Chemistry & Biology
Article

Benzodiazepine Biosynthesis in *Streptomyces refuineus*

Yunfeng Hu,[1] Vanessa Phelan,[1] Ioanna Ntai,[1] Chris M. Farnet,[2] Emmanuel Zazopoulos,[2] and Brian O. Bachmann[1,*]
[1] Department of Chemistry, Vanderbilt University, Nashville, TN 37204, USA
[2] Ecopia Biosciences, Inc., Frederick-Banting, Montreal, Quebec, Canada
*Correspondence: brian.bachmann@vanderbilt.edu
DOI 10.1016/j.chembiol.2007.05.009

SUMMARY

Anthramycin is a benzodiazepine alkaloid with potent antitumor and antibiotic activity produced by the thermophilic actinomycete *Streptomyces refuineus* sbsp. *thermotolerans*. In this study, the complete 32.5 kb gene cluster for the biosynthesis of anthramycin was identified by using a genome-scanning approach, and cluster boundaries were estimated via comparative genomics. A λ-RED-mediated gene-replacement system was developed to provide supporting evidence for critical biosynthetic genes and to validate the boundaries of the proposed anthramycin gene cluster. Sequence analysis reveals that the 25 open reading frame anthramycin cluster contains genes consistent with the biosynthesis of the two halves of anthramycin: 4 methyl-3-hydroxyanthranilic acid and a "dehydroproline acrylamide" moiety. These nonproteinogenic amino acid precursors are condensed by a two-module nonribosomal peptide synthetase (NRPS) terminated by a reductase domain, consistent with the final hemiaminal oxidation state of anthramycin.

INTRODUCTION

Benzodiazepine-containing natural products are produced by a variety of bacteria and fungi of terrestrial and aquatic origin and are notable for their potent antimicrobial and antitumor activities. The marked biological activity in this family of natural products is perhaps unsurprising considering the ubiquity of peripheral benzodiazepine receptors, which are linked to a number of significant biochemical processes [1, 2], and the ability of some benzodiazepines to react irreversibly with DNA. Anthramycin, the first known bacterially produced benzodiazepine alkaloid, is a pyrrolo [1,4]-benzodiazepine from *Streptomyces refuineus* sbsp. *thermotolerans* [3–5]. Anthramycin activity results from noncovalent binding to TA-rich regions in the minor groove of DNA followed by irreversible covalent modification of the N^2-amino group of a vicinal guanine [6]. This mode of action is shared by a family of pyrrolobenzodiazepine containing natural products that includes sibiromycin and tomaymycin and is likely responsible for their varied antibiotic, antiviral, and antitumor activities [7]. While the inherent cardiotoxicity of anthramycin precludes its broad application in the clinic, new classes of antitumor benzodiazepine class compounds have been discovered that, to date, do not posses this disadvantage. For instance ECO-4601, a farnesylated benzodiazepinone natural product from *Micromonospora* with activity against several cancer cell lines, is currently in clinical trials for the treatment of brain, breast, prostate, and pancreatic tumors [8–10].

Hurley et al. have demonstrated that anthramycin is composed of two halves, each derived by oxidative ring opening of an aromatic amino acid followed by subsequent biosynthetic tailoring steps [11] (Figure 1). The 4-methyl-3-hydroxyanthranilic acid half of anthramycin is derived from tryptophan via the primary metabolic kynurenine pathway followed by *S*-adenosylmethionine dependent aromatic *C*-methylation [12]. The "dehydroproline acrylamide" moiety is derived from tyrosine via oxidative ring opening of DOPA [13]. The pattern of tyrosine incorporation of this amino acid into anthramycin is highly analogous to the pattern of incorporation in the antibiotic metabolite lincomycin, which contains an analogous 4-propyl-4,5-dehydroproline moiety [14]. The parallel nature of these pathways is also supported by incorporation of CH_3-labeled methionine into the C-terminal carboxamide of anthramycin and terminal methyl group of lincomycin propyl side chain (Figure 1).

Despite these insightful precursor incorporation studies, insight toward understanding the pathway for formation of the benzodiazepine skeleton has not progressed since the original forays. Here, we report the sequencing, validation, and analysis of the gene cluster for anthramycin biosynthesis. The 32.5 kb gene cluster is comprised of unambiguous gene cassettes for the biosynthesis of the two noncanonical amino acid halves of anthramycin in addition to a reductase-terminated bimodular nonribosomal peptide synthetase system for benzodiazepine ring formation. This study provides, to our knowledge, the first bacterial biosynthetic gene cluster for a biogenic benzodiazepine and provides an NRPS megasynthetase system for a known natural product in a thermophilic actinomycete.

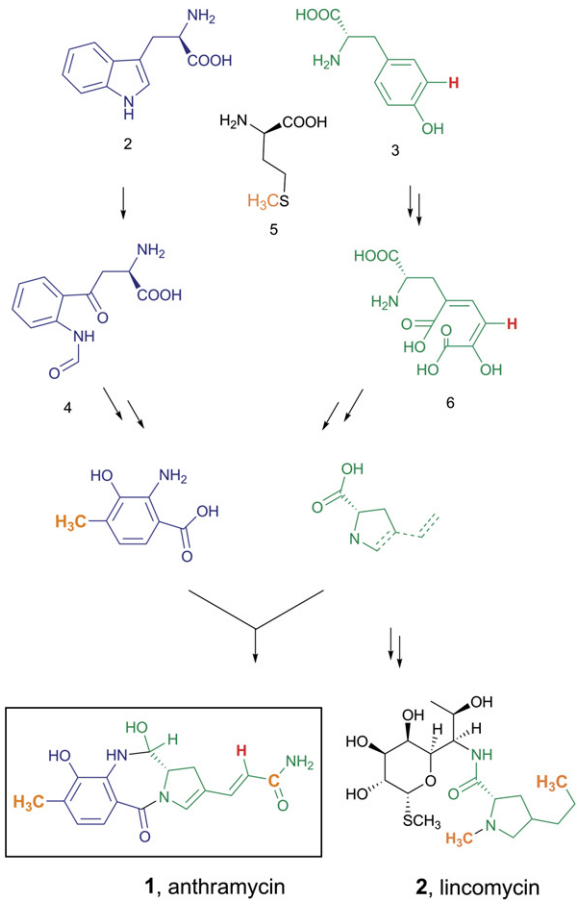

Figure 1. Biosynthetic Overview Based on Isotopic Incorporation Studies

RESULTS

Identification and Cloning of the Anthramycin Gene Cluster from *S. refuineus*

The complete gene cluster for anthramycin biosynthesis was identified by a rapid genome-scanning method [15]. Briefly, a two-tiered genomic DNA library (of 1.5–3 kb and 30–50 kb fragments) was constructed from sonically and enzymatically sheared high-molecular-weight DNA, respectively, isolated from *Streptomyces refuineus* sbsp. *thermotolerans*. The short insert library was cloned into a pBluescript SK+ derivative, and a total of 486 genomic sequence tags (GSTs) were generated by sequencing randomly selected clones. Clones were analyzed for sequence similarity to secondary metabolic genes in the NCBI nonredundant protein database. In this manner, a single GST was identified containing sequence similarity to the predicted requisite NRPS. An 864 member cosmid insert library, corresponding to approximately 3- to 4-fold coverage, was created from the large fragment DNA, and the NRPS GST was labeled and used as colony-hybridization probe for the identification of anthramycin cluster-containing cosmid candidates.

Sequencing and Overview of the Anthramycin Biosynthetic Gene Cluster

Two cosmids were identified by colony hybridization with the NRPS GST and entirely sequenced by a shot-gun sequencing approach. Neighboring cosmids were identified by colony hybridization to provide two overlapping cosmids that putatively spanned a gene cluster comprising genes predicted to be required for anthramycin biosynthesis, including genes with translated similarity to known enzymes in 3-hydroxyanthranillic acid biosynthesis, proteins associated with pyrrolyl-functional metabolite biosynthesis (lincomycin) and NRPS enzymatic systems (Figure 2 and Table 1). The boundaries of the gene cluster were deduced by comparative genomics. Genes with high-translated sequence similarity to bacterial primary metabolism in the nonredundant GenBank database, and not to secondary metabolism genes in GenBank, were putatively assigned to primary metabolism.

Gene-Replacement Methodology and Boundary Validation

Estimated boundaries of the anthramycin gene cluster were validated by a series of PCR-targeted gene-replacement experiments. The two cosmids containing the putative anthramycin biosynthetic genes were retrofitted with an origin of intergeneric transfer, *oriT*, in order to facilitate subsequent conjugation into *S. refuineus*. The general gene-replacement strategy was based on the well-established λ-Red (Red/ET) system (Figure 3) [16, 17]. Briefly, the apramycin resistance gene *aac(3)IV*, or a cassette containing the apramycin-resistance gene *aac(3)IV* and *oriT*, was amplified from plasmid pIJ773 with the 39 bp

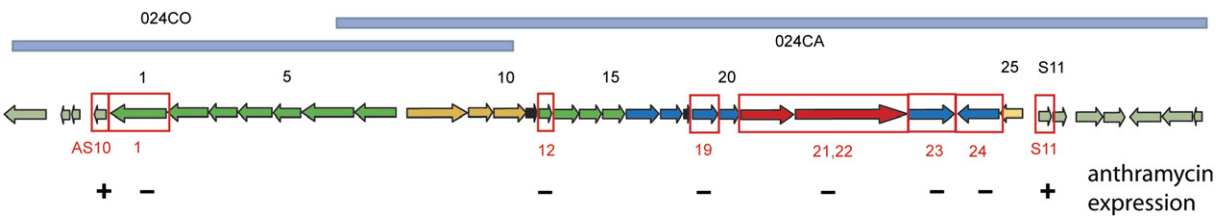

Figure 2. Proposed Biosynthetic Gene Cluster for Anthramycin Is Contained on Two Cosmids 024CO and 024CA

Red boxes summarize gene replacement mutants created in this study; +/− indicates presence or absence of anthramycin in extracts of replacement mutants as analyzed by bioactivity and HPLC/MS.

Table 1. Open Reading Frame Analysis for Anthramycin Gene Cluster

ORF	#AA	GenBank Homology	% Identity/Similarity	Proposed Function
AS10	138	AAU19321.1	46/68	unknown, hypothetical protein
1	624	BAB12569.1	57/68	amidotransferase
2	500	CAD30313.1		aldehyde dehydrogease
3	354	EAO60654.1	50/65	alcohol dehydrogenase
4	410	CAJ23858.1	49/63	Cytochrome P-450 hydroxylase
5	352	487713	79/87	lmbA, methyltranferase
6	621	CAA55746.1	73/80	lmbW, unknown function
7	487	ABF90321.1	38/54	FAD oxidoreductase
8	764	AAL06654.1	59/75	drug-resistence pump
9	256	CAB55527.1	27/40	putative hydroxylase/glyoxylase
10	377	EAL16816.1	41/66	transporter
11	89	–	–	none
12	169	CAA55747.1	48/63	lmbB1, L-DOPA 2,3-dioxygenase
13	302	CAA55748.1	42/54	lmbB2, L-tyrosine 3-hydroxylase.
14	297	CAA55772.1	50/63	lmbY, unk. lincomycin biosynth.
15	276	CAA55771.1	34/41	lmbX, unk. lincomycin biosynth.
16	413	42543461	37/57	kynureninase
17	261	4753870	37/52	tryptophan 2,3-dioxygenase
18	58	–	–	none
19	348	37542638	43/63	aromatic *C*-methyltransferase
20	296	EAU10758.1	36/49	aryl formamidase
21	600	CAD92850.1	35/51	NRPS
22	1146	ABF90459.1	32/46	NRPS
23	500	ABF39686.1	32/49	kynurenine 3-monooxygenase
24	475	ABF87356.1	36/50	flavin-containing oxidoreductase
25	273	BAC79018.1	46/64	repressor-response regulator
S11	174	BAC69182.1	63/68	Zn-dependent hydrolase
S12	199	EAU12469.1	31/53	unknown, hypothetical protein

flanking primers homologous to the area for replacement (see Table S1 in the Supplemental Data available with this article online). Transformation of the linear PCR product into an *E. coli* BW25113/pIJ790 strain bearing the target cosmid and expressing λ-RED recombinase facilitated efficient double crossover recombination within the target cosmid. Gene replacement was assayed by selecting for apramycin resistance, and correct insertion was confirmed by PCR amplification. These suicide cosmids were then transferred into *S. refuineus* sbsp. *thermotolerans* by intergeneric transfer and double crossover events were selected by assaying for resistance to apramycin and sensitivity to the cosmid vector kanamycin resistance. PCR analysis of sensitive colonies was performed as described in Experimental Procedures through genomic DNA amplification with primers internal to the *aac(3)IV* gene and the adjacent anthramycin biosynthetic genes (Table S2).

The genomic deletion of NRPS genes by replacement of *orf21*/*orf22* with the *aac(3)IV* gene resulted in a mutant incapable of producing anthramycin. Duplicate butanol extracts of cultures of (*orf21*/*orf22::aac(3)IV*) were compared to control cultures of the wild-type strain via TLC bioautography with *B. subtillis* and HPLC/MS. Similarly, cluster boundary determinations were supported by gene replacement of *orfAS10* and *orf1*, on the upstream side of the anthramycin gene cluster, and *orfS11*, *orf23*, and *orf24* on the downstream side. The *orf25* gene was not selected for gene replacement due to its high-sequence similarity to known secondary metabolite regulatory genes. Loss of anthramycin production was observed only in (*orf1::aac(3)IV*/*oriT*), (*orf23::aac(3)IV*/*oriT*), and (*orf24::aac(3)IV*/*oriT*) gene-replacement mutants, indicating that sequence-based analysis correctly approximated the boundaries of the anthramycin biosynthetic gene cluster.

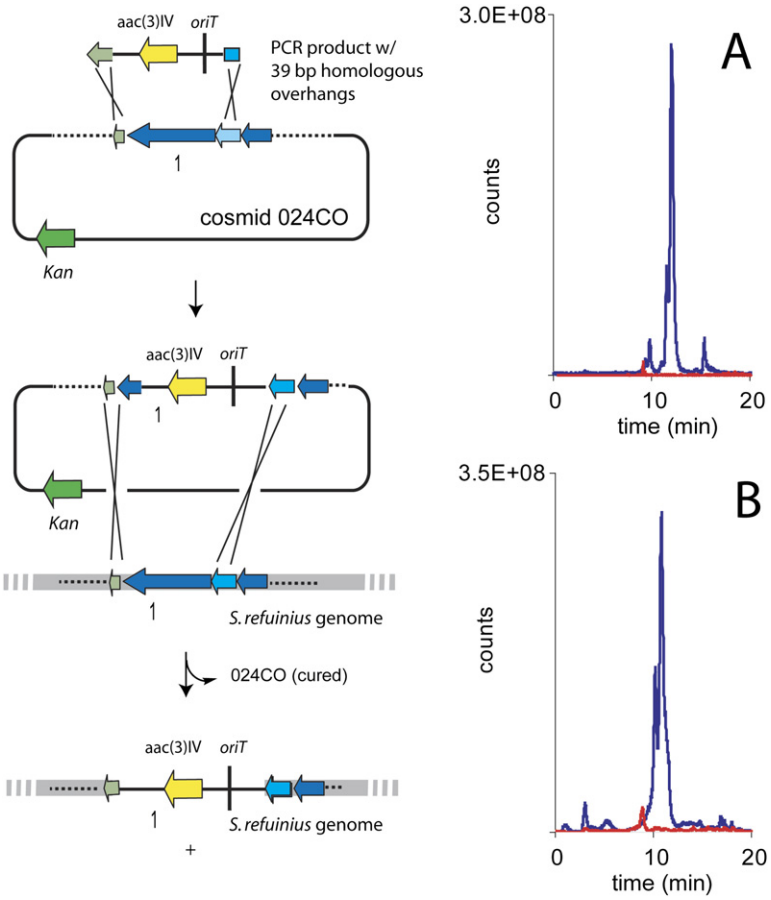

Figure 3. Gene-Replacement Strategy and Detection of Anthramycin

(A) Anthramycin production (ESI+ = 316) in WT (blue) versus (*orf1::aac(3)IV*) (red).

(B) Anthramycin production in (*orf24::aac(3)IV/oriT*) (red) with chemical complementation of compound **9** (blue).

Chemical Complementation

The retinue of putative anthranilic acid biosynthetic mutants created by gene replacement was analyzed by a series of chemical-complementation experiments. Disrupted genes hypothesized to be involved in the biosynthesis of the anthranilate-derived pathway of anthramycin biosynthesis (see below) include *orf23*, *orf24*, and *orf19*. Four possible intermediates, L-kynurenine **8**, 3-hydroxy-L-kynurenine **9**, hydroxyanthranilic acid **10**, and synthetically prepared 4-methyl-3-hydroxyanthranilic acid **11** [18], were tested for their ability to complement the three gene-replacement mutants in precursor feeding experiments. Extracts prepared from gene-replacement mutants grown in the presence of precursors were analyzed by TLC-bioautography (anti-*Bacillus*) and HPLC/MS. Detection of benzodiazepine mass ions unambiguously demonstrated the presence or absence of anthramycin (ES+ = 316 Da) or desmethylanthramycin (ES+ = 302 Da) in culture extracts (Figure 3). Anthramycin equilibrates in the extracts between several isomeric forms in methanol, including the aldehyde (open), hemiaminal, and methyl ether (closed) forms, all of which were observable under our chromatographic conditions. The results of the anthranilate pathway chemical-complementation experiments are summarized in Figure 4. In addition to providing evidence for the involvement of *orf23* and *orf24* in the biosynthesis of the anthranilate-derived portion of anthramycin, these data demonstrate that in the case of *orf19*, *orf23*, and *orf24*, these disruptions resulted in minimal polar effects in the anthramycin pathway.

DISCUSSION

With the complete gene cluster for anthramycin in hand, the outline for the pathway of assembly of the benzodiazepine structure of anthramycin begins to take shape. The anthramycin biosynthetic gene cluster is comprised of 25 genes including a bimodular NRPS system, genes for 4-methyl-3-hydroxy-anthranilic acid biosynthesis, and a series of genes for the biosynthesis of the pyrrolylacrylamide moiety. Additionally, gene candidates for regulation, resistance, and transport can be suggested by sequence analysis.

The two amino acid-derived halves of anthramycin are likely to be activated via adenylation and thioesterification on the bimodular megasynthetase. While the precise identities of the two amino acid NRPS substrates are yet to be determined, it is likely that the first module activates the 4-methyl-3-hydroxyanthranilic acid, or a closely related precursor, which becomes thioesterified on the first T domain, and the second module activates a dehydroproline acrylamide precursor. *C*-methylation of anthranilic acid

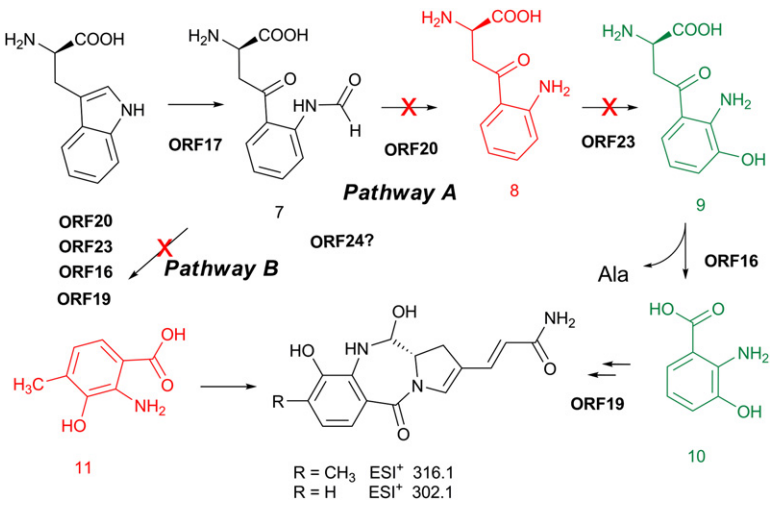

Figure 4. Above, a Possible Mechanism of Methylanthranilic Acid Biosynthesis, and below, Summary of Chemical Complementation Experiments

Theoretical versus observed prediction of masses is based on this scheme, though other schemes are possible.

may occur prior or subsequent to benzodiazepine-core formation. Similarly, the dehydroproline acrylamide moiety may be fully elaborated prior to activation and thioesterification, as apparently occurs in the biosynthesis of the structurally similar lincomycin natural products, or multiple modifications (oxidation, amidation) may take place following peptide-bond synthesis. Somewhat unusually, the anthramycin NRPS contains a putative terminal NADH-binding domain in place of the traditional thioesterase domain. This is consistent with the oxidation state of the hemiaminal of anthramycin and suggests that the dipeptide is released by a reductive mechanism. Intramolecular addition of the anthranillate-derived arylamine to the dehydroproline acrylamide aldehyde most likely completes the formation of the seven-membered hemiaminal benzodiazepine skeleton of anthramycin.

4-Methyl-3-Hydroxyanthranilic Acid Biosynthesis

The 4-methyl-3-hydroxyanthranilic acid moiety in anthramycin is shared by a number of peptide secondary metabolites, notably the venerable *Streptomyces* secondary metabolite actinomycin [19]. Several of the genes putatively assigned for the biosynthesis of 4-methyl-3-hydroxyanthranilic acid are comprised of homologs of primary metabolic genes involved in 3-hydroxyanthranilic acid biosynthesis (*orf16*, *orf17*, *orf20*, and *orf23*) (Figure 4). As this metabolite is of central importance in primary metabolism as a progenitor of the nicotinate-derived cofactors, these genes may represent additional copies of these essential housekeeping genes. Alternatively, the substrate specificity of these genes and/or the order of their application in the pathway may differ from the primary metabolic paradigm. Due to clustering and sequence analysis, we speculated that ORF24 may also be involved in 4-methyl-3-hydroxyanthranilic acid biosynthesis. However, the precise role of the *orf24* gene in anthramycin biosynthesis is less clear. The translated gene product has high similarity to flavin-containing L-amine oxidases, and we speculate that ORF24 may also play a role in the oxidative generation of **8** and **9** via an oxidative cleavage mechanism.

The only remaining activity required to biosynthesize the aromatic half of anthramycin is a *C*-methyltransferase to append the aryl methyl group. There are two putative methyltransferase genes in the anthramycin gene cluster, *orf5* and *orf19*. ORF5 is likely to be involved in dehydroproline acrylamide biosynthesis as it is 73% identical to LmbW of lincomycin 4-propyl-4,5-dehydroproline biosynthesis (see below). Hence, by process of elimination, ORF19 is a likely candidate for the requisite aromatic *C*-methylase. This hypothesis is supported by the sequence similarity of ORF19 to SfcF (43%/63%), a putative aromatic *C*-methyltransferase implicated in the biosynthesis of safracin in *Pseudomonas fluorescens* [20]. 3-hydroxyanthranilic acid is a known primary metabolite, and we have identified four genes with high translated sequence similarity to the primary metabolic pathway enzymes. However, the question of timing of events, particularly methylation, remains unknown. In Figure 4, we delineate two general possibilities. In pathway A, *C*-methylation is postulated to occur subsequent to 3-hydroxyanthranilic acid formation, preserving the order of primary metabolic precedent. In this case, introduction of aryl *C*-methyl may occur via intermediate **11**, by direct methylation of **10**, or may occur during or after dipeptide assembly. In pathway B, methylation is inserted at an unknown position in the primary anthranilic acid pathway, causing downstream divergence of enzymes specificity in the biosynthesis of **11**.

To support the roles of these genes in 4-methyl-3-hydroxyanthranilic acid biosynthesis and to gain insight into the timing of methylation, gene-deletion mutants were supplemented with exogenously added anthranilate pathway intermediates and **8**, **9**, **10**, and **11**. 3-hydroxy-L-kynurenine **9** and 3-hydroxyanthranilic acid **10** were demonstrated to restore benzodiazepine biosynthesis when added to cultures of the (*orf19::aac(3)IV*), (*orf23::aac(3)IV/oriT*), and (*orf24::aac(3)IV/oriT*) mutants, as demonstrated by analysis of culture extracts by TLC bioautography and confirmed with HPLC/MS, whereas compound **8** and 3-methyl-4-hydroxyanthranilic acid **11** failed to complement any of the mutants. For pathway A to be

operative and to encompass this data, either the transport of **8** must be unfavorable under our fermentation conditions, or the order of aryl oxidation must differ slightly from primary metabolism (for instance, **7** may be the substrate for hydroxylation instead of **8**). Furthermore, in the context of pathway A, these data suggest that *C*-methylation does not proceed via intermediate **11**. For pathway B to be operative and to encompass this data, it must be postulated that the transport of intermediate **11** is unfavorable under our fermentation conditions. Unfortunately, we were unable to detect accumulated intermediates (with one notable exception below) in our gene-disrupted strains. However, since the intermediacy of compound **9** is legitimized by its ability to complement anthramycin biosynthesis, we favor pathway A. Further biochemical experiments are needed to fully resolve the issue of the timing of methylation and the intermediacy of **11**.

While feeding studies for the anthranilate-derived half of anthramycin may be obfuscated by the similarity of this pathway to primary metabolism, the possibility of shunt pathways, and issues pertaining to transport, it is unambiguous that benzodiazepine biosynthesis can be restored in all three mutants by the addition **9** and **10**, indicating that targeted gene disruptions resulted in minimal polar effects. In the case of complementation of the *orf23* and *orf24* mutants, but not the *orf19* mutant, with **10**, we were able to confirm the production of a major shunt antibiotic desmethylanthramycin (previously referred to as ''yellow pigment'' [5]) that accumulated to high levels in this experiment. We therefore propose that ORF19 is the secondary metabolic aromatic *C*-methyltransferase using an unknown aromatic intermediate as a substrate. It may be noted that this activity has been previously demonstrated in actinomycin-producing *Streptomyces*, but the sequence of the isolated enzyme has yet to be reported [18].

Dehydroproline Acrylamide Biosynthesis

The dehydroproline acrylamide moiety biosynthetic cluster is evidenced by a group of genes with high-sequence similarity to six genes involved in 4-propyl-4,5-dehydroproline biosynthesis in lincomycin [21]. The similarity is not surprising considering the correspondence in precursor incorporation experiments in these two pathways, which demonstrate that the respective dihydropyrrolyl components of these natural products are derived from tyrosine and that the acrylamide carbonyl and the propyl methyl group are both derived from methionine. The precise sequence of biotransformations leading to dehydroproline acrylamide remains to be determined, but it is likely that the first two steps of dehydroproline acrylamide biosynthesis are analogous to the corresponding pathway in lincomycin biosynthesis, in which 2,3-extradiol cleavage of dihydroxyphenylalanine results in the formation of imine **12**. While functional analysis of the majority of lincomycin biosynthetic genes has not been performed, it has been demonstrated that *lmbB2* and *lmbB1* (similar to *orf13*, *orf12*) encode enzymes that convert L-tyrosine to L-dihydroxyphenylalanine and catalyze 2,3-extradiol cleavage, respectively [22, 23]. Labeling studies have demonstrated loss of glyoxlyate or formal equivalent in the lincomycin pathway, and we suggest that anthramycin ORF5, with 79% sequence identity to lincomycin methyltransferase LmbA, likely methylates intermediate **13** to α,β-unsaturated imine **14**, a possible branch point at which the biosynthetic pathways of lincomycin and anthramycin diverge (Figure 5). In the case of lincomycin, the 4-propyl-4,5-dehydroproline precursor **17** may be generated by conjugate reduction of unsaturated imine **14**. In the case of anthramycin, tautomerization of imine **14** may yield a dieneamine as an appropriate substrate for allylic methyl oxidation by ORF4. This putative P-450 hydroxylase is suggestively clustered with a series of enzymes (*orf1-3*-encoding proteins) with high sequence similarity to characterized alcohol dehydrogenase, aldehyde dehydrogenase, and amidotransferase enzymes, which are all formally required to complete the synthesis of acrylamide **16** from allylic alcohol **15**.

Of note, however, no accumulation of dehydroproline acrylamide (or any other new major metabolite) was observed in HPLC/MS analysis of duplicate butanol extracts of anthranilate disruption mutants. This may be a reflection of inefficient partitioning of this zwitterionic compound in butanol, low levels of accumulation of the intermediate, or the possibility that the true substrate for the NRPS catalyzed condensation reaction may in fact be a precursor of amide **16** in final structure anthramycin.

Nonribosomal Peptide Synthetase

The NRPS system is the preferred paradigm for the biosynthesis of small peptides in bacteria [24–26]. In these systems, a modular thiotemplate mechanism is used for assembly of structurally diverse peptides. Domains for adenylation (A domain), amide bond condensation (C domain), and thioesterification (T domain) are concatenated, often collinearly, into ''modules,'' each module corresponding to a peptide homologation step. The release of bound peptide intermediates is catalyzed by terminal thioesterase domains (TE). NRPS systems have been shown to accommodate an increasing inventory of domains for modification of the elongating peptide intermediates including heterocylization, oxidation, and methylation domains, to name a few. The anthramycin cluster contains two NRPS genes encoding a compact series of six catalytic domains in colinear order on two separate modules *orf20* (A-T) and *orf21* (C-A-T-Re) (Figure 6). Notably, the incorporation of an unusual C-terminal reductase domain (Re) in the second module suggests that release from the megasynthetase is accomplished by reduction of the phosphopantathienyl tethered dipeptide, rather than the hydrolysis of a benzodiazepinone intermediate, which would be evidenced by a terminal thioesterase (TE) domain. This reductive release mechanism has also been proposed to be operative in saframycin [27], mycochelin [28], nostocyclopeptide [29], and equisetin [30] NRPS proteins, which also encode terminal reductase domains. Analysis of the anthramycin NRPS A-domains by the established

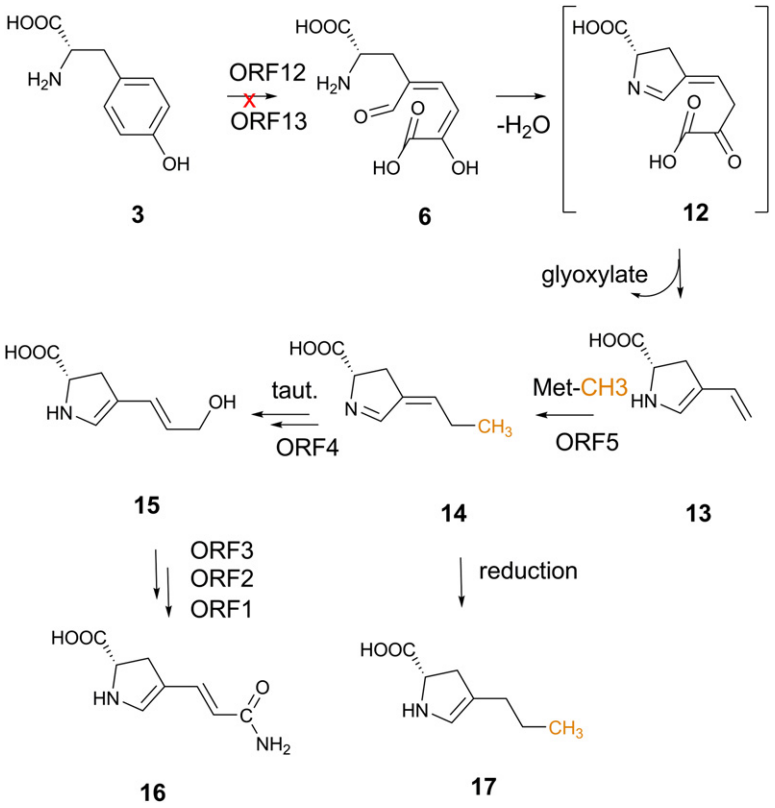

Figure 5. Proposed Dehydroproline Acrylamide Pathway in which 14 Is Hypothesized to Be the Common Intermediate in Anthramycin and Lincomycin

ORF12, ORF13, ORF14, and ORF5 possess significant sequence similarity to the lincomycin biosynthesis pathway.

predictive homology modeling method [31, 32] demonstrates relatively low similarity to known A domains including, notably, actinomycin synthetase, which has been previously demonstrated to activate 4-methyl-3-hydroxy-anthranilic acid [33]. A notable feature of the homology model for the first A domain is that it is missing critical Asp-235, which is generally believed to interact with the α-amino group of the amino acid substrate. This is unsurprising considering the structural divergence of β-amino acid that is proposed to serve as substrate.

Regulation and Resistance

The similarity of ORF8 to the UvrA [34] component of the excision nucleotide repair system and the daunorubicin-resistance protein DrrC [35] suggests a role in resistance. Adjacent genes *orf9* and *orf10* may also be components of the resistance system, encoding proteins with similarity to bleomycin-resistance proteins and drug-efflux proteins, respectively. *orf25* encodes an amino acid sequence with N-terminal sequence similarity to the receiver domain and C-terminal similarity to the DNA-binding domain of a bacterial two-component response system [36] and thus may play a role in coordinating anthramycin biosynthesis with environmental signals.

SIGNIFICANCE

The benzodiazepine natural products are broadly cytotoxic and have been found to have marked selectivity for various cancer cell lines in addition to their antibacterial and antiviral properties. Due to issues of general cytotoxicity and cardiotoxicity, the anthramycin class of natural products has not found human clinical application. However, new promising members of benzodiazepine natural products continue to be discovered. For example, ECO-4601 is a benzodiazepinone with high selectivity for human glioblastoma cell lines and is currently in Phase I clinical trials for treatment of various types of cancer. The deduction of the anthramycin gene cluster provides insight into the biosynthesis of this important class of natural products and will facilitate the identification of new benzodiazepine compounds from microbial genome projects. In addition, most known actinomycetes are strictly mesophillic, with an optimum growth temperature range of 26°C–30°C. The anthramycin-producing organism *Streptomyces refuineus*, originally isolated from an exothermically decaying compost heap, is moderately thermophilic, preferring to grow and to produce anthramycin at 47°C. However, prior to this study, methods for the transformation and genetic manipulation of thermophilic actinomycetes have not been described. Thermophilic actinomycetes may have utility as hosts for heterologous expression of natural product gene clusters. Moreover, the practical advantages for structural studies with enzymes from thermophilic microorganisms are well precedented.

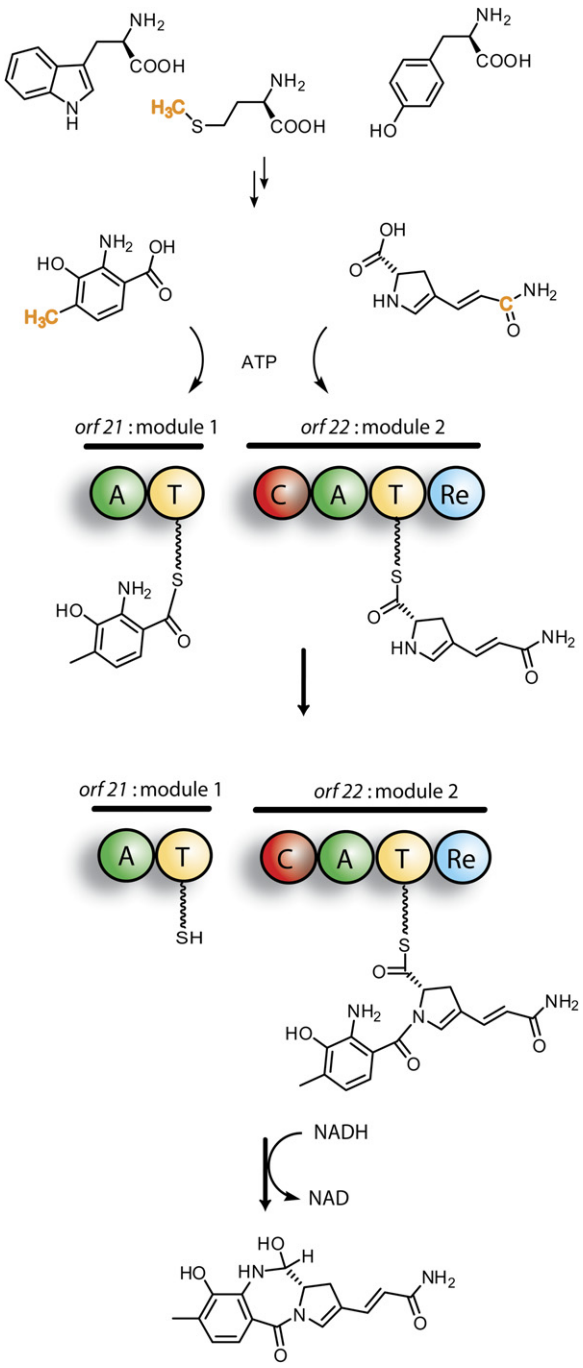

Figure 6. Nonribosomal Assembly and Reductive Release of Anthramycin

EXPERIMENTAL PROCEDURES

Bacterial Strains and Culture Conditions

Streptomyces refuineus var. *thermotolerans* (NRRL 3143) and its derivatives were maintained and grown on either ISP4 medium or TSB medium with appropriate antibiotics at 37°C. For anthramycin production, *S. refuineus* and its derivatives were cultured in production medium (1% corn starch, 2% peptonized milk, and 0.3% yeast extract at pH 7.0) at 47°C. *Bacillus* sp. *TA* (NRRL B-3167) was cultured in nutrient agar and used as a test organism for the antibacterial activity of *S. refuineus* and its derivatives. DH10B (Invitrogen) served as host for *S. refuineus* genomic library construction. *E. coli* BW25113 containing plasmid pIJ790 was used for targeted gene disruption in *S. refuineus*. *E. coli* ET12567 containing the RP4 derivative pUZ8002 was used for intergeneric conjugation between *E. coli* and *S. refuineus*. *E. coli* strains were grown in LB medium supplemented with appropriate antibiotics for selection of plasmids.

Plasmids and General DNA Procedures

SuperCos-1 derivatives were used to construct genomic libraries of *S. refuineus*. DNA isolation and manipulation, cosmid preparation, and gel electrophoresis were conducted according to standard methods. Cosmid DNA was isolated from *E. coli* strains by using Qiagen miniprep kits. Isolation of DNA fragments from agarose was carried out by using Qiagen gel extraction kit. Genomic DNA from *S. refuineus* and its derivatives were isolated by using Wizard genomic DNA purification kit (Promega).

Preparation of Possible Biosynthetic Intermediates of Anthramycin

L-kynurenine, 3-hydroxy-L-kynurenine, hydroxyanthranilic acid were purchased directly from Sigma Corporation. 4-methyl-3-hydroxyanthranillic acid was prepared by catalytic hydrogenation of 4-methyl-3-hydroxy-2-nitrobenzoic acid (Aldrich, Inc.) with Pd/C and H_2 at 1 atm in ethanol [18].

Genomic Sampling Library Construction and Screening

High molecular weight DNA was isolated by established protocols [37], and its integrity was verified by field inversion gel electrophoresis (FIGE) with preset program number 6 (switch time ramp 0.4–1.5 s, linear shape, forward voltage 180, reverse voltage 120) of the FIGE MAPPER power supply (BIORAD, Inc.).

To generate the small insert DNA library, genomic DNA was randomly sheared by sonication. DNA fragments having a size range between 1.5 and 3 kb were fractionated on agarose. The ends of the DNA fragments were repaired by using T4 DNA polymerase (Roche) as described by the supplier. The repaired DNA fragments were subcloned into a derivative of pBluescript SK+vector (STRATAGENE). Plasmid DNA carrying the *Streptomyces refuineus* genomic DNA fragments was extracted by the alkaline lysis method, and the insert size of 1.5 to 3 kb was confirmed by electrophoresis on agarose gels. Using this procedure, a library of small-size random genomic DNA fragments representative of the entire genome of *Streptomyces refuineus* was generated.

The small insert DNA library was analyzed by sequence determination of the cloned genomic DNA inserts to generate genomic sequence tags, GSTs. GST sequencing was performed by using a 3700 ABI capillary electrophoresis DNA sequencer (Applied Biosystems). The average length of the DNA sequence reads was ~700 bp. Further analysis of the obtained GSTs was performed by sequence homology comparison to various protein-sequence databases. The DNA sequences of the obtained GSTs were translated into amino acid sequences and compared to the National Center for Biotechnology Information (NCBI) nonredundant protein database by using known algorithms [38, 39].

A total of 486 *Streptomyces refuineus* GSTs were generated and analyzed by sequence comparison with BLAST. Sequence alignments displaying an E value of at least e-5 were considered as significantly homologous and retained for further evaluation. One GST clone identified by BLAST analysis as encoding a fragment of a nonribosomal peptide synthetase (NRPS) enzyme was selected for the generation of an oligonucleotide probe, which was then used to identify the gene cluster harboring this specific NRPS gene(s) in the large insert cosmid library.

Cosmid Library Construction and Screening

A cosmid library was constructed from the *Streptomyces refuineus* high molecular weight genomic DNA by using a SuperCos-1-derived cosmid vector (STRATAGENE). The cosmid arms were prepared as specified by the manufacturer. The cosmid library consisted of 864 isolated cosmid clones in *E. coli* DH10B (Invitrogen). These clones were picked and inoculated into nine 96-well microtiter plates containing LB broth, which were grown overnight and then adjusted to contain a final concentration of 25% glycerol. These microtiter plates were stored at −80°C and served as glycerol stocks of the cosmid library. Duplicate microtiter plates were arrayed onto nylon membranes and crosslinked onto the membranes by UV irradiation by using a GS GENE LINKER UV Chamber (BIORAD, Inc). Considering an average size of 8 Mb for an actinomycete genome and an average size of 35 kb of genomic insert in the cosmid library, this library represents roughly a 4-fold coverage of the microorganism's genome.

A GST clone identified by BLAST analysis as encoding a fragment of a nonribosomal peptide synthetase (NRPS) enzyme was selected for the generation of an oligonucleotide probe, which was then used to identify the gene cluster harboring this specific NRPS gene(s) in the cosmid library.

Hybridization oligonucleotide probes were radiolabeled with P^{32} by using T4 polynucleotide kinase (New England Biolabs, Inc.) in 15 μl reactions containing 5 pM of oligonucleotide and 6.6 pM of γ-P^{32} ATP in the kinase reaction buffer supplied by the manufacturer. Positive clones were identified, cosmid DNA was extracted from 30 ml cultures by the alkaline lysis method, and the inserts were entirely sequenced by a shotgun sequencing approach [40]. Two overlapping cosmid clones that were detected by the oligonucleotide probe derived from the original NRPS GST clone were completely sequenced to provide approximately 60 Kb of information.

Retrofitting Cosmids with Origin of Transfer

To enable gene deletion, cosmids 024CO and O24CA were retrofitted with *oriT* by replacing kanamycin resistance gene (*neo*) with *oriT* and streptomycin resistance-gene (*aad(A)*) cassette from plasmid pIJ778 with primers 5′-TCAAGATCTGATCAAGAGACAGGATGAGGATCGTT TCGCGCGGCATCTTATTTGCCGAC-3′ and 5′-GCGTCG CTTGGTCG GTCATTTCGAACCCCAGAGTCCCGCGTTCCCGCCAGCCTCGCAG-3′ (pIJ778 homologous sequence underlined) and then replacing *aad(A)* with *neo* with primers 5′-TATGCAGAGG CCGAGGC-3′ and 5′-TATATCGTGCGAAAAAGGA TGGATATACCGAAAA AATCGCCA GAGTCCCGCTCAGAAG-3′ (*neo* homologous sequence underlined) through PCR targeting system. Cosmids 024CO and O24CA containing the *oriT* insertion are designated 024OCO and 024OCA, respectively, and used for the disruption of *orfS10*, *orf12*, *orf19*, and NRPS (*orf21*/*orf22*).

Targeted Replacement of *orfS10*, *orf12*, *orf19*, and NRPS *orf21*/*orf22*

Apramycin resistance gene (*aac(3)IV*) was amplified from plasmid pIJ773 with forward and reverse primers (Table S1) with flanking regions appropriate for gene replacement. The resulting PCR products were transformed into *E. coli* BW25113/pIJ790 containing cosmid 024OCO or 024OCA. *orfS10*, *orf12*, *orf19* and NRPS (*orf21*/*orf22*) were replaced separately by *aac(3)IV*, and the resulting cosmids pBOB3001, pBOB3002, pBOB3003, and pBOB3004 were transformed into *E. coli* ET12567/pUZ8002 and then introduced into wild-type *S. refuineus* strain by intergeneric conjugation. The double-crossover homologous recombination mutant strains were selected for an apramycin-resistant and kanamycin-sensitive phenotype. *orfS10*, *orf12*, *orf19*, and NRPS (*orf21*/*orf22*) mutants were tested by PCR amplification with the primers internal to apramycin-resistance marker *aac(3)IV* and to regions external to the replaced genes (Table S2). The corresponding mutants are designated BOB3001, BOB3002, BOB3003, and BOB3004, separately.

Targeted Replacement of *orf1* and *orfAS11*

The cassette comprising the apramycin-resistance gene (*aac(3)IV*) and *oriT* was amplified from plasmid pIJ773 with appropriate forward primers and reverse primers (Table S1). The resulting PCR products were transformed into *E. coli* BW25113/pIJ790 containing cosmid 024CO (024CA). *orf1* and *orfAS11* were replaced separately by *aac(3)IV* and *oriT* cassette and the resulting cosmids pBOB3005 and pBOB3006 were transformed into *E. coli* ET12567/pUZ8002 and then introduced into wild-type *S. refuineus* strain by intergeneric conjugation. The double-crossover homologous recombination mutant strains were selected for by using apramycin-resistant and kanamycin-sensitive phenotypes. *orf1* and *orfAS11* mutants were tested by PCR amplification with the primers internal to apramycin resistance marker *aac(3)IV* and to regions external to the replaced genes (Table S2). The corresponding mutants were designated BOB3005 and BOB3006, separately.

Targeted Replacement of *orf23* and *orf24*

The apramycin resistance gene cassette, containing (*aac(3)IV*) and *oriT*, was amplified from plasmid pIJ773 with forward primers and reverse primers (Table S1). The resulting PCR products were transformed into *E. coli* BW25113/pIJ790 containing cosmid 024CA. Three hundred and sixty-one base pairs inside *orf23* and 360 bp inside *orf24* were replaced separately by *aac(3)IV* and *oriT* cassette, and the resulting cosmids pBOB3007 and pBOB3008 were transformed into *E. coli* ET12567/pUZ8002 and then introduced into wild-type *S. refuineus* strain by intergeneric conjugation. Apramycin-resistant and kanamycin-sensitive *orf23* and *orf24* disruption mutants were tested by PCR amplification with the primers internal to apramycin resistance marker *aac(3)IV* and to regions external to the replaced genes (Table S2). The corresponding mutants were designated BOB3007 and BOB3008.

Production and Detection of Anthramycin

S. refuineus and its mutant derivatives were cultured in 50 ml seed medium at 47°C for 24 hr. A 5% inoculum was then added to 50 ml production medium (for chemical complementation, corresponding compounds were added directly to the production medium to a final concentration of 2 mM beginning of fermentation) and cultured at 47°C for 24 hr. Anthramycin was extracted from the production medium with 50 ml butanol. Butanol extracts were concentrated in vaccuo and redissolved in MeOH.

Antibacterial activity of anthramycin was detected by thin-layer chromatography bioautography. Anthramycin butaol extracts (dissolved in MeOH) were run on 25DC-Alufolien 20 × 20 cm Kieselgel plates with solvent MeOH:$CHCl_3$ (1:9); LB agar containing indicator strain *Bacillus* sp. *TA* was overlaid on TLC plates and cultured at 37°C for 20 hr to detect anti-*Bacillus* activity (evidenced by growth inhibition zones) of anthramycin.

Anthramycin production was further confirmed by HPLC/MS. Mass spectrometry was performed by using ThermoFinnigan (San Jose, CA) TSQ Quantum triple quadrupole mass spectrometer equipped with a standard electrospray ionization source outfitted with a 100 μm I.D. deactivated fused Si capillary. The injection volume was 10 μl. Anthramycin was separated from cometabolites by using a Jupiter minibore 5 μm C18 column (2.0 mm × 15 cm) with a linear water-acetonitrile gradient (ranging from 95:5 to 5:95 H_2O:CH_3CN) containing 10 mM ammonium acetate. The flow rate was 0.2 ml/min. The mass spectrometer was operated in the positive (or negative) ion mode, and the electrospray needle was maintained at 4,200 V. The ion transfer tube was operated at 35 V and 342°C (−35 V and 300°C for negative). The tube lens voltage was set to 85 V (−220 V for negative). Source CID (offset voltage between skimmer and the first ion guide, Q00) was used at 15 V.

The mass spectrometer was operated in full scan mode with Quad 1. The mass spectral resolution was set to a peak width of 0.70 u (full width at half maximum, FWHM). Full scan spectra were acquired from m/z 150.0 to 700.0 (m/z 150.0 to 1200.0 for negative) over

1.0 s. Data were acquired in profile mode. The electron multiplier gain was set to 3×10^5.

Supplemental Data

Supplemental Data for this article is available online at http://www.chembiol.com/cgi/content/full/14/6/691/DC1/.

ACKNOWLEDGMENTS

Funding was provided by the Vanderbilt Institute of Chemical Biology. Plasmids and the λ-Red recombination kit were graciously provided by Keith F. Chater from John Innes Centre, UK. Cluster identification, sequencing, and initial annotation were performed at Ecopia biosciences, and all genetic manipulation, gene knockouts, complementation studies, and subsequent analyses was performed at Vanderbilt University.

Received: March 12, 2007
Revised: May 15, 2007
Accepted: May 18, 2007
Published: June 22, 2007

REFERENCES

1. Decaudin, D. (2004). Peripheral benzodiazepine receptor and its clinical targeting. Anticancer Drugs *15*, 737–745.
2. Zisterer, D.M., and Williams, D.C. (1997). Peripheral-type benzodiazepine receptors. Gen. Pharmacol. *29*, 305–314.
3. Tendler, M.D., and Korman, S. (1963). "Refuin": a non-cytotoxic carcinostatic compound proliferated by a thermophilic actinomycete. Nature *199*, 501.
4. Leimgruber, W., Stefanov, V., Schenker, F., Karr, A., and Berger, J. (1965). Isolation and characterization of anthramycin a new antitumor antibiotic. J. Am. Chem. Soc. *87*, 5791–5793.
5. Leimgruber, W., Batcho, A.D., and Schenker, F. (1965). Structure of anthramycin. J. Am. Chem. Soc. *87*, 5793–5795.
6. Kopka, M.L., Goodsell, D.S., Baikalov, I., Grzeskowiak, K., Cascio, D., and Dickerson, R.E. (1994). Crystal structure of a covalent DNA-drug adduct: anthramycin bound to C-C-A-A-C-G-T-T-G-G and a molecular explanation of specificity. Biochemistry *33*, 13593–13610.
7. Baraldi, P.G., Bovero, A., Fruttarolo, F., Preti, D., Tabrizi, M.A., Pavani, M.G., and Romagnoli, R. (2004). DNA minor groove binders as potential antitumor and antimicrobial agents. Med. Res. Rev. *24*, 475–528.
8. Bachmann, B.O., McAlpine, J.B., Zazopoulos, E., Farnet, C.M., and Piraee, M. 2006. U.S. patent 7,101,872.
9. Charan, R.D., Schlingmann, G., Janso, J., Bernan, V., Feng, X.D., and Carter, G.T. (2004). Diazepinomicin, a new antimicrobial alkaloid from a marine *Micromonospora* sp. J. Nat. Prod. *67*, 1431–1433.
10. Farnet, C.M., Dimitriadou, V., and Bachmann, B.O. March 2007. U.S. patent 7,186,713.
11. Hurley, L.H. (1980). Elucidation and formulation of novel biosynthetic pathways leading to the pyrrolo[1,4]benzodiazepine antibiotics anthramycin, tomaymycin, and sibiromycin. Acc. Chem. Res. *13*, 263–269.
12. Hurley, L., Das, N., Gairola, C., and Zmijewski, M. (1976). Biosynthetic incorporation of Dl-tryptophan-(5-H-3) into anthramycin, sibiromycin and tomaymycin-Nih shift produced by actinomycetes. Tetrahedron Lett. *18*, 1419–1422.
13. Hurley, L.H., Zmijewski, M., and Chang, C.J. (1975). Biosynthesis of anthramycin—determination of labeling pattern by use of radioactive and stable isotope techniques. J. Am. Chem. Soc. *97*, 4372–4378.
14. Brahme, N.M., Gonzalez, J.E., Rolls, J.P., Hessler, E.J., Mizsak, S., and Hurley, L.H. (1984). Biosynthesis of the lincomycins. 1. Studies using stable isotopes on the biosynthesis of the propyl-L-hygric and ethyl-L-hygric acid moieties of lincomycin-a and lincomycin-B. J. Am. Chem. Soc. *106*, 7873–7878.
15. Zazopoulos, E., Huang, K., Staffa, A., Liu, W., Bachmann, B.O., Nonaka, K., Ahlert, J., Thorson, J.S., Shen, B., and Farnet, C.M. (2003). A genomics-guided approach for discovering and expressing cryptic metabolic pathways. Nat. Biotechnol. *21*, 187–190.
16. Gust, B., Challis, G.L., Fowler, K., Kieser, T., and Chater, K.F. (2003). PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc. Natl. Acad. Sci. USA *100*, 1541–1546.
17. Sosio, M., Giusino, F., Cappellano, C., Bossi, E., Puglia, A.M., and Donadio, S. (2000). Artificial chromosomes for antibiotic-producing actinomycetes. Nat. Biotechnol. *18*, 343–345.
18. Keller, U., Kleinkauf, H., and Zocher, R. (1984). 4-Methyl-3-hydroxyanthranilic acid activating enzyme from actinomycin-producing *Streptomyces chrysomallus*. Biochemistry *23*, 1479–1484.
19. Dalgliesh, C.E., and Todd, A.R. (1949). Actinomycin. Nature *164*, 820.
20. Velasco, A., Acebo, P., Gomez, A., Schleissner, C., Rodriguez, P., Aparicio, T., Conde, S., Muñoz, R., de la Calle, F., Garcia, J.L., and Sánchez-Puelles, J.M. (2005). Molecular characterization of the safracin biosynthetic pathway from *Pseudomonas fluorescens* A2-2: designing new cytotoxic compounds. Mol. Microbiol. *56*, 144–154.
21. Peschke, U., Schmidt, H., Zhang, H.Z., and Piepersberg, W. (1995). Molecular characterization of the lincomycin-production gene-cluster of *Streptomyces lincolnensis*-78-11. Mol. Microbiol. *16*, 1137–1156.
22. Novotna, J., Honzatko, A., Bednar, P., Kopecky, J., Janata, J., and Spizek, J. (2004). L-3,4-dihydroxyphenyl alanine-extradiol cleavage is followed by intramolecular cyclization in lincomycin biosynthesis. Eur. J. Biochem. *271*, 3678–3683.
23. Neusser, D., Schmidt, H., Spizek, J., Novotna, J., Peschke, U., Kaschabeck, S., Tichy, P., and Piepersberg, W. (1998). The genes lmbB1 and lmbB2 of *Streptomyces lincolnensis* encode enzymes involved in the conversion of L-tyrosine to propylproline during the biosynthesis of the antibiotic lincomycin A. Arch. Microbiol. *169*, 322–332.
24. Challis, G.L., and Naismith, J.H. (2004). Structural aspects of nonribosomal peptide biosynthesis. Curr. Opin. Struct. Biol. *14*, 748–756.
25. Finking, R., and Marahiel, M.A. (2004). Biosynthesis of nonribosomal peptides1. Annu. Rev. Microbiol. *58*, 453–488.
26. Schwarzer, D., Finking, R., and Marahiel, M.A. (2003). Nonribosomal peptides: from genes to products. Nat. Prod. Rep. *20*, 275–287.
27. Pospiech, A., Bietenhader, J., and Schupp, T. (1996). Two multifunctional peptide synthetases and an O-methyltransferase are involved in the biosynthesis of the DNA-binding antibiotic and antitumour agent saframycin Mx1 from *Myxococcus xanthus*. Microbiol. *142*, 741–746.
28. Gaitatzis, N., Kunze, B., and Muller, R. (2001). In vitro reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. Proc. Natl. Acad. Sci. USA *98*, 11136–11141.
29. Kopp, F., Mahlert, C., Grunewald, J., and Marahiel, M.A. (2006). Peptide macrocyclization: the reductase of the nostocyclopeptide synthetase triggers the self-assembly of a macrocyclic imine. J. Am. Chem. Soc. *128*, 16478–16479.

30. Sims, J.W., Fillmore, J.P., Warner, D.D., and Schmidt, E.W. (2005). Equisetin biosynthesis in *Fusarium heterosporum*. Chem. Commun. (Camb.) *2*, 186–188.
31. Challis, G.L., Ravel, J., and Townsend, C.A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem. Biol. *7*, 211–224.
32. Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem. Biol. *6*, 493–505.
33. Jones, G.H. (1993). Combined purification of actinomycin synthetase-I and 3-hydroxyanthranilic acid 4-methyltransferase from *Streptomyces antibioticus*. J. Biol. Chem. *268*, 6831–6834.
34. Liu, W., Christenson, S.D., Standage, S., and Shen, B. (2002). Biosynthesis of the enediyne antitumor antibiotic C-1027. Science *297*, 1170–1173.
35. Lomovskaya, N., Hong, S.K., Kim, S.U., Fonstein, L., Furuya, K., and Hutchinson, R.C. (1996). The *Streptomyces peucetius* drrC gene encodes a UvrA-like protein involved in daunorubicin resistance and production. J. Bacteriol. *178*, 3238–3245.
36. Wang, L., and Vining, L.C. (2003). Control of growth, secondary metabolism and sporulation in *Streptomyces venezuelae* ISP5230 by jadW(1), a member of the afsA family of gamma-butyrolactone regulatory genes. Microbiol. *149*, 1991–2004.
37. Kieser, T., Hopwood, D.A., Wright, H.M., and Thompson, C.J. (1982). pIJ101, a multi-copy broad host-range *Streptomyces* plasmid: functional analysis and development of DNA cloning vectors. Mol. Gen. Genet. *185*, 223–228.
38. Altschul, S.F. (1993). A protein alignment scoring system sensitive at all evolutionary distances. J. Mol. Evol. *36*, 290–300.
39. Altschul, S.F., and Lipman, D.J. (1990). Protein database searches for multiple alignments. Proc. Natl. Acad. Sci. USA *87*, 5509–5513.
40. Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science *269*, 496–512.

