The biosynthetic gene cluster for the 26-membered ring polyene macrolide pimaricin

A NEW POLYKETIDE SYNTHASE ORGANIZATION ENCODED BY TWO SUBCLUSTERS SEPARATED BY FUNCTIONALIZATION GENES

(Received for publication, December 22, 1998)

Jesús F. Aparicio‡, Angel J. Colina, Elvira Ceballos§, and Juan F. Martín¶

From the Institute of Biotechnology INBIOTEC, Science Park of León, Avda. del Real, n° 1, 24006 León, Spain

The biosynthetic gene cluster for the 26-membered ring of the polyene macrolide pimaricin extends for about 110 kilobase pairs of contiguous DNA in the genome of Streptomyces natalensis. Two sets of polyketide synthase (PKS) genes are separated by a group of small polyketide-functionaizing genes. Two of the polyketide synthase genes, pimS0 and pimS1, have been fully sequenced and disrupted proving the involvement of each of these genes in pimaricin biosynthesis. The pimS0 gene encodes a relatively small acetate-activating PKS (−193 kDa) that appears to work as a loading protein which "presents" the starter unit to the second PKS subunit. The pimS1 gene encodes a giant multienzyme (−710 kDa) harboring 15 activities responsible for the first four cycles of chain elongation in pimaricin biosynthesis, resulting in formation of the polyene chromophore.

Polyketides are a large and highly diverse group of natural products (1) that includes antibacterial, antifungal, anticancer, antiparasitic, and immunosuppressant activities, among others. Despite their structural diversity, these metabolites are synthesized by a common pathway in which units derived from acetate, propionate, or butyrate are condensed onto the growing chain by a polyketide synthase (PKS) in a process resembling fatty acid biosynthesis (2–4). In this pathway, the β-keto function introduced at each elongation step may undergo all, part, or none of a reductive cycle comprising β-ketoreduction, dehydration, and enoyl reduction (5). The structural variety in this class of products arises from choice of monomers, the extent of β-ketoreduction, and dehydration and the stereochemistry of each chiral center. Yet further diversity is produced by functionalization of the polyketide chain by the action of glycosylases, methyltransferases, and oxidative enzymes.

Macrocyclic polyketides are produced mainly by Streptomyces and related filamentous bacteria through the action of so-called type I modular PKSs (5–8). These enzymes usually consist of several extremely large polypeptides in which different modules (sets) of enzymatic activities catalyze each successive round of elongation (6, 7). Characterization of such PKS systems is not only an important scientific challenge but a necessary start to the study of factors that control the specificity and extent of chain extension, which are largely unknown (4, 9, 10).

Polyene macrolides are a group of macrocyclic polyketides that interact with membrane sterols and are, therefore, active against fungi but not bacteria (11, 12). Pimaricin (Fig. 1) represents a prototype molecule of glycosylated polyenes (13) important for antifungal therapy and promising for antiviral activity, stimulation of the immune response, and action in synergy with other antifungal drugs or antitumor compounds (14). It is produced by Streptomyces natalensis and widely utilized in the food industry to prevent mold contamination of cheese and other nonsterile foods (i.e. cured meats). The macrolide rings of polyene antibiotics are larger (up to twice in size) than those of standard 14- or 16-membered nonpolyene macrolides (11). Their rings include a chromophore of conjugated double bonds (the characteristic polyene structure). The genetic determination of the length of the polyene and the chromophore and the ring size of these giant macrolide rings is intriguing.

Despite the general interest of polyene macrolides, very little is known about their biosynthetic routes and the gene clusters encoding them. We have, therefore, undertaken the cloning of a large Streptomyces gene cluster involved in the biosynthesis of the polyketide backbone that forms the 26-membered tetraene macrolide ring of pimaricin. pimS1, the gene responsible for the first four rounds of chain elongation has been sequenced, and its modular nature has been established. pimS0, a gene required for chain initiation has been also studied. Gene disruption studies provide evidence for the involvement of pimS0 and pimS1 in pimaricin biosynthesis.

EXPERIMENTAL PROCEDURES

Bacterial Strains, Cloning Vectors, and Phages—S. natalensis ATCC 27448 was used as the source of DNA in the construction of the genomic library. Escherichia coli strain XLI-Blue MR was used for obtaining SuperCos 1 cosmids (Stratagene) recombinant derivatives and also served as a host for plasmid subcloning in plasmids pBluescript (Stratagene) pUC18 and pUC19. Candida utilis (synonym Pichia jadinii) CECT 1061 was used for bioassay experiments. Phage KC515 (c′ attP-tsr-vph), a OC31-derived phage (15), was used for gene disruption experiments (16). Streptomyces lividans JII 1326 (17) 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. (16).

S. natalensis was routinely grown in YEME medium (16). 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 (2.5% (w/v) soy peptone, 0.5 mM ZnSO4.7H2O, 2% (w/v) glucose, pH 7.5) as described by Martin et al. (18).
by sequence comparison alignments with other modular PKSs (see below). Because of the modular nature of PKS genes, the use of a highly conserved probe at moderate stringency hybridization gave us a gross indication of the extent of the PKS-coding DNA. Thus, about 13 kb of the rightmost sequences of Cos 37 (Fig. 2) did not hybridize with the probe, suggesting that the end of the PKS part of the gene cluster is located within cosmID 37. This hybridization approach also indicated that approximately 9 kb of the leftmost sequences of Cos 37 (Fig. 2) were beyond the PKS portion of the cluster. In addition, no hybridization signal was observed in the central area of the cluster (Fig. 2), suggesting that a possible non-PKS region (−10 kb) is located between sets of PKS genes in this central area.

To confirm these results, all the NotI fragments within the 110 kb of cloned chromosomal DNA were subcloned into pBlueScript, and their ends were sequenced. Fig. 2B shows the function to which translated peptides from the sequence showed similarity and the direction of transcription. The first set of PKS genes pimSO and pimS1 (Fig. 2) were fully sequenced. The second set of PKS genes are transcribed in opposite orientation to the first set and appears to encode two additional modular PKSs. Several non-PKS- accompanying proteins were found to be encoded between the two large sets of PKS domains, further corroborating the above results. One of these activities showed significant sequence similarity with peroxisomes within the ATP-dependent ABC transporter superfamily, whereas another one showed high similarity to cholesterol oxidases of other Streptomyces species. Other activities with counterparts in the protein data base included a sensory transduction protein, a ferredoxin, and a cytochrome P-450 monooxygenase. All these genes can be plausibly assigned roles in pimaricin biosynthesis. On the right side of the cluster, a cytochrome P-450 monooxygenase-like gene (pimD) was also identified and fully sequenced. No other open reading frames with known homologies were detected on the left side of the 110-kb region studied.

Functional Inactivation of the PKS Blocks Pimaricin Production—The involvement of the cloned DNA region in pimaricin biosynthesis was tested by gene disruption. Because S. natalensis ATCC 27448 has so far proven to be highly resistant to transformation with commonly used Streptomyces plasmids, a different approach for delivering of DNA into this strain was used. We took advantage of the ability of phage KC515, an attP-defective 0C31 derivative (15), to infect S. natalensis to introduce DNA into this strain. A 3.3-kb NotI subclone in pBlueScript containing an internal coding region to PKS domains on the right side of the cluster (fragment b in Fig. 2A) was used as a source of DNA for subcloning into KC515 (Fig. 3). A 3.25-kb SacI fragment (containing most of the insert and the SacI site of the polylinker of...
pBluescript) was cloned into KC515, and the recombinant phage DNA was transfected into S. lividans protoplasts, as described elsewhere (16). Further infection of S. natalensis with the recombinant phage allowed the selection for lysogen formation. Because the KC515 phage lacks attP, it can form lysogens only by homologous recombination into the chromosome (Fig. 3A).

Several lysogens of S. natalensis were obtained by selection for thiostrepton resistance, and their pimaricin production was tested by bioassay against C. utilis. No antibiotic activity was observed in the tested lysogens (Fig. 3C) as compared with the halos of antifungal activity produced by the wild type strain. Specificities were assigned by comparison with divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases (27). Open circles indicate deduced AT domains whose specificity could not be established. Numbers over the circles indicate the order in which each AT is used during the biosynthesis of the pimaricin polyketide backbone. 0 refers to the AT presumably involved in the loading of the starter unit (see text).

The Pimaricin Polyketide Synthase

FIG. 2. Organization of the pimaricin biosynthetic gene cluster. A, shading indicates NotI, BamHI, EcoRI, or SphI fragments hybridizing with probe a (coding for most of module 1 KS domain). Fragments b and c were used for gene disruption (see Fig. 3). Only NotI (N) restriction sites are indicated on the map. The cosmid clones studied are shown. B, nucleotide sequence information obtained from the ends of the NotI fragments is indicated on the map. The direction of transcription is indicated by the arrowheads. pimS1, pimD, and pimS0 have been completely sequenced and analyzed (see text). KR, β-ketoreductase; DH, dehydratase; Cho Ox, cholesterol oxidase. C, organization of AT domains involved in pimaricin biosynthesis. Ac, AT domain that incorporates an acetate extender unit into the growing polyketide chain; Pr, AT domain specific for propionate. Numbers over the circles indicate the order in which each AT is used during the biosynthesis of the pimaricin polyketide backbone.
The dehydratase domain from 6-deoxyerythronolide B synthase module 4 and the dehydratase domains from the RAPS proteins of the rapamycin synthase (8); the active-site motif HXXXGXXXXP (24, 25) is present. In addition, KS domains are highly conserved (70–80% identity over the whole domain) and contain the two invariant His residues located 135 and 173 amino acids C-terminal of the active-site Cys, as previously stated for other PKS ketosynthase domains (Ref. 8 and references therein). The β-ketoreductase domains of PIMS contain a potential motif for NADP(H) binding (GXGXXGXXA) at the N-terminal end of the domain (26), as occurs with all the previously published β-ketoreductase domains except for those of the rapamycin synthase (8). ACP domains in PIMS are also highly conserved. Remarkably, ACP from PIMS module 1 has the active-site sequence QGFDS, whereas all previously identified prokaryotic ACPs have Leu adjacent to the Gly.

AT domains of type I modular PKSs fall into two distinct groups, depending upon whether the substrate for the AT is 2S-methylmalonyl-CoA (9) or malonyl-CoA, as discussed in detail elsewhere (27). When PIMS1 AT domains were compared with the divergent motifs found in AT domains of type I PKSs, a convincing match was seen with the malonyl-CoA (“acetate”) consensus (Pim 01, 02, 03, and 04 in Fig. 5), thus suggesting that these AT domains incorporate acetate extender units into the growing polyketide chain.

Alignment of other AT domains of the cluster derived from the DNA sequence of subclone ends (when the sequence covered the stretch of 20 amino acids that corresponds to the divergent AT motifs) allowed the identification of another six acetate-incorporating ATs (Pim 00, 05, 06, 08, 09, and 10) (two others remain to be determined to complete the known chain length of pimaricin) and one AT that incorporates propionate extender units (Pim 07, Figs. 2 and 5). Based on the proposed biosynthetic units of pimaricin (11) (Fig. 1), the polyketide synthase should harbor only one such propionate-specific AT, the one comprised in module 7 and 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, as occurs in lucensomycin biosynthesis (28). The rare Leu codon TTA has been proposed to serve as a device for global regulation in Streptomyces coelicolor (29).

There is one such codon in the pimS1 gene, which may allow this regulatory mechanism to control expression of pimS1.

pimS0 Encodes a Separate Module Also Required for Pimaricin Biosynthesis—Lying downstream of pimS1 and convergently transcribed, there is another open reading frame (pimS0) encoding a protein of 1847 amino acids. The molecular mass of the protein encoded by pimS0 (hereafter named
PIMS0 was calculated to be 193,407 Da as a monomer. PIMS0 contains in its C-terminal end a whole PKS module with no domains for reduction or dehydration (i.e. only KS, AT, and ACP; Fig. 4C). Its KS domain lacks the active-site cysteine residue (Fig. 6). Besides, the sequence of its first domain in the N-terminal region, containing about 560 amino acid residues, reveals up to 28% amino acid sequence identity to ATP-dependent carboxylic acid:CoA ligases, including acetyl-CoA synthases from *Bacillus subtilis* (EMBL L17309), coumarate-CoA ligases from plants (EMBL X13325, L43362, D39405, U50845), 2,3-dihydroxybenzoate-AMP ligases from *E. coli* (EMBL X15058), and *B. subtilis* (U26444), firefly luciferases (EMBL A26772, S33788, S29354), and the long chain fatty acid-carboxylic acid:CoA ligase from *Mycobacterium tuberculosis* (EMBL Q10776), among others. An ACP domain following the carboxylic acid:CoA ligase domain completes the number of domains housed within this protein (Fig. 4C). Remarkably, the C-terminal ACP has the active-site sequence MGINS instead of the signature sequence LGXDS.

To rule out the possibility that *pimS0* could be a remnant of a former gene with no real involvement in pimaricin biosynthesis, its disruption was accomplished. A 1.9-kb BamHI-PstI fragment (containing the coding region for most of the carboxylic acid:CoA ligase domain and the N-terminal end of the KS domain of *pimS0*) was subcloned into KC515. The method to achieve disruption was as described above for *pimS1*. Selected disruptants showed no production of pimaricin (not shown), indicating that *pimS0* is strictly required for pimaricin biosynthesis.

**DISCUSSION**

Polyene macrolides contain very large macrolide rings (sometimes up to 38-membered rings). It is unknown if such large macrolides are synthesized by single PKSs as occurs with 14- or 16-membered macrolides. In this paper we describe the first extensive genetic characterization of a polyene macrolide biosynthetic gene cluster. Southern hybridization experiments were used as means to identify and isolate cosmids clones containing the *S. natalensis* gene cluster for the 26-membered ring of the macrocyclic polyketide pimaricin.

The most striking feature of the DNA region cloned (110 kb) is that the polyketide synthase genes turned out to be split into two subclusters separated by genes involved in the functionalization of the polyketide backbone. This is the first example of two separate PKSs required for the biosynthesis of different fragments of a single polyketide. In the rapamycin biosynthetic gene cluster, the PKS genes are separated by a pipecolate-activating gene that is directly involved in the formation of the aglycone core of rapamycin (5). Assembly of a complex polyketide core by non-PKS enzymes occurs also in *Aspergillus nidulans*, where a gene for the biosynthesis of a preformed fatty acid precursor of sterigmatocystin is separated from the type II PKS genes (30). However, this organization of a modular PKS gene cluster separated by genes involved in functionalization of the polyketide backbone is unprecedented.

The arrangement in separate subclusters could suggest that this huge macrolide has been assembled by putting together genes encoding two different polyketide synthases. Future
The Pimaricin Polyketide Synthase

**FIG. 5.** Alignment of the sequences for the divergent motifs and the active sites found in AT domains of type I polyketide synthases. Pim 1 to 10 indicate AT domains present in modules 1 to 10 of the PKS for pimaricin biosynthesis. Ery01, Ole 5, Sor 5 and Rap02, Rap10, and Rap14 indicate AT domains present in the corresponding modules of the modular PKSs for erythromycin (6, 7), oleandomycin (36), soraphen A (37), and rapamycin (5, 8) biosynthesis, respectively. Pim00 indicates the AT domain present in PIMS0 (see text). Amino acids identical to the divergent motifs are shown in bold. 

| AT Motif | Active Site |
|----------|-------------|
| Pim01    | Pim02       | Pim03       | Pim04       | Pim05       |
| (Val)    | Cys         | Cys         | Cys         | Cys         |
| Pim06    | Pim07       | Pim08       | Pim09       | Pim10       |
| (Val)    | Cys         | Cys         | Cys         | Cys         |
| Consensus| Pim00       | Pim01       | Pim02       | Pim03       |
| (Val)    | Cys         | Cys         | Cys         | Cys         |

**FIG. 6.** Alignment of the sequences around the active site Cys (in bold letters) found in KS domain of type I polyketide synthases. Pim00 to Pim04 correspond to the KS domains present in modules 0 to 4 of the PKS for pimaricin biosynthesis. Ery 1, Ole 5, and Rap05 and Rap10 correspond to KS domains present in the modules of PKSs for erythromycin, oleandomycin, and rapamycin, respectively, as shown in Fig. 5. Note that Pim00 lacks the active-site Cys (see text).

| KS Motif | Active Site |
|----------|-------------|
| Pim00    | Pim01       | Pim02       | Pim03       | Pim04       |
| (Val)    | Cys         | Cys         | Cys         | Cys         |
| Pim05    | Pim06       | Pim07       | Pim08       | Pim09       |
| (Val)    | Cys         | Cys         | Cys         | Cys         |
| Consensus| Pim00       | Pim01       | Pim02       | Pim03       |
| (Val)    | Cys         | Cys         | Cys         | Cys         |

**FIG. 7.** Domain organization and biosynthetic intermediates for PIMS0 and PIMS1. The pims0 gene product could constitute a loading tool for the acetate starter unit (see text). For domain designations see the abbreviations footnote and the following: CoL, carboxylic acid:CoA ligase; KR, β-ketoreductase; DH, dehydratase. The KS in PIMS0 is presumably inactive (see text).
domain, thus making necessary the interaction of module 1 KS either with free acetyl-CoA or, more likely, with a separate loading protein that would present the starter unit to module 1. Fig. 7 shows the latter possibility. The gene product of pimS0 could be such a loading protein. Its first domain (carboxylic acid-CoA ligase) could activate acetyl-CoA, forming the acyladenylate that would be transferred by the adjacent ACP to the KS of PIMS0. This KS, being inactive, would not catalyze any condensation reaction but would allow the transfer of the growing polyketide chain to the C-terminal ACP, which would feed the acetyl moiety to module 1 KS of PIMS1 for the first elongation step (Fig. 7). The cloning and specific alteration of the biosynthesis genes for this polyene might allow the engineering of novel analogs with improved antifungal properties.

Acknowledgments—We thank M. R. Rodicio for providing KC515 phage and P. F. Leadlay and J. A. Gil for the kind gift of probes from the rapamycin and the candididin biosynthesis clusters, respectively.

REFERENCES
1. O’Hagan, D. (1991) The Polyketide Metabolites, Ellis Horwood Ltd., Chichester, UK.
2. Hopwood, D. A., and Sherman, D. H. (1990) Annu. Rev. Genet. 24, 37–66
3. Staunton, J. (1991) Angew. Chem. Int. Ed. Engl. 30, 1302–1306
4. Katz, L., and Donadio, S. (1988) Annu. Rev. Microbiol. 47, 875–912
5. Scheweke, T., Aparicio, J. F., Molinár, Y., König, A., Khaw, L. E., Haydock, S. F., Oliynyk, M., Caffrey, P., Cortés, J., Lester, J. B., Böhm, G. A., Staunton, J., and Leadlay, P. F. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 7839–7843
6. Donadio, S., Slaver, M. J., McAlpine, J. B., Swanson, S. J., and Katz, L. (1991) Science 252, 675–679
7. Cortés, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J., and Leadlay, P. F. (1990) Nature 348, 176–178
8. Aparicio, J. F., Molinár, I., Scheweke, T., König, A., Haydock, S. F., Khaw, L. E., Staunton, J., and Leadlay, P. F. (1996) Gene 169, 9–16
9. Marsden, A. F. A., Caffrey, P., Aparicio, J. F., Loughran, M. S., Staunton, J., and Leadlay, P. F. (1994) Science 263, 378–380
10. Aparicio, J. F., Caffrey, P., Marsden, A. F. A., Staunton, J., and Leadlay, P. F. (1994) J. Biol. Chem. 269, 8524–8528
11. Martin, J. F. (1977) Annu. Rev. Microbiol. 31, 12–38
12. Gil, J. A., and Martin, J. F. (1997) in Biotechnology of Antibiotics (Strohl, W. R., ed) 2nd Ed., pp. 551–576, Marcel Dekker, Inc., New York
13. Omura, S., and Tanaka, H. (1984) in Macrolide Antibiotics, Chemistry, Biology, and Practice (Omura, S., ed) pp. 351–404, Academic Press, Inc., NY
14. Bolard, J. (1986) Biochim. Biophys. Acts 719, 11–17
15. Rodicio, M. R., Bruton, C. J., and Chater, K. F. (1985) Gene 34, 283–292
16. 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., and Schrepf, H. (1985) in Genetic Manipulation of Streptomyces: A Laboratory Manual, John Innes Foundation, Norwich, UK
17. Chater, K. F., Bruton, C. J., and Suárez, J. E. (1981) Gene 14, 183–194
18. Martin, J. F., and Dearn, A. L. (1975) Biochem. Biophys. Res. Commun. 71, 1103–1109
19. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
20. Hanahan, D. (1985) in DNA Cloning: A Practical Approach (Glover, D. M., ed) pp. 109–132, IRL Press at Oxford University Press, Oxford
21. Criado, L. M., Martin, J. F., and Gil, J. A. (1993) Gene 126, 135–139
22. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
23. Devereux, J., Haeberli, P., and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395
24. Bevitt, D. J., Cortés, J., Haydock, S. F., and Leadlay, P. F. (1992) Eur. J. Biochem. 204, 39–49
25. Donadio, S., and Katz, L. (1992) Gene 11, 51–60
26. Scrutton, N. S., Berry, A., and Perham, R. N. (1990) Nature 343, 38–43
27. Haydock, S., Aparicio, J. F., Molinár, I., Scheweke, T., König, A., Marsden, A. F. A., Galloway, I. S., Staunton, J., and Leadlay, P. F. (1995) FEBS Lett. 374, 246–248
28. Maxwelling, D. G., Rickards, R. W., Gaudiano, G., and Nicoletta, V. (1966) J. Antibiot. (Tokyo) 22, 545–550
29. Leskiw, B. K., Lawlor, E. J., Fernández-Abalos, J. M., and Chater, K. F. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 2461–2465
30. Brown, D. W., Adams, T. H., and Keller, N. P. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 14873–14877
31. MacNeil, D. J., Oeci, J. L., Gewain, K. M., MacNeil, T., Gibbons, P. H., Ruby, C. L., and Danis, S. J. (1992) Gene 115, 119–125
32. Waki, S. J. (1988) Biochemistry 27, 4523–4530
33. Albertini, A. M., Caramori, T., Scoffone, F., Scotti, C., and Galizi, A. (1995) Microbiology 141, 299–309
34. Meurer, G., and Hutchinson, C. R. (1995) J. Bacteriol. 177, 477–481
35. Staunton, J., Caffrey, P., Aparicio, J. F., Roberts, G. A., Béthell, S. S., and Leadlay, P. F. (1996) Nat. Struct. Biol. 3, 188–192
36. Swan, D. G., Rodríguez, A. M., Vilches, C., Méndez, C., and Salas, J. A. (1994) Mol. Gen. Genet. 242, 358–362
37. Schupp, T., Toupet, C., Cluzel, B., Neff, S., Hill, S., Beck, J. M., and Ligon, J. M. (1995) J. Bacteriol. 177, 3673–3679