Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate '*Streptomyces maritimus*': evidence for the derailment of an aromatic polyketide synthase

Jörn Piel [1,*], Christian Hertweck [1], Paul R Shipley [2], Deanna M Hunt [1], Mark S Newman [1] and Bradley S Moore [1,2]

Background: Polycyclic aromatic polyketides, such as the tetracyclines and anthracyclines, are synthesized by bacterial aromatic polyketide synthases (PKSs). Such PKSs contain a single set of iteratively used individual proteins for the construction of a highly labile poly-β-carbonyl intermediate that is cyclized by associated enzymes to the core aromatic polyketide. A unique polyketide biosynthetic pathway recently identified in the marine strain '*Streptomyces maritimus*' deviates from the normal aromatic PKS model in the generation of a diverse series of chiral, non-aromatic polyketides.

Results: A 21.3 kb gene cluster encoding the biosynthesis of the enterocin and wailupemycin family of polyketides from '*S. maritimus*' has been cloned and sequenced. The biosynthesis of these structurally diverse polyketides is encoded on a 20 open reading frames gene set containing a centrally located aromatic PKS. The architecture of this novel type II gene set differs from all other aromatic PKS clusters by the absence of cyclase and aromatase encoding genes and the presence of genes encoding the biosynthesis and attachment of the unique benzoyl-CoA starter unit. In addition to the previously reported heterologous expression of the gene set, in vitro and in vivo expression studies with the cytochrome P-450 EncR and the ketoreductase EncD, respectively, support the involvement of the cloned genes in enterocin biosynthesis.

Conclusions: The enterocin biosynthesis gene cluster represents the most versatile type II PKS system investigated to date. A large series of divergent metabolites are naturally generated from the single biochemical pathway, which has several metabolic options for creating structural diversity. The absence of cyclase and aromatase gene products and the involvement of an oxygenase-catalyzed Favorskii-like rearrangement provide insight into the observed spontaneity of this pathway. This system provides the foundation for engineering hybrid expression sets in the generation of structurally novel compounds for use in drug discovery.

[1]Department of Chemistry, P.O. Box 351700, University of Washington, Seattle, WA 98195-1700, USA
[2]Division of Medicinal Chemistry, College of Pharmacy, P.O. Box 210207, University of Arizona, Tucson, AZ 85721-0207, USA

*Present address: Max Planck Institute of Chemical Ecology, Carl-Zeiss-Promenade 10, 07745 Jena, Germany.

Correspondence: Bradley S Moore
E-mail: more@pharmacy.arizona.edu

Keywords: Antibiotic biosynthesis; Benzoic acid; Marine bacterium; Oxygenase; Polyketide synthase

Received: **14 June 2000**
Revisions requested: **19 July 2000**
Revisions received: **27 September 2000**
Accepted: **10 October 2000**

Published: **1 November 2000**

Chemistry & Biology 2000, **7**:943–955

1074-5521/00/$ – see front matter

PII: S1074-5521(00)00044-2

Introduction

Bacterial aromatic polyketides, such as the clinically important metabolites oxytetracycline, doxorubicin and mithramycin, represent a large group of biologically active natural products. Although well over 500 in number, these multicyclic aromatic polyketides have largely been characterized from actinomycetes and belong to just a few common structural types that include benzoisochromanequinones, tetracyclines, angucyclines, anthracyclines, tetracenomycins and aureolic acids. Representative biosynthetic gene clusters have been cloned and sequenced from each group and belong to a homologous family called the aromatic or type II polyketide synthases (PKSs) [1,2]. Each PKS contains a single set of iteratively used individual proteins (the two β-ketoacyl:acyl carrier protein (ACP) synthase subunits KS_α and KS_β, an ACP, and possibly a malonyl-CoA:ACP transacylase) for the construction of an unmodified polyketide chain. Associated proteins reduce (via ketoreductases (KRs)) and cyclize (via cyclases and aromatases) the highly labile linear poly-β-ketoacyl thioester intermediate. Nature only provides a few different core polyketide structures by varying the chain length, position of the ketoreduction, and regiochemistry of the cyclization events. Structural diversity within the aromatic polyketides is rather dictated by tailoring enzymes, such as oxygenases, methylases, and glycosylases, during post-cyclization events.

We recently described an unprecedented type II PKS gene set from the marine isolate '*Streptomyces maritimus*' that

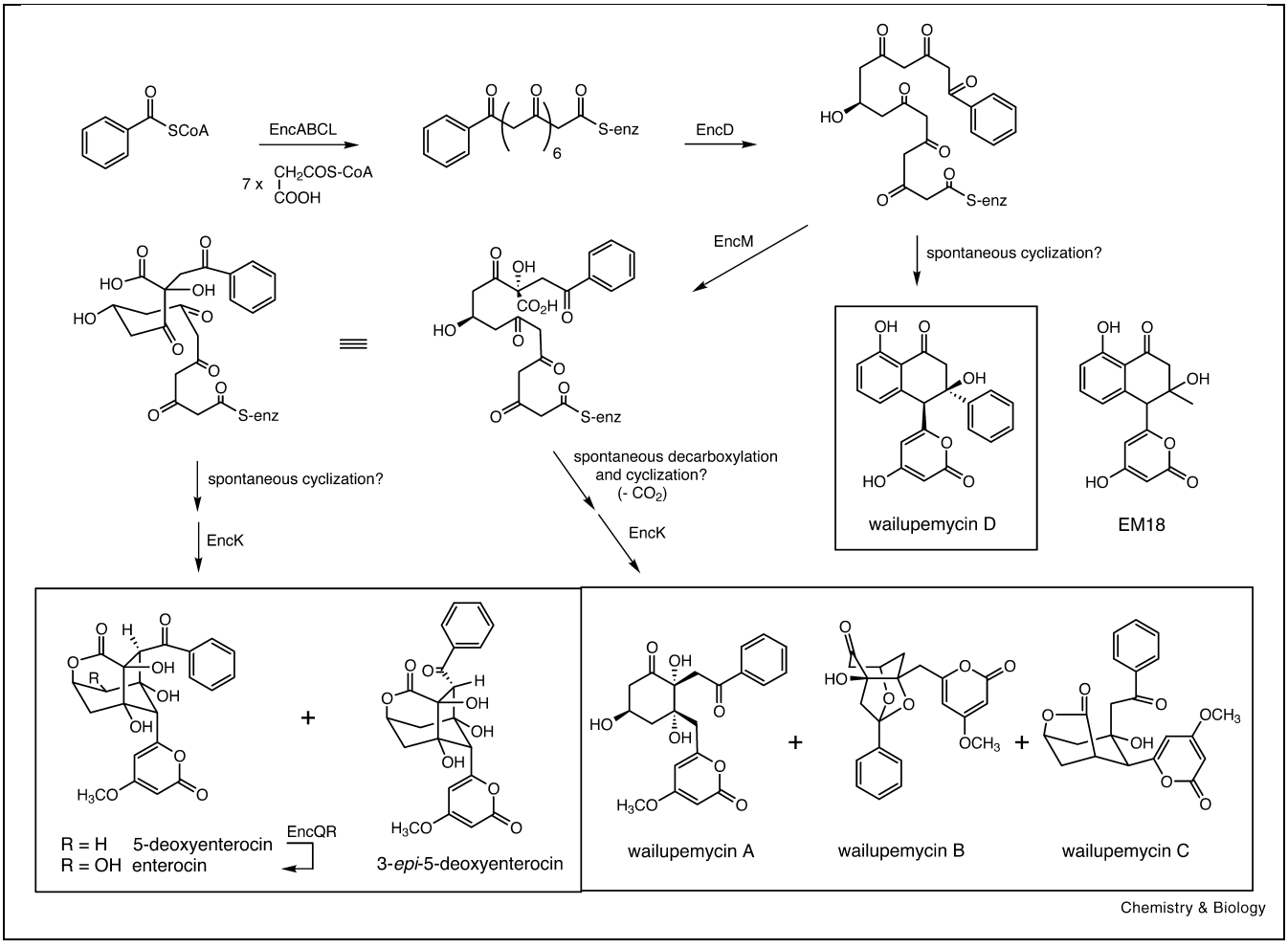

Figure 1. Proposed biosynthetic pathway from benzoyl-CoA to enterocin and the wailupemycins.

encodes the biosynthesis of a diverse series of three-dimensionally, rigid polyketides known as the enterocins and wailupemycins (Figure 1) [3]. This series of structurally diverse bacteriostatic agents is derived from a single biosynthetic pathway with numerous metabolic options. The branched octaketide enterocin is the major metabolite produced by '*S. maritimus*' [4] and has additionally been identified from edaphic streptomycetes [5–7] and a marine invertebrate [8]. Enterocin is derived from an uncommon benzoate starter unit and seven malonate molecules and undergoes a rare Favorskii-like rearrangement [7]. 5-Deoxyenterocin and 3-*epi*-5-deoxyenterocin differ from enterocin only in hydroxyl group functionalization at C5 and the stereochemistry at C3. Three additional α-pyrone containing metabolites, wailupemycins A–C [4], are probably derived from a common branched intermediate that decarboxylates to yield the structurally diverse set. Wailupemycin D, on the other hand, is derived from a common pre-rearranged polyketide intermediate [3].

Heterologous expression of the enterocin (*enc*) biosynthesis gene cluster from '*S. maritimus*' in the genetically engineered host strain *Streptomyces lividans* K4-114 [9] verified that all of the genetic information for the production of the enterocin–wailupemycin series was contained on the cosmid clone pJP15F11 [3]. The *enc* cluster represents the most versatile type II PKS system investigated to date as at least six divergent compounds are derived from this single cluster. Its novel features provide the foundation for engineering hybrid expression systems with more typical type II PKS gene sets in the generation of diverse ranges of novel compounds for use in drug discovery. Hence, we performed a detailed analysis of the *enc* biosynthetic gene cluster in '*S. maritimus*'. The sequence includes not only the genes encoding polyketide assembly, but also for the formation of the novel benzoate starter unit. In addition, this is the first report of the genetic organization of a secondary metabolic gene cluster from a marine actinomycete.

Results

Cloning and sequencing of the *enc* cluster

The common α-pyrone residue in the enterocins and wailupemycins is a regular feature of many mutant and recombinant polyketides generated from type II PKS systems and served to indicate that this series of metabolites is probably synthesized via an aberrant type II rather than a modular type I polyketide pathway. Consequently, we initially speculated that the enterocin/wailupemycin biosynthesis gene cluster contained homologues to *act*I (minimal PKS) and *act*III (KR) genes from the actinorhodin pathway of *Streptomyces coelicolor* A3(2) [10]. A '*S. maritimus*' cosmid library was thus constructed in SuperCos I and screened with a digoxigenin-labeled *act*I-ORF1 (KS_α) probe. Restriction mapping of the positive cosmids revealed three distinct type II PKS gene sets in the genome. Two of these clusters additionally hybridized to the *act*III probe. Spot sequencing of both clusters verified the presence of type II PKS genes and further revealed the presence of putative genes encoding a phenylalanine ammonia lyase (PAL) and a benzoyl-CoA ligase on cosmid pSS9A6. Since these genes are probably involved in the synthesis of the unique benzoyl-CoA starter unit, we selected pSS9A6 for further analysis. The identification of the *enc* cluster on this cosmid was confirmed by the successful expression of the entire enterocin gene cluster from the pSS9A6 homologue pJP15F11, a derivative of the *Escherichia coli–Streptomyces* shuttle cosmid pOJ446, in *S. lividans* K4-114 [3].

A contiguous 21.3 kb region of pSS9A6 was fully sequenced on both strands. The nucleotide (nt) sequence was deposited at GenBank under the accession number AF254925. The sequence was analyzed for putative open reading frames (ORFs) with the FRAME program [11] and aligned with homologous sequences in the databases. Twenty putative complete ORFs were designated *encA* through *encT* and consist of at least 14 probable structural genes and four regulation and resistance genes (Figure 2, Table 1). All ORFs are preceded by putative ribosome-binding sites with common consensus motifs [12]. The sequence showed the typical GC content for *Streptomyces* DNA at 68.8% in the coding region. Flanking the *encE* and *encT* genes are ORFs with high homology to hypothetical proteins of unknown functions and to housekeeping proteins, thus delineating the boundaries of the *enc* gene cluster.

Probable PKS genes

The minimal *enc* PKS, EncABC, is encoded by a set of genes architecturally similar to most other type II PKS clusters and is centrally located in the cluster. EncA is either translated from a 1272 nt ORF, starting with a GTG codon, or from an ATG codon 51 nt further upstream. Both start codons are preceded by convincing putative ribosome-binding sites at appropriate distances. The protein sequence of the KS_α homologous EncA contains a candidate residue for the active cysteine residue at position 170 or 187, respectively, that anchors the acyl group during chain procession, as well as the conserved acyltransferase (AT) motif GHS 178 aa further downstream [10]. Overlapping the *encA* stop codon by 4 nt, *encB* encodes a protein of 407 aa in length with strong homology to KS_β enzymes. As in other type II PKS systems, the KS_β EncB exhibits se-

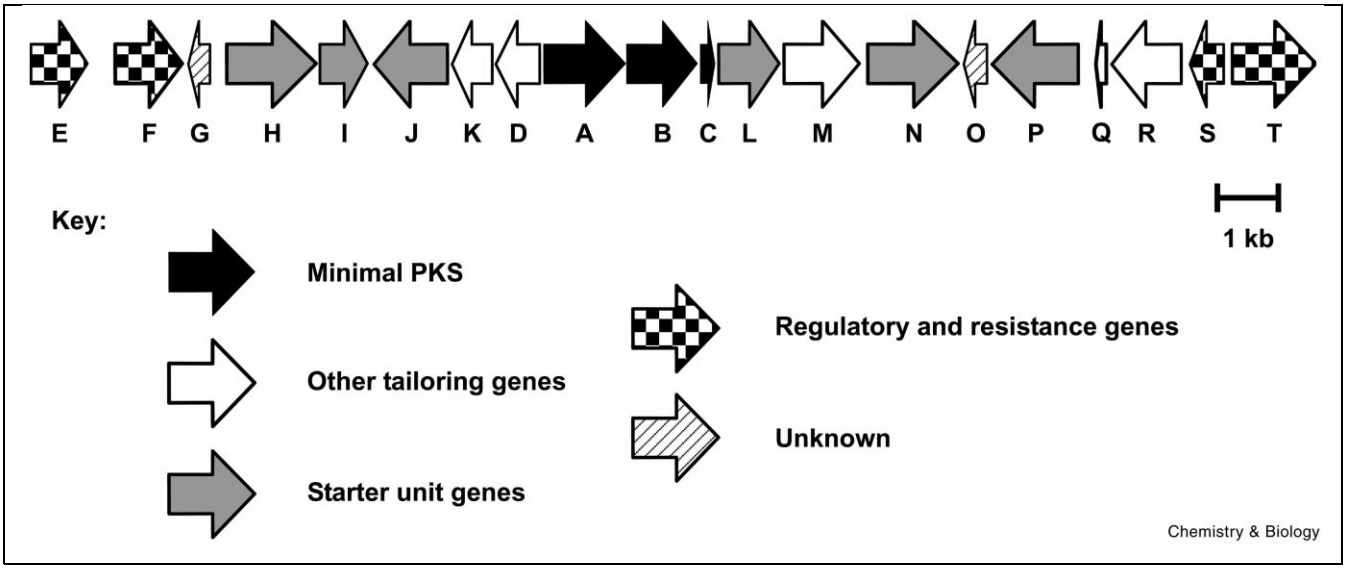

Figure 2. Organization of the enterocin biosynthetic gene cluster (*enc*) in '*S. maritimus*'. Each arrow represents the direction of transcription of an ORF.

Table 1
Deduced functions of the ORFs in Figure 2.

Protein	Amino acids (no.)	Proposed function	Sequence similarity (protein, origin)	Similarity/identity (%)	Protein accession number	Reference
EncA	423	KS_α	NcnA, *Streptomyces arenae*	80%/66%	AAD20267	[68]
EncB	407	KS_β	Frn2, *S. roseofulvus*	68%/57%	AAA19617	[22]
EncC	82	ACP	ActI ORF3, *S. coelicolor*	65%/43%	CAA45045	[10]
EncD	266	KR	AknIII, *Streptomyces galilaeus*	73%/59%	BAA03128	[69]
EncE	163	Regulatory protein	NolA, *Bradyrhizobium elkanii*	64%/44%	P50329	[70]
EncF	283	Transcriptional activator	FrnG, *S. roseofulvus*	62%/46%	AAC18102	[22]
EncG	163	Unknown	MtmX, *Streptomyces argillaceus*	51%/37%	S58172	[38]
EncH	535	Acyl-CoA ligase	CaiC, *E. coli*	51%/33%	I41013	[71]
EncI	258	Enoyl-CoA hydratase	YngF, *B. subtilis*	52%/35%	CAA74218	[72]
EncJ	400	β-Oxoacyl-CoA thiolase	PaaJ, *E. coli*	42%/29%	CAA66099	[16]
EncK	241	Methyltransferase (MT)	PemT, *A. aceti*	40%/26%	BAA34057	
EncL	340	AT	DpsD, *S. peucetius*	57%/43%	AAA65209	[21]
EncM	464	FAD-dependent oxygenase	MitR, *S. lavendulae*	48%/34%	AAD28454	[25]
EncN	522	Aryl-CoA ligase	BadA, *R. palustris*	57%/41%	AAA92151	[17]
EncO	130	Unknown				
EncP	523	PAL	HutH, *B. subtilis*	58%/37%	P10944	[73]
			PAL, *Trifolium subterraneum*	52%/31%	P45734	[74]
EncQ	81	Ferredoxin	RapO, *Streptomyces hygroscopicus*	54%/44%	CAA60464	[28]
EncR	401	Cytochrome P-450 hydroxylase	Orf3, *S. lavendulae*	60%/47%	AAD28449	[25]
EncS	215	Regulatory protein	BetI, *E. coli*	43%/26%	P17446	[75]
EncT	482	Efflux protein	CmcT, *S. lactamdurans*	55%/43%	Q04733	[33]

quence similarity to the KS_α EncA and has a glutamine substituted for the conserved cysteine residue. In agreement with the structure of enterocin, EncB is most homologous to KS_β proteins involved in the biosynthesis of other octaketides, such as actinorhodin, frenolicin, and nogalamycin. The third enzyme of the minimal *enc* PKS is encoded by *encC*, an ORF of 249 nt, which bears significant similarity on an amino acid level to ACPs. The putative 4′-phosphopantetheine-binding serine residue is located at position 39 within a highly conserved region of the protein sequence [10].

Adjacent to the putative minimal *enc* PKS in an upstream direction, *encD* shows good homology to KR genes. The protein contains the characteristic short-chain alcohol dehydrogenase signature sequence between aa 148 and 167 (PROSITE: PS00061). As with most other KRs associated with type II PKSs, EncD probably catalyzes the reduction of the C-9 keto group of the nascent polyketide chain. Notably absent, however, is the presence of cyclase or aromatase genes that typify aromatic PKS gene sets. These gene products are integral components of these systems and direct the cyclization mode of the highly reactive open chain polyketide precursor by performing regiospecific aldol reactions. This significant architectural difference between the *enc* cluster and all known bacterial type II PKS gene clusters provides partial insight to the observed spontaneity of the *enc* pathway.

Genes putatively involved in starter unit synthesis and attachment

The benzyl moiety in enterocin is derived from the rare bacterial PKS starter unit benzoic acid [7,13]. ORFs putatively involved in the biosynthesis of the benzoyl-CoA starter unit are arranged on either side of the minimal PKS on four transcripts. The first step in the proposed benzoic acid biosynthetic pathway in '*S. maritimus*' (Figure 3) involves the gene product of *encP* which shows significant homology to PALs and histidine ammonia lyases (histidases, HALs) (Table 1). EncP contains a conserved motif around Ser^{144}, which is the probable precursor of the modified dehydroalanine residue in the 4-methylidene-imidazol-5-one prosthetic group [14]. Sequence analysis further suggests that EncP functions as a PAL and not a HAL, which is consistent with benzoate synthesis in plants where PAL catalyzes the initial pathway reaction, the conversion of L-phenylalanine to *trans*-cinnamic acid [14,15]. In striking contrast to the high number of known PALs from plants and fungi, a bacterial PAL gene or protein has previously not been described. Activation of cinnamic acid to its CoA thioester is proposed as the second biosynthetic reaction in this series and is probably catalyzed by the *encH* gene product. EncH is homologous to long-chain fatty acid-CoA ligases (Figure 4) and contains a putative AMP-binding site (Pfam: PDOC00427). The deduced protein sequences of *encI* and *encJ* show high similarities to enoyl-CoA hydratases and β-ketothiolases (acyl-CoA acetyltransferase), respectively. Among the closest relatives of EncJ is the uncharacterized enzyme PaaJ from the phenyl-

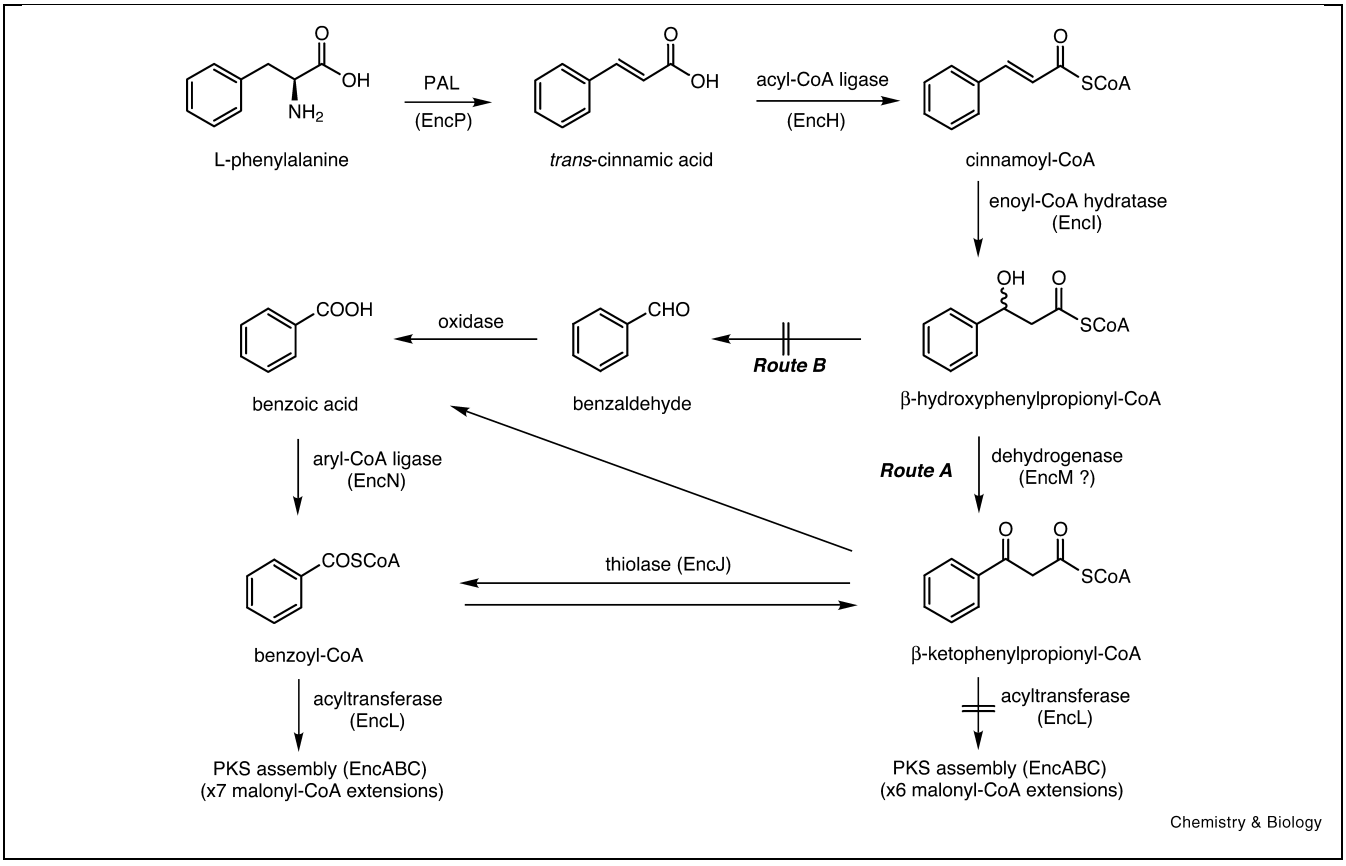

Figure 3. Possible biosynthetic pathways leading to benzoic acid in '*S. maritimus*'. Route A, β-oxidative degradation; route B, retro-aldol degradation with a possible additional chain extension performed by the thiolase EncJ.

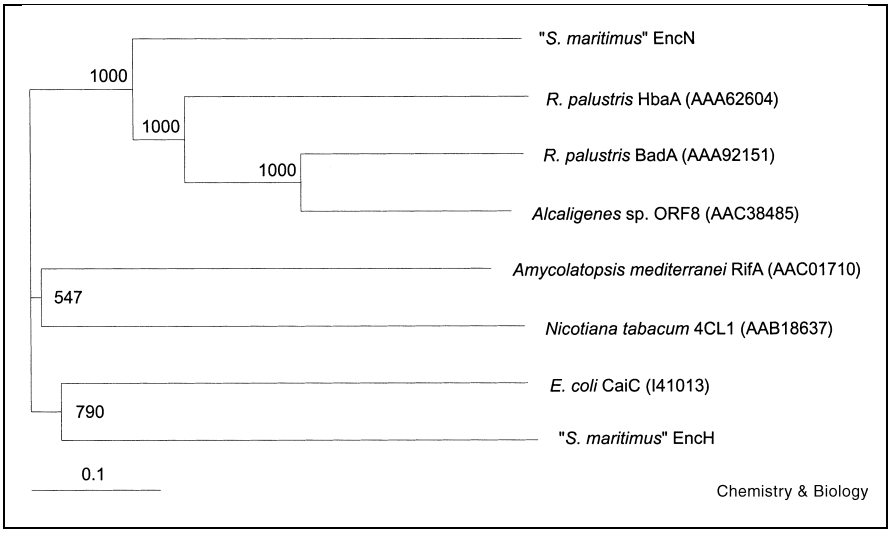

Figure 4. Phylogenetic tree of EncH, EncN and related acyl- and aryl-CoA ligases. Sequences were retrieved from GenBank (accession numbers given in parentheses) and aligned with ClustalX. The scale bar indicates 0.1 changes per nt.

acetate degradation pathway of *E. coli* [16], which suggests a structurally related substrate for EncJ. Homologues to EncH, EncI, and EncJ are typically constituents of β-oxidation pathways, implying a similar route for cinnamate degradation to benzoate. However, a corresponding ORF encoding a hydroxyacyl-CoA dehydrogenase, the fourth protein necessary for a complete β-oxidation series, was not identified in the *enc* cluster.

Database searches revealed two ORFs that may function in the activation and transfer of the benzoate starter unit. The acyl-CoA ligase EncN putatively functions in the activation of the benzoate starter unit. The putative protein has highest homology to aryl-CoA ligases, such as BadA and HbaA from *Rhodopseudomonas palustris*, which activate benzoate [17] and hydroxybenzoate [18], respectively, and ORF8 from the 3-chlorobenzoate degradation cluster in an *Alcaligenes* sp. [19] (Figure 4). EncN also shows good homology to the CoA ligase domain from RifA [20], which is involved in the selection of the 3-amino-5-hydroxybenzoic acid starter unit for the rifamycin PKS. As with EncH, the deduced protein of EncN contains the amino acid motif of an AMP-binding domain common to CoA ligases. A second enzyme that has been postulated in the selection of non-acetate starter units is the AT. A counterpart to this protein was also found in the *enc* cluster, represented by the 340 aa gene product of *encL*, located immediately downstream of the putative ACP. This gene product is similar to AT proteins or domains of polyketide and fatty acid synthases and putatively catalyzes the transfer of benzoyl-CoA onto the ACP. The sequence features the conserved GHS motif that is probably involved in the formation of the acyl-enzyme intermediate. EncL is most similar to the daunorubicin DpsD [21] and the frenolicin FrnK [22] ATs, which participate in the transfer of non-acetate starter unit acyl-CoAs in type II PKS systems.

Putative tailoring genes

The biosynthesis of enterocin probably involves two oxidative reactions that are catalyzed by separate tailoring enzymes. The Favorskii-like rearrangement is unprecedented in post-type II PKS assembly and putatively involves the oxidative rearrangement of the highly reactive poly-β-ketoacyl thioester intermediate before polyaromatic cyclization can take place. Oxidation of 5-deoxyenterocin to enterocin would require the second oxygenase. Two candidate oxygenase encoding genes, an unusual flavin-dependent monooxygenase and a cytochrome P-450 hydroxylase encoded by *encM* and *encQ*, respectively, are located downstream of the minimal *enc* PKS on separate transcripts. EncM is similar to a group of oxygenases that contain covalently bound flavin-adenine dinucleotide (FAD) (Figure 5). Position 78 contains a conserved histidine residue, which is the FAD-binding site in homologous proteins such as 6-hydroxy-D-nicotine oxidase from *Arthrobacter oxidans* [23]. Other members of this uncommon class of enzymes include McrA [24], MmcM, and MitR [25], which are involved in the resistance and biosynthesis of mitomycin, DnrW from the daunorubicin biosynthesis cluster (accession number AAD00354), and the berberine bridge enzyme from plants [26]. The diverse oxidative reactions catalyzed by these enzymes are often not well understood and unfortunately provide no direct indication of the function of EncM. The second putative oxygenase is encoded by *encR*. The deduced protein shows good homology to hydroxylases of the cytochrome P-450 family. Its sequence displays characteristic P-450 motifs [27], including the highly conserved consensus sequences associated with heme and oxygen-binding sites. The essential cysteine residue that serves as the heme iron ligand is positioned at aa 350. Downstream and apparently translationally coupled to *encR* by a probable 7 nt overlap lies *encQ*. This ORF is probably translated into a ferredoxin-type protein of 81 aa in length and is most similar to RapO from the rapamy-

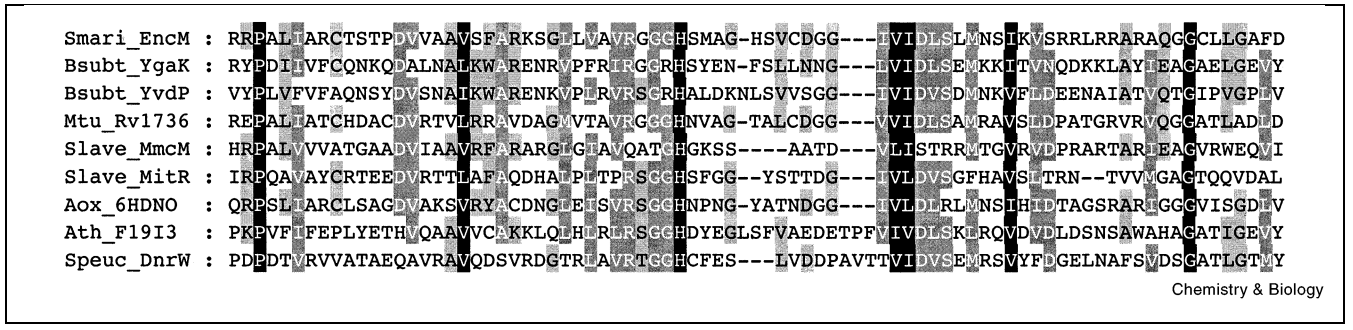

Figure 5. Alignment of homologues to EncM around the conserved FAD-binding histidine. Sequence identification codes are Smari_EncM, '*S. maritimus*' EncM; Bsubt_YgaK, *Bacillus subtilis* hypothetical protein YgaK (accession number CAB04812); Bsubt_YvdP, *B. subtilis* hypothetical protein YvdP (CAB08045); Mtu_Rv1736, *Mycobacterium tuberculosum* hypothetical protein Rv1736 (CAB03710); Slave_McmM, *Streptomyces lavendulae* McmM (AAD32736); Slave_MitR, *S. lavendulae* MitR (AAD28454); Aox_6HDNO, *A. oxidans* 6-hydroxy-D-nicotine oxidase (448109); Ath_F19I3, *Arabidopsis thaliana* putative berberine bridge enzyme (AAD25758); Speuc_DnrW, *Streptomyces peucetius* DnrW (AAD00354).

cin biosynthetic gene cluster [28]. Of the four possible iron-binding cysteine residues, three are present in the protein sequence at positions 21, 27 and 62. As with RapO, the fourth active site cysteine residue at position 24 is exchanged against serine. The coupling of *encR* and *encQ* implies that EncQ is involved in the reoxidation of EncR. A gene encoding an NADPH-ferredoxin reductase, the second enzyme necessary for P-450 reoxidation, was not found in the *enc* cluster and is probably supplied by primary metabolism.

The gene product of *encK*, which neighbors the putative KR, shows high homology to a protein from *E. coli* with unknown function (accession number P77460) and weaker homology to a phosphatidylethanolamine *N*-methyltransferase from *Acetobacter aceti* (accession number BAA34057). A probable *S*-adenosyl methionine-binding motif between positions 61 and 165 was identified (PROSITE: PS50193) that is characteristic for methyltransferases. EncK could thus be responsible for the introduction of the *O*-methyl groups of enterocin and the wailupemycins A–C. If so, this enzyme is remarkably versatile having the ability to methylate α-pyrone units belonging to significantly different substrates.

Regulatory and resistance genes

EncE resembles regulatory proteins of the MerR family. These proteins bind DNA through an N-terminal helix-turn-helix motif and turn into transcriptional activators upon binding to an inducer [29]. Separated from *encE* by a gap of 850 nt lies *encF*. The predicted protein is highly similar to transcriptional activators of the *Streptomyces* antibiotic regulatory protein (SARP) family, which are involved in morphological differentiation and in the regulation of secondary metabolism [30]. *encF* is similar to SARP genes from other type II PKS clusters, such as *act*II-ORF4 [31] and *frnG* [32]. During secondary metabolism, SARPs bind to repeats of specific DNA sequences to enhance the interaction of promoter regions with RNA polymerases. Repeats that could be potential binding sites for EncF were identified between the divergent ORFs *encF* and *encG* (i.e. a tetramer of C[GA][GC]TGG). A third protein with potential regulatory function is encoded by *encS* and contains a putative helix-turn-helix motif at aa 52–67 and a motif common to transcriptional repressors of the TetR family (Pfam: PF00440). Divergently transcribed from EncS, EncT strongly resembles efflux proteins involved in resistance mechanisms [33]. Several other determinants of resistance against antibiotics have been shown to consist of such divergent repressor–resistance gene pairs [34,35]. As in these systems, a region with dyad symmetry is present between *encS* and *encT* that probably functions as the binding site for the repressor. *encF* (63.5%) exhibits low GC content and contains a number of codons that are rare in *Streptomyces* [36]. A rare TTA codon, which is present in many SARP genes and is known to play a role in *Streptomyces* regulation and resistance [37], is present in *encT*. However no such codon is present in the SARP homologue *encF*.

Genes of unknown function

EncG is homologous to a family of proteins found in several type II PKS pathways, including MtmX (mithramycin) [38], DpsH (daunorubicin) [39], ActVI-ORFA (actinorhodin) [40], and Gra-ORF31 (granaticin) [41]. The *encG* homologues MtmX and DpsH have been implicated in the cyclization process of the linear polyketide [42,43], whereas *act*VI-ORFA has recently been proposed to stabilize the multicomponent, type II PKS complex and assist in pyran ring formation [44]. The putative gene product of the short ORF *encO* does not resemble any protein in the databases.

In vivo expression and functional analysis of *encD*

Sequence analysis of *encD* suggests a functional role of the

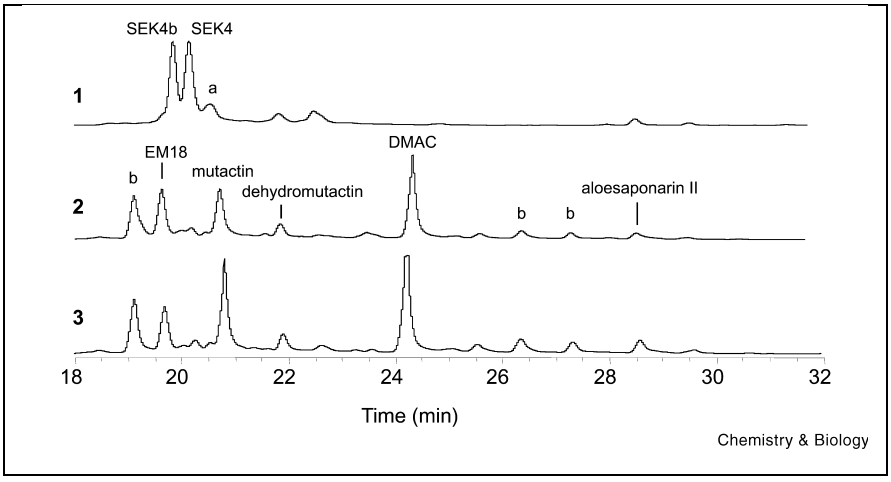

Figure 6. LC-APCI-MS traces of engineered polyketides from (1) *S. lividans* K4-114 (pSEK4), (2) *S. lividans* K4-114 (pSEK4/pCH20), and (3) *S. lividans* K4-114 (pRM5). HPLC was monitored at 254 nm. Chromatographic peaks labeled with letters 'a' ($MH^+ = 301$) and 'b' ($MH^+ = 285$) are uncharacterized polyketides.

gene product in the reduction of the C-9 carbonyl in the linear polyketide before rearrangement and subsequent cyclization (Figure 1). The apparent function of EncD is equivalent to other KRs associated with type II PKS systems, including ActIII, which catalyzes a similar reduction in actinorhodin biosynthesis [45]. In order to evaluate whether EncD is functionally equivalent to ActIII, we constructed the *encD* containing plasmid pCH20 and co-expressed it with pSEK4 [46] in the host strain *S. lividans* K4-114. pCH20 is based on the integrating bacteriophage plasmid pSET152 and was constructed by cloning a 0.9 kb *Fse*I–*Nsi*I fragment from the '*S. maritimus*' cosmid clone pSS9A6 into the pSET152 derivative pKOS010 [47]. The expression vector pSEK4 harbors the *act* PKS genes *act*I-ORFs 1–3, *act*VII, and *act*IV and is related to pRM5 which additionally carries the activated *act*III KR [46]. Liquid chromatography (LC)-APCI-mass spectrometry (MS) analysis of organic extracts from *S. lividans* K4-114 (pSEK4/pCH20) and K4-114 (pRM5) gave nearly identical profiles (Figure 6), thus verifying that EncD is a KR of similar function as ActIII. As expected [48], 3,8-dihydroxymethylanthraquinone carboxylic acid and the decarboxylated product aloesaponarin II were identified by LC-MS analysis of the *S. lividans* K4-114 (pSEK4/pCH20) and K4-114 (pRM5) extracts, along with several other known and possibly new analogues.

Overproduction and functional analysis of EncR

In order to evaluate which oxidative reaction is catalyzed by the putative cytochrome P-450 hydroxylase EncR, we first incubated '*S. maritimus*' with the monooxygenase inhibitor ancymidol [49]. This experiment resulted in the accumulation of 5-deoxyenterocin with concomitant loss of enterocin production, suggesting that the final reaction in enterocin biosynthesis involves a P-450-catalyzed hydroxylation of 5-deoxyenterocin. In a series of cell-free transformations, 5-deoxyenterocin was converted to enterocin under cytochrome P-450 conditions. The cell-free extract of '*S. maritimus*' was prepared from mycelia obtained after passage of cells through a French press. The desalted ammonium sulfate fraction was incubated with 5-deoxyenterocin in the presence of NADPH and the P-450 electron transport proteins ferredoxin and ferredoxin-$NADP^+$ reductase. Products were analyzed by LC-MS, isolated by preparative thin layer chromatography, and compared with authentic standards of deoxyenterocin and enterocin by nuclear magnetic resonance.

The gene encoding EncR was initially cloned into the protein expression plasmid pRSETb. Unfortunately, expression of the poly-histidine-tagged EncR resulted in large quantities of insoluble protein and efforts to solubilize the protein were unsuccessful. EncR was alternatively cloned into pMAL-c2x, introducing a maltose-binding-protein (MBP) affinity tag at the amino-terminus. The resulting MBP-EncR fusion protein was purified on an amylose resin column (Figure 7) and showed a requirement for ferredoxin, ferredoxin-$NADP^+$ reductase, and NADPH. LC-MS analysis of the reaction products demonstrated conversion of 5-deoxyenterocin to enterocin (Figure 7), thus establishing EncR as a 5-deoxyenterocin hydroxylase. Furthermore, MBP-EncR was inhibited by ancymidol, con-

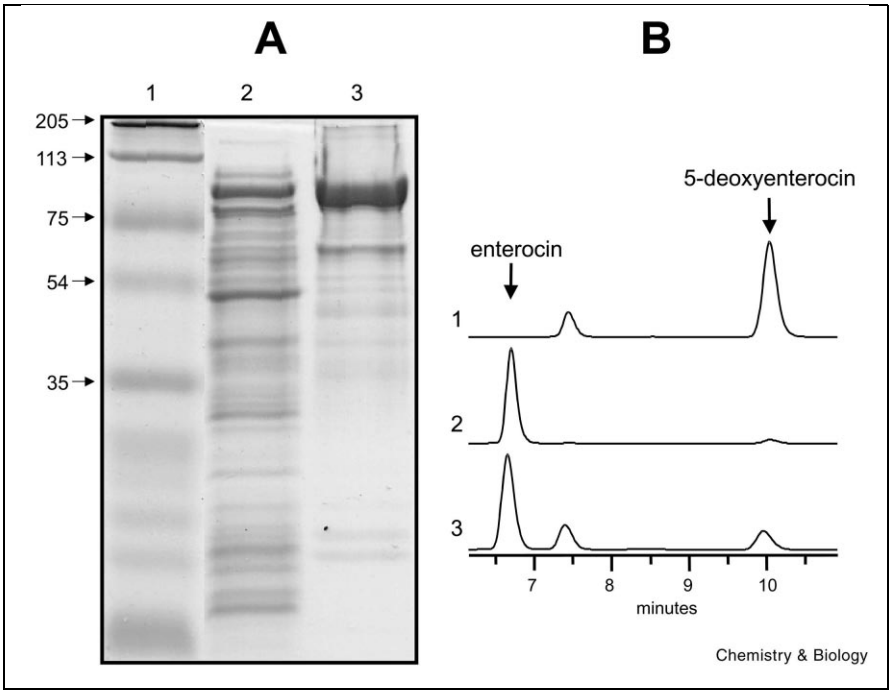

Figure 7. (A) SDS–polyacrylamide gel showing the MBP-EncR purified protein preparation. Lane 1: molecular weight markers; lane 2: cell-free preparation; lane 3: MBP-EncR (calculated molecular weight is 86.6 kDa). **(B)** HPLC traces of standards 5-deoxyenterocin (chromatogram 1) and enterocin (chromatogram 2) and the MBP-EncR catalyzed reaction mixture (chromatogram 3).

firming that the earlier observation at the in vivo level was due to ancymidol inhibition of the cytochrome P-450 EncR.

Discussion

In the work presented here, we report the sequence of the enterocin biosynthesis gene cluster that spans 21.3 kb and involves 20 ORFs (Figure 2). A detailed sequence analysis of the *enc* gene set revealed several features that are unprecedented in type II PKS catalyzed polyketide assembly and supports a proposed model of biosynthesis to this unusual set of metabolites (Figure 1). Novel biochemical features discussed at length below include the observed spontaneity of the biosynthetic pathway, the carbon rearrangement of the linear poly-β-carbonyl intermediate, and the biosynthesis and attachment of the uncommon benzoate starter unit.

The *enc* gene set contains the typical arrangement of the three components KS_α, KS_β, and ACP that comprise a minimal PKS. The first noticeable architectural departure of the *enc* cluster from all other type II PKS gene sets is the absence of cyclase and aromatase encoding genes. This key difference is mirrored in the diverse series of novel metabolites that uniquely arise from the *enc* encoded pathway [3]. Cyclases act as chaperones in catalyzing the cyclization of the labile poly-β-carbonyl intermediate away from a myriad of chemically feasible reaction products and into a single polycyclic aromatic product. When cyclases are unnaturally removed from their coupled minimal PKS, spontaneous chemistry can result as the minimal PKS relies on the stabilizing effects of the cyclase and other subunits to ensure directed assembly [50]. Thus, the natural absence of cyclases in the *enc* system provides molecular support for the metabolic diversity in this pathway as well as in engineered systems in which cyclases have been unnaturally removed [51].

The second unprecedented feature in this pathway involves the derailment of the aromatization of the highly reactive poly-β-carbonyl intermediate by a Favorskii-like rearrangement to yield non-aromatic polyketide products. This rare metabolic process has been proposed on the basis of labeling experiments in only a few other natural product pathways [52–54] and conceivably involves the action of an oxygenase. Taking the oxidative nature of this carbon rearrangement into account, the flavin-dependent oxygenase EncM is a probable candidate for catalyzing this reaction as the cytochrome P-450 EncR rather catalyzes the hydroxylation of 5-deoxyenterocin. The flavin oxygenase EncM has been tentatively assigned as the 'favorskiiase' and its role in this reaction is under evaluation. Interestingly, a similar proposal has been presented in dinoflagellate polyether biosynthesis to account for the interrupted pattern of acetate units in the polyketide chain [54].

We propose that the unprecedented α-oxidation and Favorskii rearrangement take place on the linear octaketide. The single oxygenase probably complexes with the minimal *enc* PKS since poly-β-ketides are notoriously reactive [50] and would likely spontaneously cyclize if passed onto a non-associated protein. The rearrangement, however, probably occurs after the EncD-catalyzed C-9 ketoreduction to account for the formation of wailupemycin D, a probable misadventure byproduct. This compound is structurally related to the *act/whiE* recombinant product EM18 (Figure 1) [55] and provides further evidence that the enterocin family of polyketides is derived from a rearrangement of a linear poly-β-ketide. In addition to catalyzing the Favorskii-like rearrangement, the favorskiiase may 'catalyze' the cyclization of the branched intermediate by limiting its conformational mobility to allow intramolecular aldol condensations and provide the carbon skeleton of the major metabolite enterocin. Spontaneous decarboxylation and cyclization of this highly labile intermediate would account for the branch in the pathway to provide the minor byproducts 3-*epi*-5-deoxyenterocin and wailupemycins A–C.

Another enzyme involved in post-polyketide assembly is the highly versatile methyltransferase EncK. This enzyme putatively has the remarkable ability to methylate significantly different substrates. The aberrant metabolite wailupemycin D is the only member in the characterized series whose α-pyrone ring is not methylated. EncK may alternatively act on an early general intermediate, implying an early release from the PKS. This scenario would necessitate that all cyclization events are post-assembly on the unbound substrate.

A third novel feature of the *enc* pathway involves the benzoate starter unit. The vast majority of type II PKSs utilize acetate as a starter unit. The exceptions identified to date include daunorubicin (propionate) [21], oxytetracycline (either malonamide or malonate) [56], R1128A–D (acetate, propionate, isobutyrate, and butyrate) [57], and putatively frenolicin (butyrate) [22]. Enterocin with its benzoate starter unit is the first aromatic addition to this limited set of aliphatic starter units. The process of starter unit selection is not well understood and does not seem to be catalyzed by homologous enzymes in each pathway. A conceivable mechanism could be the activation of the starter acid by an acyl-CoA ligase and the subsequent loading onto the ACP by an AT. In accordance with this proposal, AT genes have been found in the oxytetracycline, daunorubicin, R1128, and frenolicin clusters, while a dedicated AT is absent in all known clusters involved in the biosynthesis of acetate-primed polyketides. In addition to an AT, the daunorubicin cluster uniquely contains a β-ketoacyl:ACP synthase III that is involved in starter unit selection [58,59]. An acyl-CoA ligase gene has so far only been reported in the oxytetracycline cluster [56], and the expression of a subset of oxytetracycline genes missing both the CoA ligase and

the AT resulted in the production of acetate-primed compounds [56]. The starter unit selection in enterocin biosynthesis is probably determined more similarly to oxytetracycline than to daunorubicin or R1128, as the enterocin cluster contains genes homologous to an aryl-CoA ligase and an AT.

Benzoic acid is a very rare bacterial metabolite and has been implicated in only one other bacterial secondary metabolic pathway where it serves as a primer unit for the soraphen A modular type I PKS [60]. Two benzoate pathways have been characterized in plants and fungi, which involve conversion of phenylalanine to *trans*-cinnamic acid by PAL (Figure 3). The routes diverge at this intermediate, one involving a β-oxidation pathway (route A) and the second a retro-aldol path through benzaldehyde (route B). The only known bacterial benzoate pathway is anaerobic and involves transamination of phenylalanine to phenylpyruvate followed by two successive α-oxidative decarboxylations [61].

Several ORFs putatively involved in the biosynthesis of the enterocin benzoyl-CoA starter unit are arranged on either side of the minimal *enc* PKS. Sequence analysis supports the involvement of a plant-like benzoate pathway, and feeding experiments with labeled precursors support the β-oxidation route (Figure 3, route A) as all of the proposed intermediates enriched enterocin with high incorporation rates [13]. The conversion of phenylalanine through *trans*-cinnamic acid to benzoate, although very common in plants and fungi, has not been reported in bacteria. The initial biosynthetic step putatively involves a bacterial PAL catalyzed conversion by the *encP* gene product. Although ubiquitous in plants and in some fungi, surprisingly neither PAL nor its encoding gene has been isolated from bacteria. A PAL was, however, detected in a cell-free preparation of *Streptomyces verticillatus* as a component of cinnamamide biosynthesis [62]. Activation of cinnamic acid to its CoA thioester by the acyl-CoA ligase EncH followed by a series of β-oxidation reactions would directly provide the benzoyl-CoA primer unit (Figure 3, route A). Two of the three β-oxidation enzymes were identified on the sequenced fragment, a hypothetical enoyl-CoA hydratase (EncI) and a β-ketothiolase (EncJ). A dedicated β-hydroxyacyl-CoA dehyrogenase, however, is absent. This finding may imply that this protein is encoded on an unsequenced region of the cluster or on a different part of the chromosome and is functionally substituted by a resident dehydrogenase from the host *S. lividans* K4-114. Alternatively, the oxygenase EncM may have dual functions and additionally catalyze the dehydrogenation reaction. The occurrence of an aryl-CoA ligase (EncN) appears redundant, as β-oxidation of cinnamyl-CoA should directly yield benzoyl-CoA. Successful incorporation of d_5-benzoic acid into enterocin [3] suggests that EncN is indeed functional. It remains to be elucidated whether the thiolase EncJ directly generates benzoate or its CoA thioester and whether the EncN ligase accepts benzoate from an additional metabolic source. Further analysis with recombinant proteins is presently underway and will shed further light on benzoic acid biosynthesis in '*S. maritimus*' and, by analogy, in plants and fungi.

Significance

The *enc* gene cluster encodes a novel polyketide biosynthetic pathway that yields the structurally diverse set of enterocins and wailupemycins. At least 12 different structures with more than five distinct cyclization patterns result from this single biochemical pathway. The cloning and sequencing of this iterative type II biosynthetic gene set revealed several unprecedented features that account for its programmed spontaneity and novel structural characteristics. Absence of cyclases and aromatases normally associated with type II PKS clusters provides the molecular rationale for the observed reactivity as unnatural aromatic PKS systems without these gene products similarly produce randomly cyclized products. The Favorskii-like rearrangement is the probable cause for the prevention of the minimal *enc* PKS product from yielding an aromatic product and is putatively catalyzed by the oxygenase EncM. A route for the biosynthesis and attachment of the rare benzoate starter unit is proposed and involves the first prokaryotic PAL. These novel features provide a distinct opportunity to engineer structurally diverse 'unnatural' natural products for use in drug discovery by engineering hybrid expression systems with typical aromatic PKS genes, such as those from the actinorhodin, tetracenomycin and daunorubicin clusters. The engineered biosynthesis of *act* PKS products in a streptomycete transformant carrying *act* and *enc* genes on separate plasmids indicates for the first time that a multiplasmid approach with type II PKS genes is a viable and powerful technique for engineering novel products [63]. This study also represents the first example of a sequenced marine bacterial natural product biosynthetic gene cluster and its heterologous expression.

Materials and methods

Bacterial strains, plasmids, culture conditions and DNA manipulations

'*S. maritimus*' was a gift from Dr. Bradley S. Davidson (Utah State University). For sporulation, it was grown on A1 plates for 7 days at 30°C. *E. coli* XL1-Blue was used for subcloning and grown on LB plates or in LB liquid medium. *E. coli* SURE was used for cosmid library construction and was grown in LBMM liquid medium and on LB plates. pNFPK-4, pNFPK-7, and pHGF7516 were kindly provided by Dr. Tin-Wein Yu (University of Washington), and pRM5 [46] and pSEK4 by Dr. Chaitan Khosla (Stanford University). The expression plasmid pKOS010-153 [47] and the host *S. lividans* K4-114 [9] were kindly provided by Drs. Robert McDaniel and Mary C. Betlach, respectively, Kosan Biosciences, Inc., Hayward, CA, USA. pBluescript SK(−) (Stratagene) was used as vector for cloning in *E. coli* and for sequencing. SuperCos (Stratagene) and pOJ446 [64] were used for cosmid library construction. Strain and DNA manipulations as well as Southern hybridization were performed according to standard procedures [65,66].

Cosmid library construction and screening

SuperCos was digested with *Xba*I, dephosphorylated with shrimp alkaline phosphatase, and digested with *Bam*HI. '*S. maritimus*' genomic DNA was partially digested with *Sau*3AI and dephosphorylated. Genomic DNA was partially digested with *Sau*3AI and dephosphorylated. DNA with a high content of fragments between 25 and 35 kb (about 2–3 μg) was isolated by agarose gel electrophoresis and used for ligation to about 400 ng vector. Ligation was carried out overnight at 16°C using T4 DNA ligase. Packaging and titration were done using the Gigapack II XL kit (Stratagene) according to the manufacturer's instructions. After selection with carbenicillin (100 μg/ml), ~2000 colonies were picked and grown overnight in 96-well plates. Cultures were transferred to a Hybond-N membrane (Amersham Pharmacia Biotech) and grown overnight on LB agar medium containing carbenicillin. DNA was denatured and fixed according to standard procedures. For construction of the pOJ446 library, pOJ446 was digested with *Hpa*I, dephosphorylated, and digested with *Bam*HI. Further construction was as described for SuperCos, except that apramycin (100 μg/ml) was used for selection. These libraries were screened with digoxigenin-labeled *act*I-ORF1 (obtained from pNFPK-4) and *act*III (obtained from pNFPK-7) probes and ^{32}P-labeled *encP* (0.8 kb *Xma*I fragment from *encP*). Digoxigenin labeling and detection of digoxigenin-labeled positives were performed with the DIG High Prime DNA Labeling and Detection Starter kit II (Boehringer Mannheim).

Expression of enterocin-wailupemycin products

Protoplasts were prepared from *S. lividans* K4-114 and transformed with pJP15F11 according to standard procedures [66]. After protoplast regeneration became visible (12–15 h), plates were overlaid with apramycin at 1 mg/plate. Four colonies were grown on fresh plates, and cells were used for isolation of cosmid DNA. The cosmids were amplified in *E. coli*, isolated and checked by restriction analysis. *S. lividans* K4-114/pJP15F11-1 was selected for expression studies and grown on R2YE agar plates containing 100 μg/ml apramycin for 1 week. Chemical analysis of the expressed products was previously described [3].

DNA sequencing

Sequencing was carried out by the dideoxynucleotide chain termination method with BigDye Terminator Ready Mix (Applied Biosystems) and using an ABI 377X2 sequencer (Applied Biosystems). Cosmid pSS9A6 and pBluescript SK(−) containing inserts of pSS9A6 fragments generated by digestion with *Bam*HI, *Sac*I or *Xma*I were used for sequencing. Both DNA strands were sequenced using the T3 and T7 universal primers or with internal primers. Computer-aided database searching and sequence analyses were performed using the BLAST algorithm [67], MacFRAME (version 1.2, developed by Kevin Kendall (Tulane University, New Orleans, LA, USA)), and ClustalX.

Construction of pCH20

pCH20 is a derivative of pKOS010-153 [47], in which the *Pac*I–*Eco*RI *ery*AIII fragment was replaced by a 0.9 kb *Pac*I–*Eco*RI *enc*D fragment. *Pac*I and *Eco*RI were introduced by cloning the 0.9 kb *Fse*I(blunt)–*Nsi*I fragment of pSS9A6 into the *Eco*RV–*Nsi*I sites of pHGF7516, a derivative of pNEB193 (New England Biolabs) carrying a modified polylinker.

Streptomyces transformation, culture conditions, and MS analysis of polyketides

S. lividans K4-114 transformants were prepared according to standard procedures with thiostrepton and apramycin in R2YE protoplast regeneration plates [66]. Protoplasts were transformed singly with pRM5 and pSEK4 and doubly with pSEK4 and pCH20. The strains were grown on R2YE agar plates overlaid with thiostrepton (0.5 mg/plate) and/or apramycin (1 mg/plate) at 30°C for 1 week. The cultured agar from one plate (ca. 25 ml) was chopped and extracted with 95:5 EtOAc/MeOH (50 ml) at room temperature for 1–2 h. The organic solution was dried (Na_2SO_4), filtered, solvent-evaporated, dissolved in MeOH, and analyzed. A Hewlett Packard 1050 series high performance liquid chromatography (HPLC) system was linked to a Finnigan MAT TSQ 7000 mass spectrometer, using atmospheric pressure chemical ionization and operating in the positive ion mode. A Vydac 218TP52 4.6×200 mm C18 column was used at a flow rate of 0.2 ml/min with a linear solvent gradient of 0.15% trifluoroacetic acid (TFA) in water to 0.085% TFA in acetonitrile over a period of 40 min.

Ancymidol inhibition of enterocin biosynthesis

Five flasks containing 100 ml A1 medium were inoculated with 10 ml of fresh overnight '*S. maritimus*' culture. 100 μl of DMSO was added at the time of inoculation to one flask as a control, three of the remaining flasks had ancymidol added to 0.2, 0.8, and 2 mM final concentration, and the final flask received no additions. These cultures were incubated for 7 days at 28°C, 200 rpm. EtOAc fractions (100 ml) were dried ($MgSO_4$), evaporated in vacuo, dissolved in MeOH (500 μl), and analyzed (5 μl) by HPLC (4.6×150 mm YMC ODS-AQ S3 120A column; flow rate of 0.5 ml/min with a linear solvent gradient of 40% MeOH in 0.15% TFA–water to 100% MeOH over 30 min).

'*S. maritimus*' cell-free conversion of 5-deoxyenterocin to enterocin

Cells from a 5 day culture of '*S. maritimus*' in 1 l A1 media were harvested by centrifugation and resuspended in 100 mM phosphate buffer, pH 7.2, containing 20% v/v glycerol, 0.2 mM PMSF, 1 mM EDTA, and 0.2 mM dithiothreitol. The cells were lysed by a single passage through a French pressure cell, and polyethylene imine was added to a final concentration of 0.2%. The resulting suspension was centrifuged at 35 000×*g* for 35 min. The supernatant was removed and adjusted to 75% $(NH_4)_2SO_4$ saturation. The protein precipitate was resuspended in 3 ml 100 mM phosphate buffer, pH 7.2, containing 20% glycerol, desalted by gel permeation chromatography, and assayed.

100 μl of enzyme solution (0.1–0.3 mg total protein) was added to a 0.9 ml solution containing 0.01 μmol spinach ferredoxin, 0.17 μmol NADPH, 10 nmol 5-deoxyenterocin, 0.1 U ferredoxin-$NADP^+$ reductase, and 20 mM Tris pH 7.4. The reaction was incubated for 2 h at 30°C and then extracted with EtOAc (3×0.5 ml). The organic fractions were combined, dried (Na_2SO_4), and the solvent removed in vacuo. The resulting pellet was resuspended in MeOH (20 μl) and analyzed (5 μl) as above by HPLC.

Overproduction and biochemical characterization of EncR

EncR was amplified by PCR from the cosmid clone pJP15F11 with primers P450pstf (5′-GACTGCAGACCACCCATACCCAG-3′) and P450hinr2 (5′-CTAAGCTTTCACCACTTCACCGGCAG-3′). (The introduced *Pst*I and *Hin*dIII restriction sites are underlined, respectively.) The amplified product was digested and cloned into the *Pst*I–*Hin*dIII sites of pMAL-c2x (NEB) to yield pPS006. Introduction of pPS006 into *E. coli* TB1 under standard expression conditions resulted in the production of MBP-EncR. The soluble fraction of the fusion protein was partially purified on an amylose resin affinity column (NEB). Cytochrome P-450 assays were performed as above. Ancymidol inhibition was measured at 0.2, 0.8, and 2 mM concentrations.

Accession numbers

The GenBank accession number for the *enc* cluster is AF254925.

Acknowledgements

We thank Dr. S. Storkova (University of Washington) for constructing the '*S. maritimus*' SuperCos library and Drs. R. McDaniel and G. Ashley (Kosan Biosciences) for helpful discussions. This work was supported in part by the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, through the National and Washington Sea Grant programs (NA76RG01149, project no. R/B-28); and the National

Institutes of Health (AI47818). J.P. and C.H. were recipients of Feodor Lynen postdoctoral fellowships from the Alexander von Humboldt Foundation, and M.S.N. was supported in part by an undergraduate research award from the NSF (REU Site Grant CHE-9820032).

References

1. Hopwood, D.A. (1997). Genetic contributions to understanding polyketide synthases. *Chem. Rev.* **97**, 2465–2497.
2. Rawlings, B.J. (1999). Biosynthesis of polyketides (other than actinomycete macrolides). *Nat. Prod. Rep.* **16**, 425–484.
3. Piel, J., Hoang, K. & Moore, B.S. (2000). Metabolic diversity encoded by the enterocin biosynthesis gene cluster. *J. Am. Chem. Soc.* **122**, 5415–5416.
4. Sitachitta, N., Gadepalli, M. & Davidson, B.S. (1996). New α-pyrone-containing metabolites from a marine-derived Actinomycete. *Tetrahedron* **52**, 8073–8080.
5. Miyairi, N., Sakai, H.-I., Konomi, T. & Imanaka, H. (1976). Enterocin, a new antibiotic. Taxonomy, isolation and characterization. *J. Antibiot.* **29**, 227–235.
6. Tokuma, Y., Miyairi, N. & Morimoto, Y. (1976). Structure of enterocin: X-ray analysis of *m*-bromobenzoyl enterocin dihydrate. *J. Antibiot.* **29**, 1114–1116.
7. Seto, H., Sato, T., Urano, S., Uzawa, J. & Yonehara, H. (1976). Utilization of ^{13}C–^{13}C coupling in structural and biosynthetic studies. VII. The structure and biosynthesis of vulgamycin. *Tetrahedron Lett.* **17**, 4367–4370.
8. Kang, H., Jensen, P.R. & Fenical, W. (1996). Isolation of microbial metabolites from a marine ascidian of the genus *Didemnum*. *J. Org. Chem.* **61**, 1543–1546.
9. Ziermann, R. & Betlach, M.C. (1999). Recombinant polyketide synthesis in *Streptomyces*: Engineering of improved host strains. *BioTechniques* **26**, 106–110.
10. Fernández-Moreno, M.A., Martinez, E., Boto, L., Hopwood, D.A. & Malpartida, F. (1992). Nucleotide sequence and deduced functions of a set of cotranscribed genes of *Streptomyces coelicolor* A3 including the polyketide synthase for the antibiotic actinorhodin. *J. Biol. Chem.* **267** (2), 19278–19290.
11. Bibb, M.J., Findlay, P.R. & Johnson, M.W. (1984). The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. *Gene* **30**, 157–166.
12. Strohl, W.R. (1992). Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. *Nucleic Acids Res.* **20**, 961–974.
13. Hertweck, C. & Moore, B.S. (2000). A plant-like biosynthesis of benzoyl-CoA in the marine bacterium '*Streptomyces maritimus*'. *Tetrahedron* **56**, 9115–9120.
14. Schwede, T.F., Rétey, J. & Schulz, G.E. (1999). Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile. *Biochemistry* **27**, 5355–5361.
15. Langer, B., Rother, D. & Rétey, J. (1997). Identification of essential amino acids in phenylalanine ammonia-lyase by site-directed mutagenesis. *Biochemistry* **36**, 10867–10871.
16. Ferrandez, A., Prieto, M.A., Garcia, J.L. & Diaz, E. (1997). Molecular characterization of PadA, a phenylacetaldehyde dehydrogenase from *Escherichia coli*. *FEBS Lett.* **406**, 23–27.
17. Egland, P.G., Gibson, J. & Harwood, C.S. (1995). Benzoate-coenzyme A ligase, encoded by *bad*A, is one of three ligases able to catalyze benzoyl-coenzyme A formation during anaerobic growth of *Rhodopseudomonas palustrius* on benzoate. *J. Bacteriol.* **177**, 6545–6551.
18. Gibson, J., Dispensa, M., Fogg, G.C., Evans, D.T. & Harwood, C. (1994). 4-Hydroxybenzoate-coenzyme A ligase from *Rhodopseudomonas palustris*: purification, gene sequence, and role in anaerobic degradation. *J. Bacteriol.* **176**, 634–641.
19. Nakatsu, C.H., Straus, N.A. & Wyndham, R.C. (1995). The nucleotide sequence of the Tn5271 3-chlorobenzoate 3,4-dioxygenase genes (*cba*AB) unites the class IA oxygenases in a single lineage. *Microbiology* **141**, 485–495.
20. August, P.R., Tang, L., Yoon, Y.J., Ning, S., Müller, R., Yu, T.-W., Taylor, M., Hoffman, 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.
21. Grimm, A., Madduri, K., Ali, A. & Hutchinson, C.R. (1994). Characterization of the *Streptomyces peucetius* ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene* **151**, 1–10.
22. Bibb, M.J., Sherman, D.H., Omura, S. & Hopwood, D.A. (1994). Cloning, sequencing and deduced funcitons of a cluster of *Streptomyces* genes probably encoding biosynthesis of the polyketide antibiotic frenolicin. *Gene* **142**, 31–39.
23. Mohler, H., Bruhmuller, M. & Decker, K. (1972). Covalently bound flavin in D-6-hydroxynicotine oxidase from *Arthrobacter oxidans*. Identification of the 8-(*N*-3-histidyl)-riboflavin-linkage between FAD and apoenzyme. *Eur. J. Biochem.* **18**, 152–155.
24. Johnson, D.A., August, P.R., Shackleton, C., Liu, H.-W. & Sherman, D.H. (1997). Microbial resistance to mitomycins involves a redox relay mechanism. *J. Am. Chem. Soc.* **119**, 2576–2577.
25. Mao, Y.Q., Varoglu, M. & Sherman, D.H. (1999). Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from *Streptomyces lavendulae* NRRL2564. *Chem. Biol.* **6**, 251–263.
26. Dittrich, H. & Kutchan, T.M. (1991). Molecular cloning, expression, and induction of berberine bridge enzyme, an enzyme essential to the formation of benzophenanthridine alkaloids in the response of plants to pathogenic attack. *Proc. Natl. Acad. Sci. USA* **15**, 9969–9973.
27. Poulos, T.L., Finzel, B.C. & Howard, A.J. (1987). High-resolution crystal structure of P450cam. *J. Mol. Biol.* **195**, 687–700.
28. Schwecke, T., Aparicio, J.F., Molnar, I., Koenig, A., Khaw, L.E., Haydock, S.F., Oliynyk, M., Caffrey, P., Cortes, J., Lester, J.B., Boehm, G.A., Staunton, J. & Leadlay, P.F. (1995). The biosynthesis gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. USA* **92**, 7839–7843.
29. Helmann, J.D., Ballard, B.T. & Walsh, C.T. (1990). The MerR metalloregulatory protein binds mercuric ion as a tricoordinate, metal-bridged dimer. *Science* **23**, 946–948.
30. Horinouchi, S. & Beppu, T. (1992). Regulation of secondary metabolism and cell differentiation in *Streptomyces*: A-factor as a microbial hormone and the AfsR protein as a component of a two-component regulatory system. *Gene* **15**, 167–172.
31. Fernández-Moreno, M.A., Caballero, J.L., Hopwood, D.A. & Malpartida, F. (1991). The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces*. *Cell* **66**, 769–780.
32. Stutzman-Engwall, K.J., Otten, S.L. & Hutchinson, C.R. (1992). Regulation of secondary metabolism in *Streptomyces* spp. and overproduction of daunorubicin in *Streptomyces peucetius*. *J. Bacteriol.* **174**, 144–154.
33. Coque, J.J., Liras, P. & Martin, J.F. (1993). Genes for a β-lactamase, a penicillin-binding protein and a transmembrane protein are clustered with the cephamycin biosynthetic genes in *Nocardia lactamdurans*. *EMBO J.* **12**, 631–639.
34. Guilfoile, P.G. & Hutchinson, C.R. (1992). Sequence and transcriptional analysis of the *Streptomyces glaucescens tcmAR* tetracenomycin C resistance and repressor gene loci. *J. Bacteriol.* **174**, 3651–3658.
35. Hillen, W. & Berens, C. (1994). Mechanisms underlying expression of Tn10 encoded tratracycline resistance. *Annu. Rev. Microbiol.* **48**, 345–369.
36. Wright, F. & Bibb, M.J. (1992). Codon usage in the G+C-rich *Streptomyces* genome. *Gene* **113**, 55–65.
37. Leskiw, B.K., Bibb, M.J. & Chater, K.F. (1991). The use of a rare codon specifically during development. *Mol. Microbiol.* **5**, 2861–2867.
38. Lombo, F., Blanco, G., Fernandez, E., Mendez, C. & Salas, J.A. (1996). Characterization of *Streptomyces argillaceus* genes encoding a polyketide synthase involved in the biosynthesis of the antitumor mithramycin. *Gene* **172**, 87–91.
39. Lomovskaya, N., Otten, S.L., Doi-Katayama, Y., Fonstein, L., Liu, X.C., Takatsu, T., Inventi-Solari, A., Filippini, S., Torti, F., Colombo, A.L. & Hutchinson, C.R. (1999). Doxorubicin overproduction in *Streptomyces peucetius*: Cloning and characterization of the *dnrU* ketoreductase and *dnrV* genes and the *doxA* cytochrome P-450 hydroxylase gene. *J. Bacteriol.* **181**, 305–318.
40. Taguchi, T., Itou, K., Ebizuka, Y., Malpartida, F., Hopwood, D.A., Surti, C.M., Booker-Milburn, K.I., Stephenson, G.R. & Ichinose, K. (2000). Chemical characterisation of disruptants of the *Streptomyces coelicolor* A3 *act*VI genes involved in actinorhodin biosynthesis. *J. Antibiot.* **53** (2), 144–152.
41. Ichinose, K., Bedford, D.J., Tornus, D., Bechthold, A., Bibb, M.J.,

Revill, W.P., Floss, H.G. & Hopwood, D.A. (1998). The granaticin biosynthetic gene cluster of *Streptomyces violaceoruber* Tü22: sequence analysis and expression in a heterologous host. *Chem. Biol.* **5**, 647–659.
42. Rohr, J., Mendez, C. & Salas, J.A. (1999). The biosynthesis of aureolic group antibiotics. *Biorg. Chem.* **27**, 41–54.
43. Gerlitz, M., Meurer, G., Wendt-Pienkowski, E., Madduri, K. & Hutchinson, C.R. (1997). Effect of the daunorubicin *dpsH* gene on the choice of starter unit and cyclization pattern reveals that type II polyketide synthases can be unfaithful yet intriguing. *J. Am. Chem. Soc.* **119**, 7392–7393.
44. Taguchi, T., Ebizuka, Y., Hopwood, D.A. & Ichinose, K. (2000). Identification of a novel shunt product produced by a disruptant of the *act*VI-ORFA gene involved in the biosynthesis of actinorhodin in *Streptomyces coelicolor* A3. *Tetrahedron Lett.* **41** (2), 5253–5256.
45. Fu, H., Ebert-Khosla, S., Hopwood, D.A. & Khosla, C. (1994). Engineered biosynthesis of novel polyketides: Dissection of the catalytic specificity of the *act* ketoreductase. *J. Am. Chem. Soc.* **116**, 4166–4170.
46. McDaniel, R., Ebert-Khosla, S., Hopwood, D.A. & Khosla, C. (1993). Engineered biosynthesis of novel polyketides. *Science* **262**, 1546–1550.
47. Xue, Q., Ashley, G., Hutchinson, C.R. & Santi, D.V. (1999). A multiplasmid approach to preparing large libraries of polyketides. *Proc. Natl. Acad. Sci. USA* **96**, 11740–11745.
48. McDaniel, R., Ebert-Khosla, S., Fu, H., Hopwood, D.A. & Khosla, C. (1994). Engineered biosynthesis of novel polyketides: Influence of a downstream enzyme on the catalytic specificity of a minimal aromatic polyketide synthase. *Proc. Natl. Acad. Sci. USA* **91**, 11542–11546.
49. Oikawam, H., Ichihara, A. & Sakamura, S. (1988). Biosynthetic study of betaenone B. Origin of the oxygen atoms and accumulation of a deoxygenated intermediate using P-450 inhibitor. *J. Chem. Soc. Chem. Commun.* 600–602.
50. Shen, Y., Yoon, P., Yu, T.-W., Floss, H.G., Hopwood, D.A. & Moore, B.S. (1999). Ectopic expression of the minimal *whiE* polyketide synthase generates a library of aromatic polyketides of diverse sizes and shapes. *Proc. Natl. Acad. Sci. USA* **96**, 3622–3627.
51. Moore, B.S. & Piel, J. Engineering biodiversity with type II polyketide synthase genes. *Antonie van Leeuwenhoek J. Microbiol.*, in press.
52. Simpson, T.J. & Holker, J.S.E. (1975). Biosynthesis of a pyrone metabolite of *Aspergillus melleus*. Application of long-range ^{13}C–^{13}C coupling constants. *Tetrahedron Lett.* **16**, 4693–4696.
53. Kakinuma, K., Uzawa, J. & Uramoto, M. (1982). Biosynthesis of the 17-membered carbocyclic ring of lankacidin antibiotics. *Tetrahedron Lett.* **23**, 5303–5306.
54. Wright, J.L.C., Hu, T., McLachlan, J.L., Needham, J. & Walter, J.A. (1996). Biosynthesis of DTX-4: confirmation of a polyketide pathway, proof of a Baeyer-Villiger oxidation step, and evidence for an unusual carbon deletion process. *J. Am. Chem. Soc.* **118**, 8757–8758.
55. Alvarez, M.A., Fu, H., Khosla, C., Hopwood, D.A. & Bailey, J.E. (1996). Engineered biosynthesis of novel polyketides: Properties of the *whiE* aromatase/cyclase. *Nat. Biotechnol.* **14**, 335–338.
56. Fu, H., Ebert-Khosla, S., Hopwood, D.A. & Khosla, C. (1994). Relaxed specificity of the oxytetracycline polyketide synthase for an acetate primer in the absence of a malonamyl primer. *J. Am. Chem. Soc.* **116**, 6443–6444.
57. Marti, T., Hu, Z., Pohl, N.L., Shah, A.N. & Khosla, C. (2000). Cloning, nucleotide sequence, and heterologous expression of the biosynthetic gene cluster for R1128, a non-steroidal estrogen receptor antagonist. Insights into an unusual priming mechanism. *J. Biol. Chem.* **275**, 33443–33448.
58. Bao, W., Sheldon, P.J., Wendt-Pienkowski, E. & Hutchinson, C.R. (1999). The *Streptomyces peucetius dpsC* gene determines the choice of starter unit in biosynthesis of the daunorubicin polyketide. *J. Bacteriol.* **181**, 4690–4695.
59. Bao, W., Sheldon, P.J. & Hutchinson, C.R. (1999). Purification and properties of the *Streptomyces peucetius dpsC* β-ketoacyl:acyl carrier protein synthase III that specifies the propionate-starter unit for type II polyketide biosynthesis. *Biochemistry* **38**, 9752–9757.
60. Schupp, T., Toupet, C., Cluzel, B., Neff, S., Hill, S., Beck, J.J. & Ligon, J.M. (1995). A *Soranium cellulosum* (myxobacterium) gene cluster for the biosynthesis of the macrolide antibiotic soraphen A. Cloning, characterization and homology to the polyketide synthase genes from actinomycetes. *J. Bacteriol.* **177**, 3673–3679.
61. Schneider, S., Mohamed, M.E. & Fuchs, G. (1997). Anaerobic metabolism of L-phenylalanine via benzoyl-CoA in the denitrifying bacterium *Thauera aromatica*. *Arch. Microbiol.* **168**, 310–320.
62. Bezanson, G.S., Desaty, D., Emes, A.V. & Vining, L.C. (1970). Biosynthesis of cinnamide and detection of phenylalanine ammonia-lyase in *Streptomyces verticillatus*. *Can. J. Microbiol.* **16**, 147–151.
63. Hertweck, C. (2000). The multiplasmid approach: a new perspective for combinatorial biosynthesis. *ChemBioChem* **1**, 103–106.
64. Bierman, J., Logan, R., O'Brien, K., Seno, E.T., Rao, R.N. & Schoner, B.E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **116**, 43–49.
65. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning, a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
66. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. & Hopwood, D.A. (2000). *Practical Streptomyces Genetics*. The John Innes Foundation, Norwich.
67. 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.
68. Bruenker, P., McKinney, K., Sterner, O., Minas, W. & Bailey, J.E. (1999). Isolation and characterization of the naphthocyclinone gene cluster from *Streptomyces arenae* DSM40737 and heterologous expression of the polyketide synthase genes. *Gene* **227**, 125–135.
69. Tsukamoto, N., Fujii, I., Ebizuka, Y. & Sankawa, U. (1994). Nucleotide sequence of the *aknA* region of the aklavinone biosynthetic gene cluster of *Streptomyces galilaeus*. *J. Bacteriol.* **176**, 2473–2475.
70. Dobert, R.C., Breil, B.T. & Triplett, E.W. (1994). DNA sequence of the common nodulation genes of *Bradyrhizobium elkanii* and their phylogenetic relationship to those of other nodulating bacteria. *Mol. Plant. Microbe Interact.* **7**, 564–572.
71. Eichler, K., Bourgis, F., Buchet, A., Kleber, H.P. & Mandrand-Berthelot, M.A. (1994). Molecular characterization of the *cai* operon necessary for carnitine metabolism in *Escherichia coli*. *Mol. Microbiol.* **13**, 775–786.
72. Tosato, V., Albertini, A.M., Zotti, M., Sonda, S. & Bruschi, C.V. (1997). Sequence completion, identification and definition of the fengycin operon in *Bacillus subtilis* 168. *Microbiology* **143**, 3443–3450.
73. Oda, M., Sugishita, A. & Furukawa, K. (1988). Cloning and nucleotide sequence of histidase and regulatory genes in the *Bacillus subtilis hut* operon and positive regulation of the operon. *J. Bacteriol.* **170**, 3199–3205.
74. Howles, P.A., Arioli, T. & Weinman, J.J. (1994). Characterization of a phenylalanine ammonia-lyase multigene family in *Trifolium subterraneum*. *Gene* **138**, 87–92.
75. Lamark, T., Kaasen, I., Eshoo, M.W., Falkenberg, P., McDougall, J. & Strom, A.R. (1991). DNA sequence and analysis of the *bet* genes encoding the osmoregulatory choline-glycine betaine pathway of *Escherichia coli*. *Mol. Microbiol.* **5**, 1049–1064.