem.* **274**, 10133–10139.
8. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. & Katz, L. (1991). Modular organisation of genes required for complex polyketide biosynthesis. *Science* **252**, 675–679.
9. Cortés, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J. & Leadlay, P.F. (1990). An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348**, 176–178.
10. Aparicio, J.F., Caffrey, P.C., Marsden, A.F.A., Staunton, J. & Leadlay, P.F. (1994). Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase. *J. Biol. Chem.* **269**, 8524–8528.
11. Marsden, A.F.A., Caffrey, P., Aparicio, J.F., Loughran, M.S., Staunton, J. & Leadlay, P.F. (1994). Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. *Science* **263**, 378–380.
12. Haydock, S.F., Aparicio, J.F., Molnár, I., Schwecke, T., König, A., Marsden, A.F.A., Galloway, I.S., Staunton, J. & Leadlay, P.F. (1995). Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA: acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* **374**, 246–248.
13. O'Hagan, D. (1995). Biosynthesis of fatty acid and polyketide metabolites. *Nat. Prod. Rep.* **12**, 1–32.
14. Martín, J.F. (1977). Biosynthesis of polyene macrolide antibiotics. *Annu. Rev. Microbiol.* **31**, 12–38.
15. Bolard, J. (1986). How do the polyene macrolide antibiotics affect the cellular membrane properties? *Biochim. Biophys. Acta* **864**, 257–304.
16. Brautaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strom, A.R., Valla, S. & Zotchev, S.B. (2000). Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway. *Chem. Biol.* **7**, 395–403.
17. Wright, F. & Bibb, M.J. (1992). Codon usage in the G+C-rich *Streptomyces* genome. *Gene* **113**, 55–65.
18. Caffrey, P., Green, B., Packman, L.C., Rawlings, B.J., Staunton, J. & Leadlay, P.F. (1991). An acyl-carrier-protein-thioesterase domain from the 6-deoxyerythronolide B synthase of *Saccharopolyspora erythraea*. *Eur. J. Biochem.* **195**, 823–830.
19. Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. & Omura, S. (1999). Organisation of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proc. Natl. Acad. Sci. USA* **96**, 9509–9514.
20. Molnár, I., Schupp, T., Ono, M., Zinkle, R.E., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D.D., Gorlach, J., Mayo, J.M., Hu, A., Goff, S., Schmid, J. & Ligon, J.M. (2000). The biosynthetic gene cluster for the microtubule-stabilising agents epothilones A and B from *Sorangium cellulosum* So ce90. *Chem. Biol.* **7**, 97–109.
21. Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. & Julien, B. (2000). Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640–642.
22. Cortés, J., Wiesmann, K.E.H., Roberts, G.A., Brown, M.J., Staunton, J. & Leadlay, P.F. (1995). Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science* **268**, 1487–1489.
23. August, P.R., Tang, L., Yoon, Y.J., Ning, S., Müller, R., Yu, T.-W., Taylor, M., Hoffmann, D., Kim, C.-G., Zhang, X., Hutchinson, C.R. & Floss, H.G. (1998). Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. *Chem. Biol.* **5**, 69–79.
24. Böhm, I., Holzbaur, I.E., Hanefeld, U., Cortés, J., Staunton, J. & Leadlay, P.F. (1998). Engineering of a minimal modular polyketide synthase, and targeted alteration of the stereospecificity of polyketide chain extension. *Chem. Biol.* **5**, 407–412.
25. Staunton, J., Caffrey, P., Aparicio, J.F., Robets, G.A., Bethell, S.S. & Leadlay, P.F. (1996). Evidence for a double-helical structure for modular polyketide synthases. *Nat. Struct. Biol.* **3**, 188–192.
26. Bevitt, D.J., Cortés, J., Haydock, S.F. & Leadlay, P.F. (1992). 6-Deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea*. *Eur. J. Biochem.* **204**, 39–49.
27. Donadio, S. & Katz, L. (1992). Organisation of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. *Gene* **111**, 51–60.
28. Aparicio, J.F., Molnár, I., Schwecke, T., König, A., Haydock, S.F., Khaw, L.E., Staunton, J. & Leadlay, P.F. (1996). Organisation of the biosynthetic gene cluster for rapamycin in *Streptomyces hygrosco-*

picus: analysis of the enzymatic domains in the modular polyketide synthase. *Gene* **169**, 9–16.
29. Swan, D.G., Rodríguez, A.M., Vilches, C., Méndez, C. & Salas, J.A. (1994). Characterisation of a *Streptomyces antibioticus* gene encoding a type I polyketide synthase which has an unusual coding sequence. *Mol. Gen. Genet.* **242**, 358–362.
30. Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G. & Lindquist, Y. (1998). Crystal structure of β-ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzyme. *EMBO J.* **17**, 1183–1191.
31. Kakavas, S.J., Katz, L. & Stassi, D. (1997). Identification and characterisation of the niddamycin polyketide synthase genes from *Streptomyces caelestis*. *J. Bacteriol.* **179**, 7515–7522.
32. Xue, Y., Zhao, L., Liu, H.-W. & Sherman, D.H. (1998). A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: Architecture of metabolic diversity. *Proc. Natl. Acad. Sci. USA* **95**, 12111–12116.
33. Bisang, C., Long, P.F., Cortés, J., Westcott, J., Crosby, J., Matharu, A.L., Cox, R.J., Simpson, T.J., Staunton, J. & Leadlay, P.F. (1999). A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **401**, 502–505.
34. Witkowski, A., Joshi, A.K., Lindqvist, Y. & Smith, S. (1999). Conversion of a β-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with a glutamine. *Biochemistry* **38**, 11643–11650.
35. Scrutton, N.S., Berry, A. & Perham, R.N. (1990). Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* **343**, 38–43.
36. Wierenga, R.K., Terpstra, P. & Hol, W.G.J. (1986). Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* **187**, 101–107.
37. Haydock, S.F., Dowson, J.A., Dhillon, N., Roberts, G.A., Cortés, J. & Leadlay, P.F. (1991). Cloning and sequencing analysis of genes involved in erythromycin biosynthesis in *Saccharopolyspora erythraea*: sequence similarities between EryG and a family of S-adenosylmethionine-dependent methyltransferases. *Mol. Gen. Genet.* **230**, 120–128.
38. Weber, J.M., Leung, J.O., Swanson, S.J., Idler, K.B. & McAlpine, J.B. (1991). An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. *Science* **252**, 114–117.
39. Stassi, D., Donadio, S., Staver, M.J. & Katz, L. (1993). Identification of a *Saccharopolyspora erythraea* gene required for the final hydroxylation step in erythromycin biosynthesis. *J. Bacteriol.* **175**, 182–189.
40. Rodríguez, A.M., Olano, C., Méndez, C., Hutchinson, C.R. & Salas, J.A. (1995). A cytochrome P450-like gene possibly involved in oleandomycin biosynthesis by *Streptomyces antibioticus*. *FEMS Microbiol. Lett.* **127**, 117–120.
41. Molnár, I., Aparicio, J.F., Haydock, S.F., Khaw, L.E., Schwecke, T., König, A., Staunton, J. & Leadlay, P.F. (1996). Organisation of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of genes flanking the polyketide synthase. *Gene* **169**, 1–7.
42. Poulos, T.L., Finzel, B.C. & Howard, A. (1987). High-resolution crystal-structure of cytochrome $P450_{CAM}$. *J. Mol. Biol.* **195**, 687–700.
43. Cupp-Vickery, J.R. & Poulos, T.L. (1995). Structure of cytochrome $P450_{ERYF}$ involved in erythromycin biosynthesis. *Nat. Struct. Biol.* **2**, 144–153.
44. O'Keefe, D.P. & Harder, P.A. (1991). Occurrence and biological function of cytochrome P450 monooxygenases in actinomycetes. *Mol. Microbiol.* **5**, 2099–2105.
45. Zotchev, S.B. & Hutchinson, C.R. (1995). Cloning and heterologous expression of the genes encoding nonspecific electron transport components for a cytochrome P450 system of *Saccharopolyspora erythraea* involved in erythromycin production. *Gene* **156**, 101–106.
46. Ohta, T., Fujishiro, K., Yamaguchi, K., Tamura, Y., Aisaka, K., Uwajima, T. & Hasegawa, M. (1991). Sequence of gene *choB* encoding cholesterol oxidase of *Brevibacterium sterolicum*: comparison with *choA* of *Streptomyces* sp. SA-COO. *Gene* **103**, 93–96.
47. Corbin, D.R., Greenplate, J.T., Wong, E.Y. & Purcell, J.P. (1994). Cloning of an insecticidal cholesterol oxidase gene and its expression in bacteria and in plant protoplasts. *Appl. Environ. Microbiol.* **60**, 4239–4244.
48. Ishizaki, T., Hirayama, N., Shinkawa, H., Nimi, O. & Murooka, Y. (1989). Nucleotide sequence of the gene for cholesterol oxidase from a *Streptomyces* sp.. *J. Bacteriol.* **171**, 596–601.
49. Criado, L.M., Martín, J.F. & Gil, J.A. (1993). The *pab* gene from *Streptomyces griseus*, encoding p-aminobenzoic acid synthase, is located between genes possibly involved in candicidin biosynthesis. *Gene* **126**, 135–139.
50. Butler, A.R., Bate, N. & Cundliffe, E. (1999). Impact of thioesterase activity on tylosin biosynthesis in *Streptomyces fradiae*. *Chem. Biol.* **6**, 287–292.
51. Tagaki, M., Takada, H. & Imanaka, T. (1990). Nucleotide sequence and cloning in *Bacillus subtilis* of the *Bacillus stearothermophilus* pleiotropic regulatory gene *degT*. *J. Bacteriol.* **172**, 411–418.
52. Dhillon, N., Hale, R.S., Cortés, J. & Leadlay, P.F. (1989). Molecular characterisation of a gene from *Saccharopolyspora erythraea* which is involved in erythromycin biosynthesis. *Mol. Microbiol.* **3**, 1405–1414.
53. Madduri, K. & Hutchinson, C.R. (1995). Functional characterisation and transcriptional analysis of the *dnrR*$_1$ locus, which controls daunorubicin biosynthesis in *Streptomyces peucetius*. *J. Bacteriol.* **177**, 1208–1215.
54. Leskiw, B.K., Lawlor, E.J., Fernández-Abalos, J.M. & Chater, K.F. (1991). TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, *Streptomyces* mutants. *Proc. Natl. Acad. Sci. USA* **88**, 2461–2465.
55. Yu, T.-W., Shen, Y., Doi-Katayama, Y., Tang, L., Park, C., Moore, B.S., Hutchinson, C.R. & Floss, H.G. (1999). Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. *Proc. Natl. Acad. Sci. USA* **96**, 9051–9056.
56. Motamedi, H. & Shafiee, A. (1998). The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. *Eur. J. Biochem.* **256**, 528–534.
57. Molnár, I., Hayashi, N., Choi, K.P., Yamamoto, H., Yamashita, M. & Murooka, Y. (1993). Bacterial cholesterol oxidases are able to act as flavoprotein-linked ketosteroid monooxygenases that catalyse the hydroxylation of cholesterol to 4-cholesten-6-ol-3-one. *Mol. Microbiol.* **7**, 419–428.
58. Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. & Schrempf, H. (1985). Genetic Manipulation of *Streptomyces*. A Laboratory Manual. John Innes Foundation, Norwich.
59. Rodicio, M.R., Bruton, C.J. & Chater, K.F. (1985). New derivatives of the *Streptomyces* temperate phage ØC31 useful for the cloning and functional analysis of *Streptomyces* DNA. *Gene* **34**, 283–292.
60. Chater, K.F., Bruton, C.J. & Suárez, J.E. (1981). Restriction mapping of the DNA of the *Streptomyces* temperate phage ØC31 and its derivatives. *Gene* **14**, 183–194.
61. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
62. Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
63. Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.