Chemistry & Biology, Vol. 10, 1225–1232, December, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2003.11.009

Iteration as Programmed Event during Polyketide Assembly; Molecular Analysis of the Aureothin Biosynthesis Gene Cluster

Jing He and Christian Hertweck*
Hans-Knoell-Institute for Natural Products Research
Department of Bioorganic Synthesis
Beutenbergstr. 11a
D-07745 Jena
Germany

Summary

Analysis of the type I modular polyketide synthase (PKS) involved in the biosynthesis of the rare nitroaryl polyketide metabolite aureothin (*aur*) from *Streptomyces thioluteus* HKI-227 has revealed only four modules to catalyze the five polyketide chain extensions required. By heterologous expression of the *aur* PKS cluster, direct evidence was obtained that these modules were sufficient to support aureothin biosynthesis. It appears that one module catalyzes two successive cycles of chain extension, one of the first examples of a PKS in which such iteration or "stuttering" is required to produce the normal polyketide product. In addition, lack of a specified loading domain implicates a novel PKS priming mechanism involving the unique p-nitrobenzoate starter unit. The 27 kb *aur* gene cluster also encodes a novel *N*-oxidase, which may represent the first member of a new family of such enzymes.

Introduction

Polyketides constitute a large and diverse group of natural products with an impressive wealth of antibiotic, anticancer, antiparasitic, and immunosuppressive activities [1]. They are assembled by polyketide synthases (PKS) in a manner that closely parallels fatty acid biosynthesis: an activated carboxylic acid starter undergoes repetitive Claisen condensations with malonyl-coenzyme A (CoA)-derived extender units. In addition to manifold tailoring reactions, polyketide metabolic diversity results from a number of programmed events involving selection of starter and extender units, carbon chain length, the degree of reduction, as well as chain termination and folding. According to their mode of operation, PKS can be generally classified into iterative and noniterative PKS. Some polyketide synthases, such as bacterial aromatic (type II) PKS, fungal, and plant PKS, operate in an iterative fashion reminiscent of fatty acid synthases. In these systems, catalytic domains are reused until the polyketide chain has reached its full length. In stark contrast, bacterial type I PKS [2, 3] represent giant multienzyme systems [3–6] that accommodate distinct active sites for each step of catalysis and are thus considered as noniterative. In bacterial type I PKS, catalytic domains are physically grouped into modules that consist of a minimal set of domains for chain propagation, usually a ketosynthase (KS), an acyl transferase (AT), and an acyl carrier protein (ACP). Additional domains may be present for keto group processing, such as ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) [2, 3]. While the first module is usually fused to a loading domain at its N terminus, the last PKS module terminates polyketide assembly with an additional thioesterase (TE) domain. According to the molecular assembly line character of modular PKS, order and architecture of the modules usually reflect the chain length as well as degree of reduction of the resultant polyketide [2, 3]. This one-to-one correspondence has been demonstrated experimentally [7] and provides the basis for rationally engineering type I PKS biosynthetic pathways. By a variety of targeted genetic mutations, such as inactivation, substitution, and addition of catalytic domains or even entire modules, polyketide structures have been successfully modified. In nature, evolution of polyketide metabolic diversity may be rationalized by analogous random mutations [8]. During the last decade, a huge body of knowledge about polyketide synthases has been established by molecular and biochemical approaches. However, one of the unresolved problems of polyketide biosynthesis is the molecular basis for iterative or noniterative action of polyketide synthases and their evolutionary relationship.

In a structure-guided search for novel biosynthetic activities, we were prompted by the unusual structures of the nitrophenyl-substituted polyketides aureothin and neoaureothin (Figure 1). Aureothin is a metabolite of the soil isolate *Streptomyces thioluteus* [9] with antitumor, antifungal, and insecticidal properties [10]. The homolog neoaureothin, produced by *S. orinoci*, exhibits antifungal and antiviral activities [11]. By molecular analysis of the requisite gene clusters, we expected new insights into the machinery and evolution of polyketide synthases and to provide the basis for engineering novel aureothin derivatives. In this work, we report on the cloning, sequencing, and heterologous expression of the aureothin biosynthesis gene cluster, which has revealed an aberrant, nonlinear type I polyketide synthase.

Results

Identification, Cloning, and Heterologous Expression of the Aureothin Biosynthesis Gene Cluster

The structure of aureothin and early feeding experiments [12] suggested a mixed biosynthetic pathway involving a type I polyketide synthase (Figure 2). A *S. thioluteus* HKI-227 cosmid library was constructed and screened for type I PKS genes. Cosmid mapping, PCR, and hybridization experiments revealed the presence of several type I PKS gene clusters in the *S. thioluteus* HKI-227 genome. In order to identify the correct cluster, a complementary screening was required. Recently, we have elucidated the nature and biosynthetic origin of the rare nitro aryl moiety of aureothin. We have shown that a

*Correspondence: christian.hertweck@hki-jena.de

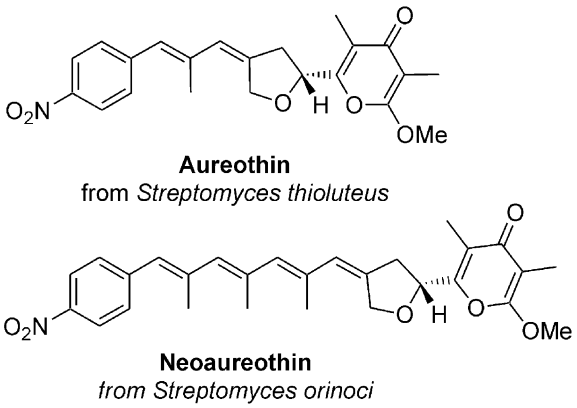

Figure 1. Rare Nitrophenyl-Substituted Polyketide Metabolites of Varying Chain Length Produced by Different Streptomyces Strains

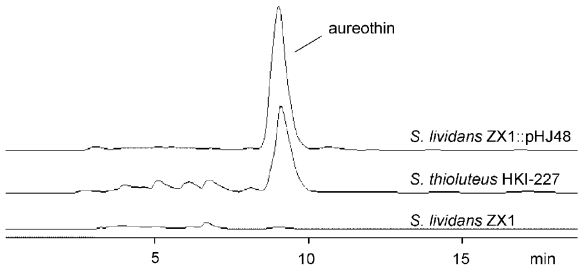

Figure 3. HPLC Analysis of Extracts from Wild-Type *S. thioluteus* HKI-227, Host *S. lividans* ZX1, and Transformant *S. lividans* ZX1::pHJ48

p-aminobenzoate (PABA) synthase is involved in the formation of a nitrobenzoyl PKS starter unit (Figure 2) (J.H. and C.H., unpublished data). By PCR using degenerate primers based on universally conserved PABA synthase motifs, we were able to amplify a 1 kb fragment, which proved to be a homolog of *pabAB* from *S. griseus* [13]. This gene probe served to identify the putative aureothin biosynthesis gene cluster out of the various *S. thioluteus* type I PKS gene clusters detected by Southern hybridization.

In order to verify the identity of the aureothin gene cluster by inactivation, various attempts to introduce foreign DNA into *S. thioluteus* HKI-227 have been undertaken, which all failed. Therefore, heterologous expression of the putative gene cluster was attempted. A ca. 40 kb SspI fragment of cosmid pST18E4, which harbors PKS genes as well as the PABA synthase gene, which included the pWEB *cos* site, was ligated into the EcoRV site of the integrative *E. coli-Streptomyces* shuttle vector pSET152 [14]. The resulting cosmid pHJ11 was packaged as phage and transfected into *E. coli* EPI 100. Restriction mapping proved that this construct was stable. By protoplast transformation, pHJ11 was introduced into *S. lividans* ZX1 [15], and the resulting transformant (*S. lividans* ZX1:pHJ11) was cultivated for 5 days. After extraction, HPLC and ESI-MS analyses unequivocally showed that aureothin is produced by the transformant at a titer comparable with the wild-type.

Since the cloned genes were sufficient to confer on *S. lividans* ZX1 the ability to make aureothin, the cosmid insert of pST18E4 was subjected to both directed and shotgun subcloning and sequencing.

Architecture of the Genomic Region Involved in Aureothin Biosynthesis

The ca. 40 kb contiguous DNA insert of pST18E4 was fully sequenced on both strands. The nucleotide sequence was analyzed for open reading frames (ORFs) by the FRAME program [16]. The boundaries of the aureothin (*aur*) gene cluster were determined by subcloning a 30 kb EcoRV fragment of pST18E4 into pSET152, yielding the integrative expression plasmid pHJ48. This surprisingly small fragment proved to be sufficient for aureothin biosynthesis by heterologous expression in *S. lividans* ZX1 (Figure 3). The deduced gene organization of the gene cluster is graphically presented in Figure 4, and the results of the sequence analyses are summarized in Table 1.

Putative functions of deduced gene products were assigned by sequence comparisons with database proteins (BLAST and PROSITE searches). Nine genes involved in aureothin biosynthesis were identified and designated *aurA* through *aurI*. In the *aur* operon, all genes are transcribed in one direction except for *aurE*, which is convergently transcribed. Aureothin biosynthesis is probably regulated by the predicted 271 aa gene product of *aurD*, which has 34% identity with AfsR, a global regulatory protein for secondary metabolite formation in *Streptomyces coelicolor* A3(2) [17]. Downstream of the *aur* biosynthetic gene cluster, a gene encoding a transposase was located. Interestingly, no candidate genes for resistance have been detected.

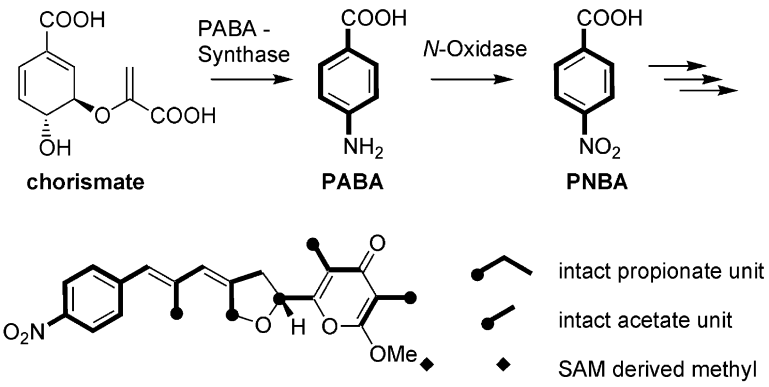

Figure 2. Biosynthetic Origin of Aureothin as Determined by Labeling Experiments

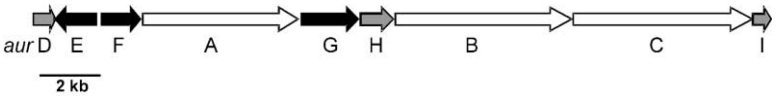

Figure 4. Organization of the Aureothin Biosynthetic Gene Cluster in *S. thioluteus* HKI-227

Each arrow indicates the direction of transcription and relative sizes of the ORFs deduced from analysis of the nucleotide sequence. White ORFs, type I PKS genes; black ORFs, genes putatively involved in starter unit biosynthesis; checked ORFs, genes encoding tailoring enzymes; hatched ORF, regulator gene.

Genes Putatively Involved in Starter Unit Synthesis and Post-PKS Processing

Besides three large ORFs encoding a type I PKS (*aurA–C*, see below), genes putatively involved in starter unit biosynthesis and post-PKS reactions were detected. By PCR and hybridization experiments, we have shown that a homolog of *pab*AB, a gene encoding a PABA synthase in *S. griseus* [13], is located on the aureothin gene cluster. The deduced 695 amino acid gene product of *aurG* has high homology to a group of bifunctional PABA synthases, which are typical for eukaryotes [18] but have also been implicated in the secondary metabolism of a few *Streptomyces* species [13, 19, 20]. In general, PABA synthases are known to catalyze the conversion of chorismic acid to p-aminobenzoic acid by transamination and subsequent hydrolysis [21]. In contrast to the monofunctional enzymes PabA and PabB involved in bacterial folate biosynthesis, the putative *aur* PABA synthase possesses two linked catalytic domains, which are functionally equivalent to PabA and PabB [13]. Homologs of AurG have been found to be involved in the biosynthesis of the polyene macrolides candicidin [13, 22] and FR-008 [23].

ORF *aurE* presumably encodes an acyl CoA ligase. The deduced 505 aa gene product contains an AMP binding site and has significant sequence homology to a number of acyl-CoA ligases, such as feruloyl-CoA synthetase from *Amycolatopsis sp*. HR167 (35% identity over 470 amino acids, GenBank accession number AX172316). In analogy to a variety of other unusual starter units [24], the *aur* PKS primer, p-nitrobenzoate (PNBA), is probably activated as the corresponding CoA thioester by AurE.

In the deduced gene product of *aurI*, an ORF located downstream of the PKS genes, a probable S-adenosyl methionine binding motif was identified (PRO-SITE: PS50193) that is characteristic for methyltransferases. Strikingly, AurI shows highest homology (36% identity over 209 amino acids) to EncK from "*Streptomyces maritimus*," which functions as a pyrone O-methyl transferase in enterocin biosynthesis [25, 26]. In analogy, AurI is assumed to be responsible for the introduction of the pyrone O-methyl group.

A putative cytochrome P-450 oxygenase encoding gene is located downstream of the PABA synthase gene *aurD* on the aureothin gene cluster, designated *aurH*. The proposed 406 aa gene product has highest homology to a hydroxylase from *Mycobacterium smegmatis*, which is required for piperidine and pyrrolidine utilization [27]. Homologous P-450 oxidases involved in polyketide tailoring reactions are, for example, the 6-deoxyerythronolide B hydroxylase EryF from *Saccharopolyspora erythraea* [28], and AveE from *S. avermitilis*, which has been shown to play a role in avermectin biosynthesis [29]. This strongly suggests that AurH is implicated in furan ring formation.

The gene (*aurF*) encoding an *N*-oxidizing enzyme has been traced by means of a PABA-PNBA biotransformation assay expressing fragments of the *aur* gene cluster. Strikingly, no homology between the deduced amino acid sequence of the remarkably arginine-rich 336 aa gene product of *aurF* and the known active sites (PROSITE) has been found. Therefore, AurF represents an unprecedented type of (*N*-) oxidase. Its function has been confirmed by heterologous expression in *S. lividans* ZX1 (J.H. and C.H., unpublished data).

In summary, all deduced gene products fit well into the biosynthetic scheme, and all gene functions except for *aurF* could be assigned by homology. Conversely, analysis of the *aur* PKS genes gave unexpected results.

Table 1. Deduced Functions of ORFs in the Aureothin Biosynthetic Gene Cluster

Protein	Amino Acids	Proposed Function	Sequence Similarity (Protein, Origin)	Identity/ Similarity, %	Protein Accession Number	Reference
AurD	271	transcriptional activator	AfsR, *Streptomyces coelicolor*	34/45	BAA14186	[17]
AurE	505	Acyl-CoA ligase	Feruloyl-CoA synthetase, *Amycolatopsis sp*. HR167	34/50	AX172316	
AurF	481	amine oxidase	none	none		
AurA Module 1	1906	polyketide synthase KS AT DH KR ACP	AVES 2, *Streptomyces avermitilis*	45/56	BAA84475	[29]
AurG	695	PABA synthase	PabAB, *Streptomyces griseus*	44/53	CAC22117	[22]
AurH	406	Cyt P-450 hydroxylase	Cytochrome P-450, *Mycobacterium smegmatis*	36/51	AF102510	[27]
AurB Module 2	2164	polyketide synthase KS AT ER DH KR ACP	NysC, *Streptomyces noursei*	48/59	AAF71776	[49]
AurC Module 3 Module 4	2201	polyketide synthase KS AT ACP KS AT ACP TE	AVES 4, *Streptomyces avermitilis*	45/60	BAA84479	[29]
AurI	230	O-Methyltransferase	EncK, *Streptomyces maritimus*	36/54	AAF81726	[25]

Table 2. Analysis of the Active Site and Substrate Specificity Motifs of the *aur* PKS AT1-4 Domains

AT	Active Site Motif	Specificity Motif	Substrate
AT1	PAAVV**GHS**QGEIAA	RVDVV	Methylmalonyl-CoA
AT2	PDLLL**G HSI**GELTA	QTGYT	Malonyl-CoA
AT3	PGAVV**GHS**QGEIAA	RVDVV	Methylmalonyl-CoA
AT4	PAFVH**G**RADGEVAA	none	inactive?

Analysis of the Aureothin PKS Genes

In the 27 kb region encoding *aur* biosynthesis, three large ORFs, *aurA–C*, encoding a bacterial type I PKS, were identified. Functional features of the deduced 1906 aa, 2164 aa, and 2201 aa gene products were determined by comparing them with known type I PKS systems and detection of signature motifs of the domains. The aureothin PKS shows some unusual characteristics with regard to PKS initiation and termination and, most remarkably, in chain elongation.

Analysis of AurA reveals a monomodular protein that is assumed to catalyze the first steps in polyketide assembly. The module harbors KS, AT, and ACP domains, as well as two keto processing domains (KR and DH). The module arrangement corresponds with unsaturated acyl intermediates, which are in fact the proposed products after the first and second elongation steps. Leadlay et al. have reported sequence differences between methylmalonate and malonate-specific acyl transferases, which allow the prediction of AT substrate specificity [30, 31]. In accord with the structure of aureothin, the AT1 signature motif indicates specificity for methylmalonyl-CoA. At its N terminus, AurA shows only very low homology to an ACP domain and the common ACP signature motif (GxDS) is absent, as reconfirmed by repeated sequencing of this region.

AurB is a monomodular enzyme consisting of KS, AT, and ACP domains and an additional set of KR, DH, and ER, which is capable of performing a full reductive cycle. The module composition and the presence of a malonate-specific AT domain provides strong evidence that this module is involved in the third round of elongation and reduction.

AurC represents a bimodular protein with a thioesterase (TE) domain located at the C terminus. The TE domain contains characteristic GxSxG and GxH motifs and has high homology to various authentic TE domains from type I PKS. Consequently, the two PKS modules encoded by *aurC* would catalyze the last two Claisen condensations. In agreement with this assumption, each module of AurC is composed of KS, AT, and ACP domains only, leaving the β-keto group unreduced. Upon release from the PKS, the resultant β,δ-diketo acid would cyclize—either spontaneously or catalyzed by the TE—to yield the corresponding pyrone moiety. It is noteworthy that in bacteria similar chemistry is only known from polyketides generated from natural and mutant type II PKS rather than from type I PKS [25].

In modules 1–4, all the PKS domains show typical conserved signature motifs except AT4 (Table 2). In contrast to ATs 1–3, which exhibit the GHSxG active site and (methyl)malonyl-CoA specificity motifs, AT4 shows only weak sequence identity to known AT domains, and both the active site and specificity motifs are aberrant. As confirmed by resequencing the AT4 domain with specific primers and the expression cosmid as template, in the typical GHSxG motif, histidine is mutated to arginine and serine is mutated to alanine. The presence of a GRADG motif most likely results in loss of acyltransferase activity, which is discussed below. Recently, anomalous and possibly inactive AT domains have also been reported for the pyoluteorin PKS [32].

In summary, the *aur* biosynthesis gene cluster encodes one bimodular and two monomodular proteins, which presumably assemble to a tetramodular megasynthase (Figure 5). The domain arrangement of modules 2–4 corresponds perfectly to the degree of reduction of the furane and pyrone moieties of aureothin. However, while the architecture of AurA logically accounts for double bond formation, it appears that only one module catalyzes two successive cycles of chain extension before transfer of the chain to module 2. Such an iterative chain elongation would be in clear contrast to the proposed and generally accepted one-to-one correspondence between encoded modularity of the enzyme activities and structure of the resultant products [33, 34].

Discussion

By reverse genetics, we have localized a 27 kb gene cluster encoding the biosynthesis of aureothin, a rare nitroaryl polyketide metabolite from *S. thioluteus* HKI-227. Detailed sequence analysis and heterologous expression of the *aur* gene set revealed several features that are unique in type I PKS catalyzed polyketide assembly, such as biosynthesis and attachment of the novel nitrobenzoate starter unit. In addition, we provide strong evidence for iteration during polyketide assembly by a type I PKS.

Biosynthesis of the Unusual Nitrobenzoate PKS Starter Unit Involves an Unprecedented N-Oxidase

The vast majority of polyketides is primed with acetate or propionate. However, a number of polyketide synthases utilize alternative starter units, such as short chain branched fatty acids, alicyclic and aromatic acids, and amino acids [24]. The nature of the primer unit often provides important structural and biological features to the molecule, and variation of the starter unit can significantly alter the activity profile of the natural product [24]. An unprecedented feature of the *aur* PKS is the utilization of p-nitrobenzoate (PNBA) as starter unit (J.H. and C.H., unpublished data). PNBA is derived from p-aminobenzoate by means of *N*-oxidation. To date, the only example of a nitro group-forming enzyme has been implicated in the context of pyrrolnitrin biosynthesis. In vitro experiments suggested that *N*-oxidation is catalyzed by a chloroperoxidase, but gene analysis disclosed a class 1A oxygenase [35, 36]. The putative *N*-oxidase required for chloramphenicol biosynthesis is as yet unknown [37]. By inactivation, we were able to demonstrate that the only deduced *aur* gene product that has sequence homology to an oxidase is not involved in the oxidation of PABA to p-nitrobenzoate. Our heterologous expression

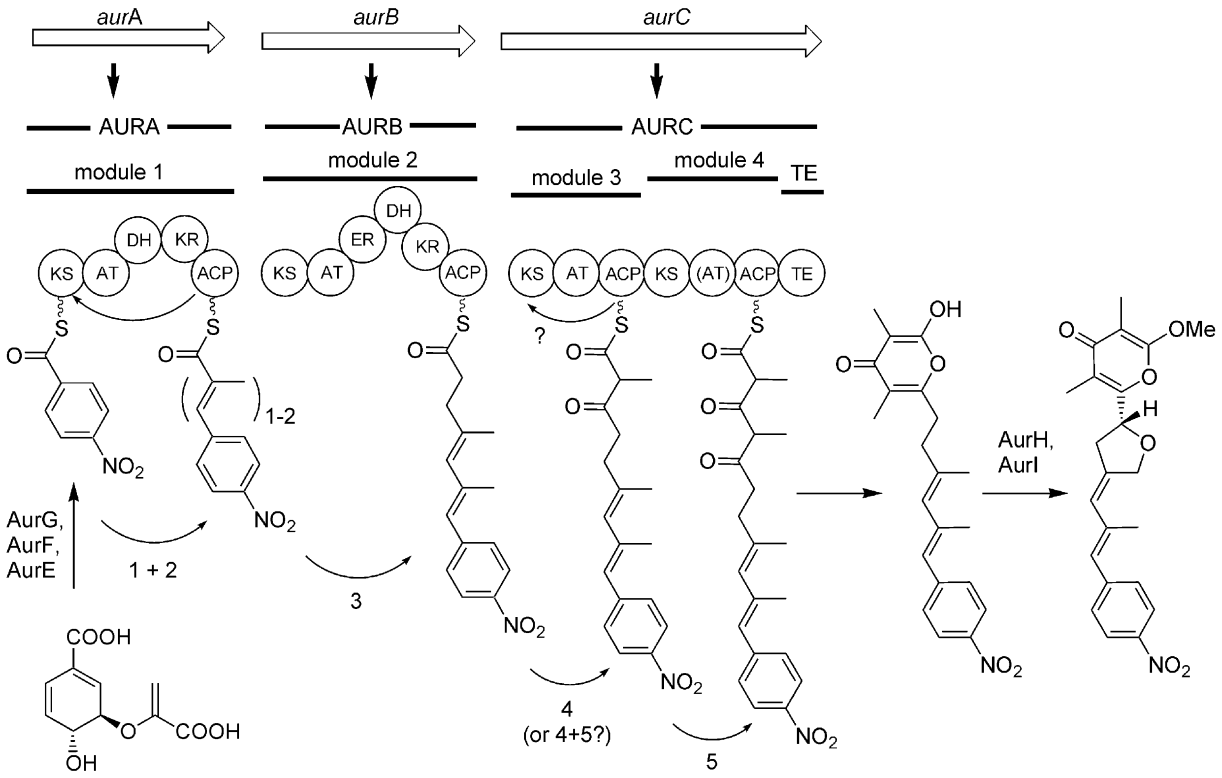

Figure 5. Model for Aureothin Biosynthesis Representing Domain Architecture of AurA, AurB, and AurC and Proposed Polyketide Chain Propagation

MCoA, malonyl-CoA; mmCoA, methylmalonyl-CoA.

experiments unequivocally proved that the *N*-oxidase involved in aureothin biosynthesis is encoded by *aurF* (J.H. and C.H., unpublished data). Surprisingly, the deduced 336 aa gene product (AurF) does not show any sequence homologies to known proteins and not even a known cofactor binding motif could be determined. Thus, the *N*-oxidase AurF could be unique.

PKS Domain Architecture Implicates a Novel Priming Mechanism

For bacterial type I PKSs that utilize starter units other than acetate or propionate, two priming mechanisms are known. Free carboxylic acids may be activated and loaded by a nonribosomal peptide synthetase (NRPS)-like adenylation-thiolation (A-T) didomain at the N terminus of the PKS. Alternatively, acyl-CoA ligases may activate the carboxylates as CoA thioesters, which are loaded onto the PKS by the type AT_L/ACP_L loading didomain [24]. Conversely, neither A-T nor AT_L/ACP_L domains could be identified at the N terminus of the first PKS module AurA. A conceivable mechanism could be the activation of PNBA by the putative acyl-CoA ligase AurE and direct transfer onto the PKS. A similar scenario could be considered for CoA ligase-acyl carrier protein ($CoL\text{-}ACP_L$)-loading didomains, which have been identified in the candicidin [22], rapamycin [4], and pimaricin [38] biosynthesis gene clusters. However, since the N terminus of AurA lacks the ACP signature motif for phosphopantheinyl binding, the priming mechanism of the aureothin PKS remains puzzling, and further biochemical studies are required to understand this unusual set up.

Five Claisen Condensations Are Catalyzed by Only Four PKS Modules

A considerable number of bacterial type I PKS gene clusters have been cloned and sequenced so far. The discovery that the structure of the polyketide metabolite has a one-to-one correspondence to the type I PKS architecture serves as a common model for understanding the programming of modular polyketide synthases and has provided the molecular basis for targeted genetic manipulations. Through detailed sequence analysis and heterologous expression of the full set of aureothin biosynthetic genes we have proven that four PKS modules are sufficient for the catalysis of five elongation and reduction cycles. Module architecture and AT specificity strongly suggests that the first module of the aureothin PKS, AurA, is used twice.

Although this finding is contrary to the principle of colinearity, it may be rationalized in the context of a few recent observations. In both natural and engineered polyketide-producing organisms, traces of metabolites with C2-elongated polyketide backbones have been detected, which do not accurately correspond to the programmed biosynthetic sequence. Höfle et al. have isolated ring-enlarged homologs of epothilone from a *Sorangium cellulosum* So ce90 culture broth [39]. Lead-

lay, Staunton, and coworkers have investigated a mutant strain of the erythromycin producer *Saccharopolyspora erythraea* that produced two novel octaketides in addition to the heptaketide aglycon DEBS, which is the major metabolite [40]. The occurrence of such side products has been interpreted by Leadlay et al. as an aberrant repeated use of a module, which is referred to as "stuttering" [40, 39]. However, these metabolites are considered as the result of error-prone programming of the requisite polyketide synthases and are thus only formed in trace amounts. In contrast, during aureothin polyketide assembly iteration is a programmed event, with exclusive formation of the pentaketide.

At this stage, we cannot rule out that AurA may be present in two identical copies (or possbly two variants of AurA) docking one behind the other, giving rise to a perfectly colinear pentamodular PKS. Alternatively, the repeated use of AurA may be truly iterative, as, for example, in fungal PKS or type I fatty acid synthases. On the basis of the current model of type I PKS as a helically arranged homodimer [41, 42] and considering the domain architecture of AurA, for this model only one mechanism seems plausible (see Figure 5): the acyl intermediate would be passed back from the ACP onto the preceding KS domain, as proposed by Leadlay and Staunton for the biosynthesis of minor side products of the erythromycin PKS [40].

In addition to AurA, we assume that iteration also occurs in AurC. As sequence analyses reveal, the AT4 domain is possibly inactive due to the lacking active site serine; furthermore, the specificity region is aberrant. Since module 3 has the same domain arrangement as module 4, one may speculate that both final chain extensions are catalyzed by module 3 alone, according to the mechanism mentioned above.

The iterative use of a single module has also been suggested by Müller, Höfle, and coworkers for the stigmatellin (*sti*) PKS, which exhibits an unusual architecture. In the sequenced region, one *sti* PKS module bearing the minimal set of KS, AT, and ACP domains is missing. Müller et al. suggest that the missing module is compensated by the iterative operation of another PKS module, either StiH or StiJ. However, despite thorough investigations on the basis of sequence analyses and inactivation experiments, the possibility of a missing module encoded somewhere else in the chromosome of *S. aurantiaca* could not be fully excluded [43].

By our heterologous aureothin PKS expression experiments we provide the yet strongest body of evidence that the iterative use of a module is an inherent property of a natural PKS. At present, it is not possible to deduce the iterative and noniterative nature of the *aur* PKS modules by comparison of sequences alone. Furthermore, it is a riddle how this unusual system has evolved. Apparently, modular PKS may have an intrinsic ability to work in an iterative fashion, as recent heterologous expression experiments of individual PKS modules demonstrate [44], and possibly the line between iterative and noniterative PKS is more blurred than commonly believed. Presumably, the iteratively used *aur* PKS modules are functionally related to both bacterial modular PKS and fungal iterative type I PKS and may well represent an evolutionary crosslink to the purely iterative type I PKS.

Comparing the structures of aureothin (diene) and neoaureothin (tetraene), it is tempting to speculate whether a homolog of AurA in *S. orinoci* catalyses four rounds of chain elongations in neoaureothin biosynthesis (as opposed to two elongations in *S. thioluteus*). We are currently undertaking detailed molecular analysis of the neoaureothin biosynthetic gene cluster, to shed more light on this question.

Significance

Modular polyketide synthases (PKS) play a significant role in the biosynthesis of a vast array of therapeutically important natural products. Typically, these giant multidomain enzymes are arranged in an assembly line fashion, harboring a catalytic domain for each biosynthetic step. Detailed sequence analysis of the aureothin (*aur*) PKS reveals several features that are highly unusual in type I PKS-catalyzed polyketide assembly, including a novel priming mechanism and the use of p-nitrobenzoate as a starter unit. The most striking and unexpected discovery is that four PKS modules are capable of catalyzing five rounds of elongation and reduction, which breaches the accepted principle of colinearity. Our expression experiments provide the best evidence yet that a natural PKS is programmed for the iterative use of a module. The aureothin synthase may thus represent an important evolutionary link between FAS and PKS and provide a unique model system to learn more about the mechanism and evolution of iterative and noniterative polyketide synthases.

Experimental Procedures

Bacterial Strains and Culture Conditions

Aureothin producer *Streptomyces thioluteus* HKI-227, obtained from the HKI strain collection, was used as the source of DNA in the construction of the genomic DNA library. *S. lividans* ZX1 [15], kindly provided by X. Zhou, served as host strain for heterologous expression experiments. For aureothin production, wild-type and mutant strains were cultivated in M10 medium (yeast extract 4 g/l, malt extract 10 g/l, glucose 4 g/l, pH 7.3) for 5 days at 28°C with shaking. *S. lividans* ZX1 was cultured on R5 agar and YEME liquid medium [45] for protoplast transformation and on 2CM [46] for all other experiments. Transformants were selected with apramycin (Sigma) at 30 μg/ml in both solid and liquid medium. *E. coli* strains EPI 100 (Epicentre) and strain DH5α [47] served as hosts for library construction and routine subcloning, respectively. *E. coli* strains were grown in LB medium [47] supplemented with ampicillin (100 μg/ml) or apramycin (50 μg/ml) for selection of plasmids.

Plasmids and General DNA Procedures

DNA isolation, plasmid preparation, restriction digests, gel electrophoresis, and ligation reactions were conducted according to standard methods [45, 47]. pBluescript II SK(−) [48] was the routine vector for subcloning and preparation of DNA templates for sequencing. The integrative *E. coli-Streptomyces* shuttle vector pSET152 [14] was used for all expression experiments in *Streptomyces*. Restriction enzyme-digested DNA fragments were recovered from agarose gel by the GFX PCR DNA and gel band purification kit (Amersham). For Southern blot hybridization, the DIG DNA labeling and detection kit (Roche) was used.

Construction and Screening of a *S. thioluteus* HKI-227 Genomic Cosmid Library

A *S. thioluteus* HKI-227 genomic cosmid library was constructed in *E. coli* EPI 100 with pWEB (Epicentre) as cosmid vector. Genomic

DNA was partially digested with Sau3AI, size fractionated by gel electrophoresis, yielding fragments with an average size greater than 35 kb, and blunted using the end repair mix (Epicentre). The library consisting of 2112 cosmid clones was screened by a PCR-based approach using two pairs of degenerate oligonucleotides, the ketosynthase (KS)-specific primers KSF1 (5′-MGNGARGCNNW-NSMNATGGAYCCNCARCANMG-3′) and KSR1 (5′-GGRTCNCCNA RNSWNGTNCCNGTNCCRTG-3′). For amplification of the ca. 1 kb fragment of the *pabAB* homolog, degenerated primers PABAF1 (5′-GACAACTACGACWSSTTCACS-3′) and PABAR1 (5′-CCTTSRKCT CGTAGCCSAG-3′) were used. PCR products were sequenced and used as DIG-labeled probes for Southern hybridization.

DNA Sequencing and Analysis

Cosmid pST18E4 was sonicated, end repaired by End-Repair Enzyme Mix (Epicentre), and size fractionated by agarose gel electrophoresis. Ca. 2 kb DNA fragments were cloned into pBluescript II SK(−), digested by EcoRV, and end sequenced. Remaining gaps were filled by targeted subcloning and primer walking.

Constructs for Expression of the Aureothin Biosynthesis Genes

In order to shuttle the aureothin biosynthetic gene cluster into the heterologous host *S. lividans* ZX1, a ca. 40 kb SspI fragment, including the whole insert and the *cos* site from cosmid pST18E4, was ligated into the EcoRV site of pSET152. The resulting cosmid pHJ11 was introduced into *S. lividans* ZX1 by protoplast transformation according to standard procedures [45]. Correct transformants were selected and grown in liquid medium to monitor secondary metabolite production. Analogously, integrative expression plasmid pHJ48 was prepared by ligating a 30 kb EcoRV fragment of pST18E4 into the EcoRV site of pSET152.

Isolation and Detection of Metabolites

Aureothin was extracted from chopped agar plates or from liquid culture broths of wild-type and recombinant strains with ethyl acetate. Extracts were concentrated in vacuo and redissolved in MeOH. Aureothin was identified by comparison of the extracts by thin-layer chromatography (TLC), high-performance liquid chromatography (HPLC), and mass spectrometry (MS) with an authentic sample. HPLC analysis was carried out on a KROMASIL C15 column (Jasco) eluted with methanol/water (75:25) in 20 min at a flow rate of 0.8 ml/min and UV detection at 345 nm.

Acknowledgments

This work was supported in part by research grant He 3469/2 by the Deutsche Forschungsgemeinschaft program SPP1125: “Evolution of metabolic diversity” (to C.H.). We thank M. Müller for assistance with HPLC analyses. We are grateful to Prof. Xiufen Zhou of Shanghai Jiaotong University for the kind gift of *S. lividans* ZX1.

Received: August 14, 2003
Revised: September 23, 2003
Accepted: September 23, 2003
Published: December 19, 2003

References

1. O’Hagan, D. (1991). The Polyketide Metabolites (Chichester, UK: Ellis Horwood).
2. Cortés, J., Haydcock, S.F., Roberts, G.A., Bevitt, D.J., and Leadlay, P.F. (1990). An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase *Saccharopolyspora erythraea*. Nature *348*, 176–178.
3. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., and Katz, L. (1991). Modular organisation of genes required for complex polyketide biosynthesis. Science *252*, 675–679.
4. Schwecke, T., Aparicio, J.F., Molnar, I., König, A., Khaw, L.E., Haydock, S.F., Oliynyk, M., Caffrey, P., Cortes, J., Lester, J.B., et al. (1995). The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. Proc. Natl. Acad. Sci. USA *92*, 7839–7843.
5. August, P.R., Tang, L., Yoon, Y.J., Ning, S., Müller, R., Yu, T.-W., Taylor, M., Hoffman, D., Kim, C.-G., Zhang, X., et al. (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.
6. Xue, Y., Zhao, L., Liu, H.W., and Sherman, D.H. (1998). A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metaboic diversity. Proc. Natl. Acad. Sci. USA *95*, 12111–12116.
7. Yu, T.W., Shen, Y., Doi-Katayama, Y., Tang, L., Park, C., Moore, B.S., Hutchinson, C.R., and Floss, H.G. (1999). Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. Proc. Natl. Acad. Sci. USA *96*, 9051–9056.
8. Staunton, J., and Weissman, K.J. (2001). Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. *18*, 380–416.
9. Maeda, K. (1953). A crystalline toxic substance of *Streptomyces thioluteus* producing aureothricin. J. Antibiot. *4*, 137–138.
10. Schwartz, J.L., Tishler, M., Arison, B.H., Shafer, H.M., and Omura, S. (1976). Identification of mycolutein and pulvomycin as aureothin and libilomycin respectively. J. Antibiot. *29*, 236–241.
11. Kakinuma, K., Hanson, C.A., and Rinehart, K.L. (1976). Spectinabilin, a new nitro-containing metabolite isolated from *Streptomyces spectabilis*. Tetrahedron *32*, 217–222.
12. Yamazaki, M., Maebayashi, Y., Katoh, H., Ohishi, J.-I., and Koyama, Y. (1975). Application of 13C-NMR to the biosynthetic investigations. II Biosynthesis of aureothin and related nitro-containing metabolites of *Streptomyces luteoreticuli*. Chem. Pharm. Bull. *23*, 569–574.
13. Criado, L.M., Martin, J.F., and Gil, J.A. (1993). The *pab* gene of *Streptomyces griseus*, encoding *p*-aminobenzoic acid synthase, is located between genes possibly involved in candicidin biosynthesis. Gene *126*, 135–139.
14. Bierman, M., Logan, R., O’Brien, K., Seno, E.T., Rao, R.N., and Schoner, B.E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces spp*. Gene *116*, 43–49.
15. Zhou, X., Deng, Z., Firmin, J.L., Hopwood, D.A., and Kieser, T. (1988). Site-specific degradation of *Steptomyces lividans* DNA during electrophoresis in buffers contaminated with ferrous iron. Nucleic Acids Res. *16*, 4341–4352.
16. Bibb, M.J., Findlay, P.R., and 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.
17. Horinouchi, S., Kito, M., Nishiyama, M., Furuya, K., Hong, S.K., Miyake, K., and Beppu, T. (1990). Primary structure of AfsR, a global regulatory protein for secondary metabolite formation in *Streptomyces coelicolor* A3(2). Gene *95*, 49–56.
18. Edman, J.C., Goldstein, A.L., and Erbe, J.G. (1993). Para-aminobenzoate synthase gene of *Saccharomyces cerevisiae* encodes a bifunctional enzyme. Yeast *9*, 669–675.
19. Brown, M.P., Aidoo, K.A., and Vining, L.C. (1996). A role for pabAB, a p-aminobenzoate synthase gene of *Streptomyces venezuelae* ISPS5230, in chloramphenicol biosynthesis. Microbiol. *142*, 1345–1355.
20. Blanc, V., Gil, P., Bamas-Jacques, N., Lorenzon, S., and Zagorec, M. (1997). Identification and analysis of genes from *Streptomycetes pristinaespiralis* encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I. Mol. Microbiol. *23*, 191–202.
21. Green, J.M., and Nichols, B.P. (1991). *p*-Aminobenzoate biosynthesis in *Escherichia coli*. Purification of aminodeoxychorismate lyase and cloning of *pabC*. J. Biol. Chem. *266*, 12972–12975.
22. Campelo, A.B., and Gil, J.A. (2002). The candicidin gene cluster from *Streptomyces griseus* IMRU 3570. Microbiol. *148*, 51–59.
23. Hu, Z.H., Boa, K., Zhou, X.F., Zhou, Q., Hopwood, D.A., Kieser, T., and Deng, Z.H. (1994). Repeated polyketide synthase modules involved in the biosynthesis of a heptaene macrolide by *Streptomyces sp*. FR-008. Mol. Microbiol. *14*, 163–172.
24. Moore, B.S., and Hertweck, C. (2002). Biosynthesis and attachment of novel bacterial polyketide synthase starter units. Nat. Prod. Rep. *19*, 70–99.
25. Piel, J., Hertweck, C., Shipley, P., Hunt, D.S., Newman, M.S., and Moore, B.S. (2000). 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. Chem. Biol. *7*, 943–955.

26. Xiang, L., Kalaitzis, J.A., Nilsen, G., Chen, L., and Moore, B.S. (2002). Mutational analysis of the enterocin Favorskii biosynthetic rearrangement. Org. Lett. *4*, 957–960.
27. Poupin, P., Ducrocq, V., Hallier-Soulier, S., and Truffaut, N. (1999). Cloning and characterization of the genes encoding a cytochrome P450 (PipA) involved in piperidine and pyrrolidine utilization and its regulatory protein (PipR) in *Mycobacterium smegmatis* mc^2115. J. Bacteriol. *181*, 3419–3426.
28. Weber, J.M., Leung, J.O., Swanson, S.J., Idler, K.B., and McAlpine, J.B. (1991). An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. Science *252*, 114–117.
29. Ikeda, H., Nonomiya, T., Usami, M., Ohta, T., and Omura, S. (1999). Organization of the biosynthetic gene cluster for the polyketide anthelminthic macrolide avermectin in *Streptomyces avermitilis*. Proc. Natl. Acad. Sci. USA *96*, 9509–9514.
30. Haydock, S.F., Aparicio, J.F., Molnar, I., Schwecke, T., Khaw, L.E., König, A., Marsden, A.F.A., Galloway, I.S., Staunton, J., and Leadlay, P.F. (1995). Divergent sequence motifs correlated with substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. FEBS Lett. *374*, 246–248.
31. Aparicio, J.F., Molnar, I., Schwecke, T., König, A., Haydcock, S.F., Khaw, L.E., Staunton, J., and Leadlay, P.F. (1996). Organisation of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. Gene *169*, 9–16.
32. Nowak-Thompson, B., Gould, S.J., and Loper, J.E. (1997). Identification and sequence analysis of genes encoding a polyketide synthase required for pyoluteorin biosynthesis in *Pseudomonas fluorescens* Pf-5. Gene *204*, 17–24.
33. Cane, D.E., and Walsh, C.T. (1999). The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. Chem. Biol. *6*, R319–R325.
34. Walsh, C.T. (2002). Combinatorial biosynthesis of antibiotics: challenges and opportunities. Chembiochem *3*, 124–134.
35. Kirner, S., and van Pee, K.-H. (1994). The biosynthesis of nitro compounds: the enzymatic oxidation to pyrrolnitrin of its amino-substituted precursor. Angew. Chem. Int. Ed. Engl. *33*, 352–354.
36. Kirner, S., Hammer, P.E., Hill, D.S., Altmann, A., Fischer, I., Weislo, L.J., Lanahan, M., van Pee, K.-H., and Ligon, J.M. (1998). Functions encoded by pyrrolnitrin biosynthetic genes from *Pseudomonas fluorescens*. J. Bacteriol. *180*, 1939–1943.
37. He, J., Magarvey, N., Piraee, M., and Vining, L.C. (2001). The gene cluster for chloramphenicol biosynthesis in *Streptomyces venezuelae* ISPS5230 includes novel shikimate pathway homologues and a monomodular non-ribosomal peptide synthetase gene. Microbiol. *147*, 2817–2829.
38. Aparicio, J.F., Fouces, R., Mendes, M.V., Olivera, N., and Martin, J.F. (2000). 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*. Chem. Biol. *7*, 895–905.
39. Hardt, I.H., Steinmetz, H., Gerth, K., Sasse, F., Reichenbach, H., and Höfle, G. (2001). New natural epothilones from *Sorangium cellulosum*, strains So ce90/B2 and So ce90/D13: isolation, structure elucidation and SAR studies. J. Nat. Prod. *64*, 847–856.
40. Wilkinson, B., Foster, G., Rudd, B.A.M., Taylor, N.L., Blackaby, A.P., Sidebottom, P.J., Cooper, D.J., Dawson, M.J., Buss, A.D., Gaisser, S., et al. (2000). Novel octaketide macrolides related to 6-deoxyerythronolide B provide evidence for iterative operation of the erythromycin polyketide synthase. Chem. Biol. *7*, 111–117.
41. Staunton, J., Caffrey, P., Aparicio, J.F., Roberts, G.A., Bethell, S.S., and Leadlay, P.F. (1996). Evidence for a double-helical structure for modular polyketide synthases. Nat. Struct. Biol. *3*, 188–192.
42. Gokhale, R.S., Lau, J., Cane, D.E., and Khosla, C. (1998). Functional orientation of the acyltransferase domain in a module of the erythromycin polyketide synthase. Biochemistry *37*, 2524–2528.
43. Gaitatzis, N., Silakowski, B., Kunze, B., Nordsiek, G., Blöker, H., Höfle, G., and Müller, R. (2002). The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. J. Biol. Chem. *277*, 13082–13090.
44. Beck, B.J., Aldrich, C.C., Fecik, R.A., Reynolds, K.A., and Sherman, D.H. (2003). Iterative chain elongation by a pikromycin monomodular polyketide synthase. J. Am. Chem. Soc. *125*, 4682–4683.
45. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000). Practical Streptomyces Genetics (Norwich, UK: The John Innes Foundation).
46. Li, A., and Piel, J. (2002). A gene cluster from a marine Streptomyces encoding the biosynthesis of the spiroketal polyketide griseorhodin A. Chem. Biol. *9*, 1017–1026.
47. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
48. Alting-Mees, M.A., and Short, J.M. (1989). pBluescript II: gene mapping vectors. Nucleic Acids Res. *17*, 9494.
49. Brautaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strom, A.R., Valla, S., and 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.

Accession Numbers

The sequences reported in this paper have been deposited in the GenBank database (accession no. AJ575648).