The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase

Liangcheng Du, César Sánchez, Mei Chen, Daniel J Edwards and Ben Shen

Background: The structural and catalytic similarities between modular nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) inspired us to search for a hybrid NRPS–PKS system. The antitumor drug bleomycin (BLM) is a natural hybrid peptide–polyketide metabolite, the biosynthesis of which provides an excellent opportunity to investigate intermodular communication between NRPS and PKS modules. Here, we report the cloning, sequencing, and characterization of the BLM biosynthetic gene cluster from *Streptomyces verticillus* ATCC15003.

Results: A set of 30 genes clustered with the previously characterized *blmAB* resistance genes were defined by sequencing a 85-kb contiguous region of DNA from *S. verticillus* ATCC15003. The sequenced gene cluster consists of 10 NRPS genes encoding nine NRPS modules, a PKS gene encoding one PKS module, five sugar biosynthesis genes, as well as genes encoding other biosynthesis, resistance, and regulatory proteins. The substrate specificities of individual NRPS and PKS modules were predicted based on sequence analysis, and the amino acid specificities of two NRPS modules were confirmed biochemically in vitro. The involvement of the cloned genes in BLM biosynthesis was demonstrated by bioconversion of the BLM aglycones into BLMs in *Streptomyces lividans* expressing a part of the gene cluster.

Conclusion: The *blm* gene cluster is characterized by a hybrid NRPS–PKS system, supporting the wisdom of combining individual NRPS and PKS modules for combinatorial biosynthesis. The availability of the *blm* gene cluster has set the stage for engineering novel BLM analogs by genetic manipulation of genes governing BLM biosynthesis and for investigating the molecular basis for intermodular communication between NRPS and PKS in the biosynthesis of hybrid peptide–polyketide metabolites.

Department of Chemistry, University of California, One Shields Avenue, Davis, CA 95616, USA

Correspondence: Ben Shen
E-mail: shen@chem.ucdavis.edu

Keywords: Biosynthesis; Bleomycin; Nonribosomal peptide synthetase; Polyketide synthase; *Streptomyces verticillus*

Received: **10 April 2000**
Revisions requested: **15 May 2000**
Revisions received: **24 May 2000**
Accepted: **30 May 2000**

Published: **1 August 2000**

Chemistry & Biology 2000, **7**:623–642

1074-5521/00/$ – see front matter

PII: S1074-5521(00)00011-9

Introduction

Nonribosomal peptides and polyketides are two large families of natural products that include many clinically valuable drugs, such as vancomycin and erythromycin (antibacterial), cyclosporin and FK506 (immunosuppresant), and bleomycin (BLM) and epothilone (antitumor). The biosyntheses of nonribosomal peptides and polyketides are catalyzed by nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs), respectively. NRPSs and PKSs use a very similar strategy for the assembly of these two distinct classes of natural products by sequential condensation of amino acids and short carboxylic acids, respectively, and utilize the same 4′-phosphopantethein prosthetic group to channel the growing peptide or polyketide intermediates during the elongation processes. Both NRPSs and type I PKSs are multifunctional proteins that are organized into modules. (A module is defined as a set of distinctive domains catalyzing all the steps for one cycle of peptide or polyketide chain elongation and associated modifications.) The number and order of modules and the type of domains within a module on each NRPS or PKS protein determine the structural variations of the resulting peptide and polyketide products by dictating the number, order, choice of the amino acid or carboxylic acid to be incorporated, and the modifications associated with a particular cycle of elongation [1–5]. These striking structural and catalytic similarities between NRPS and PKS inspired us [6] and others [7–12] to search for hybrid NRPS–PKS systems that integrate both NRPS and PKS modules. Since the modular architecture of both NRPS [13–18] and PKS [1–3,19–21] has been exploited successfully in combinatorial biosynthesis of diverse ‘unnatural’

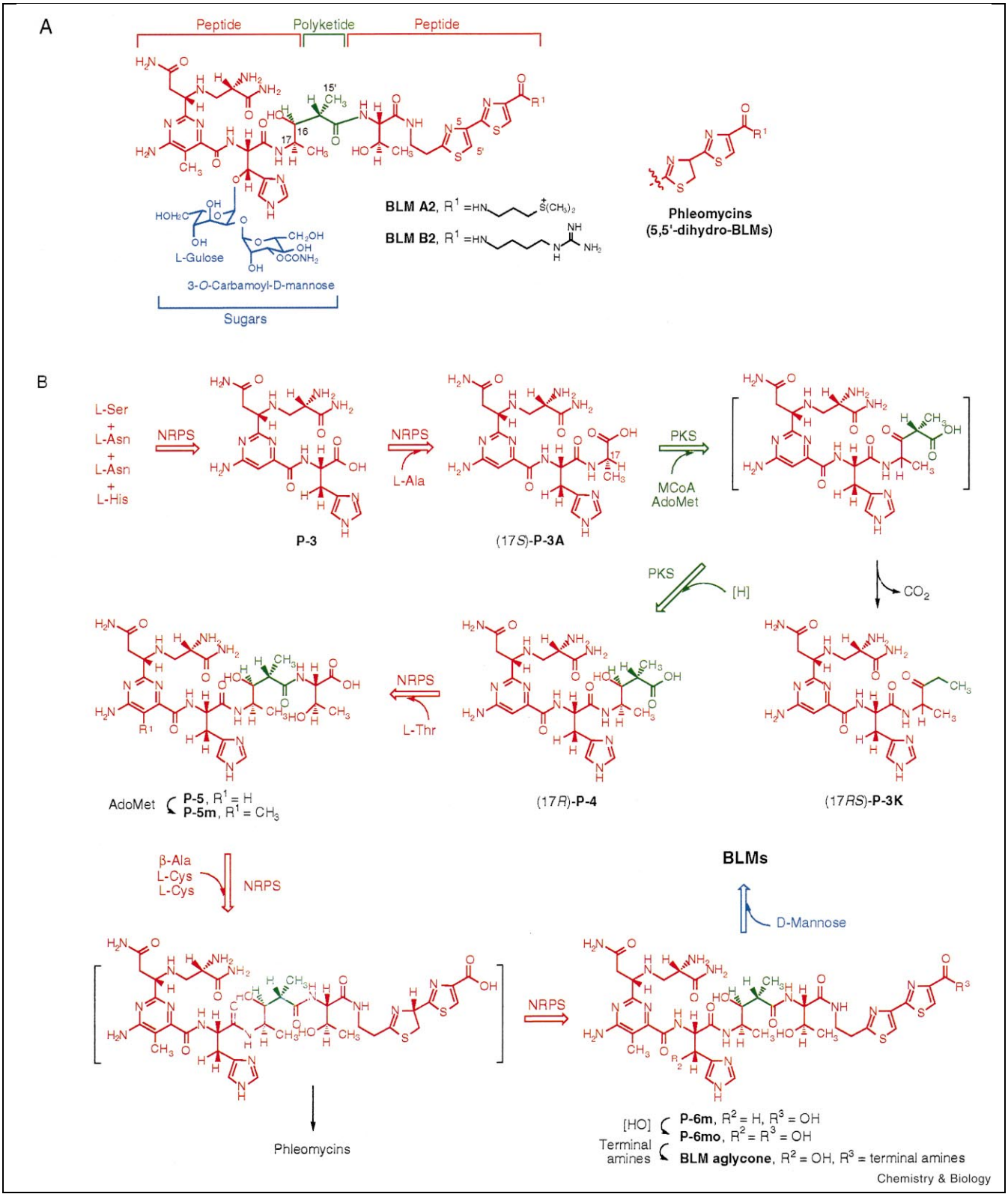

Figure 1. (A) Structures of BLMs and phleomycins and **(B)** proposed biosynthetic pathway for BLMs. Intermediates except those in brackets were identified. Color coding indicates the moiety of BLM that is of peptide (red), polyketide (green), and sugar (blue) biosynthetic origin. Three-letter amino acid designations were used. [H], reduction; [HO], hydroxylation; MCoA, malonyl coenzyme A.

natural products, it is imagined that hybrid NRPS–PKS systems, capable of incorporating both amino acids and short carboxylic acids into the final products, could lead to even greater chemical structural diversity.

The BLMs, a family of antitumor antibiotics produced by *Streptomyces verticillus*, have been the focus of intensive inquiry for more than three decades [22]. Differing structurally at the C-terminal amines of the glycopeptides, the BLMs are incorporated into current chemotherapy of several malignancies under the trade name of Blenoxane® that contains BLM A2 and B2 as the principal constituents (Figure 1A) [23]. Closely related to the BLMs are the phleomycins, the 5,5′-dihydro-analogs of BLMs. The BLMs are thought to exert their biological effects through a metal-dependent oxidative cleavage of DNA and possibly also of RNA in the presence of molecular oxygen [22]. However, wide application of this agent has been prevented by early development of drug resistance and by the dose-limiting pulmonary toxicity. Consequently, there have been continuing attempts to develop new BLM congeners to define the functional roles of the individual subunits and to search for drugs with better clinical efficacy and lower toxicity.

The biosynthesis of BLMs has been extensively studied by feeding isotope-labelled precursors and by isolating various biosynthetic intermediates and shunt metabolites [23–27]. These results unambiguously established the hybrid peptide–polyketide origin of BLMs, the aglycone of which is derived from nine amino acids, one acetate, and two *S*-adenosylmethionines (AdoMet). Driven by the belief that novel congeners could be generated by manipulating genes governing antibiotic production, attempts to clone the gene cluster for BLM biosynthesis were initiated by Sugiyama and co-workers [28] and by Calcutt and Schmidt [29], who both cloned the *blm* resistance genes from *S. verticillus* ATCC15003 in 1994, respectively.

We set out to clone the BLM gene cluster to shed light on hybrid peptide–polyketide biosynthesis, aside from its utilities in generating novel BLM congeners through combinatorial biosynthesis [6]. On the assumption that BLM biosynthesis follows the paradigm for peptide and polyketide biosynthesis, we reasoned that the Blm megasynthetase, which catalyzes the assembly of the BLM backbone from the amino acid and short carboxylic acid precursors, should bear the characteristics of both NRPS and PKS, providing a model to study the mechanism by which NRPS and PKS could be integrated into a productive biosynthetic system to synthesize a hybrid peptide–polyketide metabolite (Figure 1B). Here we report the cloning, DNA sequence analysis, and biochemical characterization of the *blm* biosynthetic gene cluster from *S. verticillus* ATCC15003. Our studies revealed several unprecedented features for hybrid peptide–polyketide biosynthesis, setting the stage to investigate the molecular basis for intermodular communication between NRPS and PKS, and supported the wisdom of combining individual NRPS and PKS modules for combinatorial biosynthesis to make novel 'unnatural' natural products from amino acids and short carboxylic acids.

Results

Cloning and identification of the *blm* gene cluster from *S. verticillus* ATCC15003

Sugiyama and co-workers previously cloned two BLM resistance genes, *blmAB*, from *S. verticillus* ATCC15003 [28]. Extensive biochemical and structural characterizations established that BlmA is a BLM-binding protein, conferring BLM resistance by drug sequestering [30,31], and that BlmB is a BLM *N*-acetyltransferase, inactivating BLM by *N*-acetylation of BLM at the primary amine of β-aminoalaninamide moiety in the presence of acetyl coenzyme A (CoA) [28,32,33]. Subsequently, Calcutt and Schmidt sequenced a 7.2-kb DNA fragment flanking the *blmAB* genes, revealing seven open reading frames (orfs) [29], however, none of which were found to encode putative NRPS or PKS enzymes (Figure 2B). Given the precedent that antibiotic production genes commonly occur as a cluster in actinomycetes [1], we set out to isolate the DNA flanking the *blmAB* resistance locus in order to identify the *blm* biosynthetic gene cluster. An *S. verticillus* ATCC15003 genomic library was constructed in the *Escherichia coli–Streptomyces* shuttle vector pOJ446 [34] and was first screened with probes made from both ends of the *blmAB* locus [29] (Figure 2A, probe-1 and -2), resulting in the isolation of overlapping cosmids extending from this locus. These experiments were then repeated with probe-3 and probe-4 made from the ends of the newly isolated cosmids, leading to the localization of a 140-kb contiguous region of DNA, covered by 16 overlapping cosmids as represented by pBS4 to pBS11 (Figure 2A). The entire 140-kb region was next analyzed by Southern hybridization under various conditions with heterologous NRPS probes to screen for putative NRPS genes. The heterologous NRPS probes were prepared from the three putative NRPS gene clusters isolated previously from *S. verticillus* ATCC15003, none of which was clustered with the *blmAB* resistance locus [6]. While no signal was detected from the 40-kb DNA region downstream of the *blmAB* genes, weak signals were evident at a low stringency hybridization condition from the cosmids within the 100-kb DNA region upstream of the *blmAB* genes [6]. The latter result prompted us to determine the approximately 77.5-kb DNA sequence of this region, covered by pBS5, pBS6, and pBS8, directly upstream of the 7.2-kb *blmAB* locus (Figure 2A).

Nucleotide sequence analysis of the *blm* cluster

The sequenced region was analyzed by the CODONPREFERENCE method available in the GCG software and revealed 40 orfs – only 30 of them are discussed in this

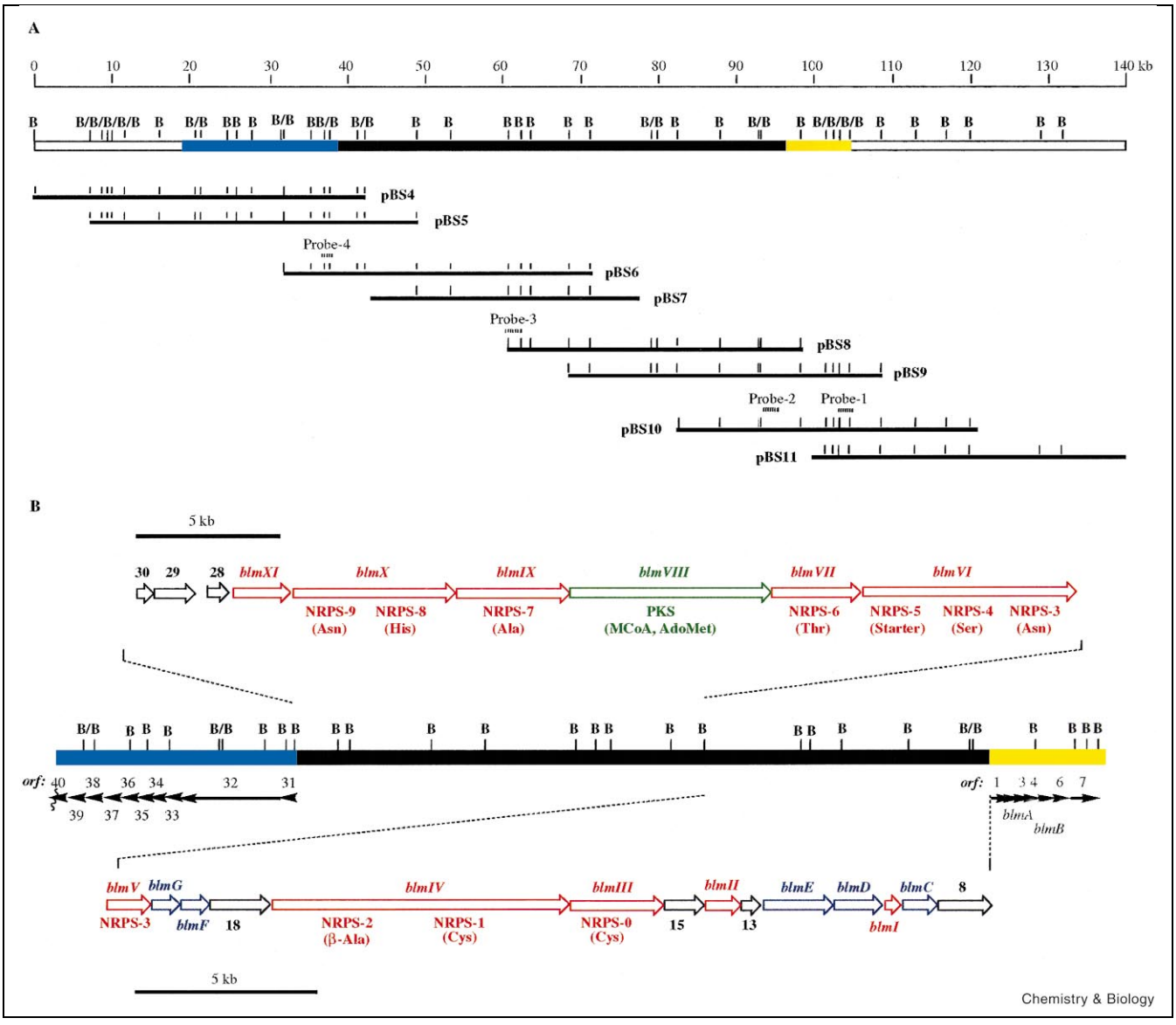

Figure 2. (A) Restriction map of the 140-kb DNA region from *S. verticillus* ATCC15003 as represented by eight overlapping cosmid clones (B, *Bam*HI. Color coding indicates the *blmAB* locus (yellow) [29], *blm* genes reported in this paper (black), and orfs sequenced but not discussed (blue). **(B)** Genetic organization of the BLM biosynthetic gene cluster. Proposed functions for individual orfs are summarized in Table 1. Color coding indicates NRPS genes (red), PKS gene (green), sugar biosynthesis gene (blue), and all other genes (black). Modules for individual NRPS and PKS are given along with their proposed substrates in parentheses.

paper (Figure 2B). The overall GC content of the sequenced region is 75.1%, characteristic for *Streptomyces* species [35]. Preliminary functional assignments of individual orfs were made by comparison of the deduced gene products with proteins of known functions in the database. Among the orfs identified from the cloned cluster, we found (1) one PKS gene (*blmVIII*), (2) 10 NRPS genes (*blmX*, *blmIX*, *blmVII*, *blmVI*, *blmV*, *blmIV*, *blmIII*, *blmI*, *blmII*, *blmXI*), one of which, *blmI*, has been previously characterized to encode a type II peptidyl carrier protein (PCP) [36], (3) five sugar biosynthesis genes (*blmC*, *blmD*, *blmE*, *blmF*, *blmG*), and (4) genes encoding other biosynthesis enzymes as well as resistance and regulatory genes (Table 1).

All the *orfs* between *orf30* and *orf7* are transcribed in the same direction, while the *orfs* at both sides of *orf30-orf7* are transcribed in opposite directions. Possible promoter regions are assumed to have been identified upstream of *orf30*, *orf28*, and *blmV* (Figure 2B). While the boundaries of the *blm* gene cluster are yet to be determined by experiments, it is reasonable to propose that *orf30* and *orf7* rep-

resent the left- and right-hand ends of the *blm* cluster, respectively, on the basis of sequence analysis. *blmX*, *blmIX*, *blmVIII*, *blmVII*, and *blmVI* are likely translationally coupled, as judged by their overlapping stop and start codons, suggesting that they may form part of the same operon. Similarly, *blmV*, *blmG*, *orf18*, *blmIV*, *blmIII*, *orf15*, *blmII*, *orf13*, *blmE*, *blmD*, *blmI*, *blmC*, *orf8*, and *orf1* may form another operon, as judged by their overlapping stop and start codons or by the lack of apparent transcriptional terminator in the short intergenic regions between *blmG* and *blmF* (13 bp), *orf18* and *blmIV* (84 bp), and *orf13* and *blmE* (70 bp). Similar operon-type organizations for both NRPS and PKS genes are well known and have been postulated to facilitate cotranslation of the genes within the operon to yield equimolar amounts of proteins for optimal interactions to form the functional NRPS megasynthetases or PKS megasynthetases [1].

The *blm* NRPS genes

The *blmIII*, *blmIV*, *blmV*, *blmVI*, *blmVII*, *blmIX*, and *blmX* genes encode modular NRPSs consisting of domains characteristic for known NRPSs, such as the adenylation (A) domain [37], PCP [38], condensation (C) domain [39], as well as thiazole-forming specific domains, such as the condensation/cyclization (Cy) domain [11,40] and the oxidation (Ox) domain [11] (Table 1). The most conserved amino acid motifs for these domains are shown in Figure 3 (A–F). BlmVI is unique among all the Blm NRPSs identified. Its N-terminal module (NRPS-5) consists of an unusual domain that bears a close resemblance to a family of acyl CoA ligases (ALs) (Figure 3E), such as the *Mycobacterium bovis* AL (53% similarity and 38% identity) [41], and an acyl carrier protein (ACP)-like domain, such as the one found in EPOS C (47% similarity and 30% identity) [11] – also known as EpoD [12]. These data suggest that

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

Gene	Amino acids	Sequence homolog[a]		Proposed function[b,c]
orf1–orf7				*blmAB* resistance locus [29]
orf8	424	YqeR (BAA12461)		Oxidase
blmC	498	RfaE (AAD49846)		NDP-sugar synthase
blmI	90	GrsB (P14688)		Type II PCP
blmD	545	MmcS (AAD32743)		Carbamoyl transferase
blmE	390	RfaF (AAC45367)		Glycosyl transferase
orf13	187	MbtH (O05821)		Unknown
blmII	462	Nrp (CAA98937)		NRPS condensation enzyme
orf15	339	SyrP (AAB63253)		Regulation
blmIII	935	EPOS P (AAF26925), McbC (P23185)		NRPS domains
			NRPS-0	A PCP Ox
blmIV	2626	HMWP2 (P48633)	NRPS-2	C A PCP
			NRPS-1	Cy A PCP
			NRPS-0	Cy
orf18	638	AsnB (P22106)		Asparagine synthetase
blmF	494	Cps14I (CAA59781)/BlmORF1 (AAB00457)		Glycosyl transferase/β-hydroxylase
blmG	325	YtcB (BBA16584)		Sugar epimerase
blmV	645	McyB (2708278)		NRPS domains
			NRPS-3	PCP C
blmVI	2675	MycA (AF184956), EPOS C (AAF26921)	NRPS-5	AL ACP
		PksD (S73014), SnbDE (CAA67249)	NRPS-4	C A PCP
			NRPS-3	C A
blmVII	1218	SyrE (3510629)	NRPS-6	C A PCP
blmVIII	1841	HMWP1 (CAA73127)		PKS domains
				KS AT MT KR ACP
blmIX	1066	SafB (1171128)		NRPS domains
			NRPS-7	C A PCP
blmX	2140	TycC (2623773)	NRPS-9	C A PCP
			NRPS-8	C A PCP
blmXI	688	SyrE (3510629)		NRPS condensation enzyme
orf28	239	Mth1026 (AAB85522)		Unknown
orf29	582	YchM (CAA73749)		Transmembrane transporter
orf30	113	SmtB (Q55940)		Regulation

[a]Protein accession numbers are given in parentheses.
[b]NRPS and PKS domains are abbreviated as follows: A, adenylation; ACP, acyl carrier protein; AL, acyl CoA ligase; AT, acyltransferase; C, condensation; Cy, condensation/cyclization; KR, ketoreductase; KS, ketoacyl synthase; MT, methyltransferase; Ox, oxidation; PCP, peptidyl carrier protein.
[c]Underlined domains are likely inactive due to the lack of putative active sites or highly conserved motifs.

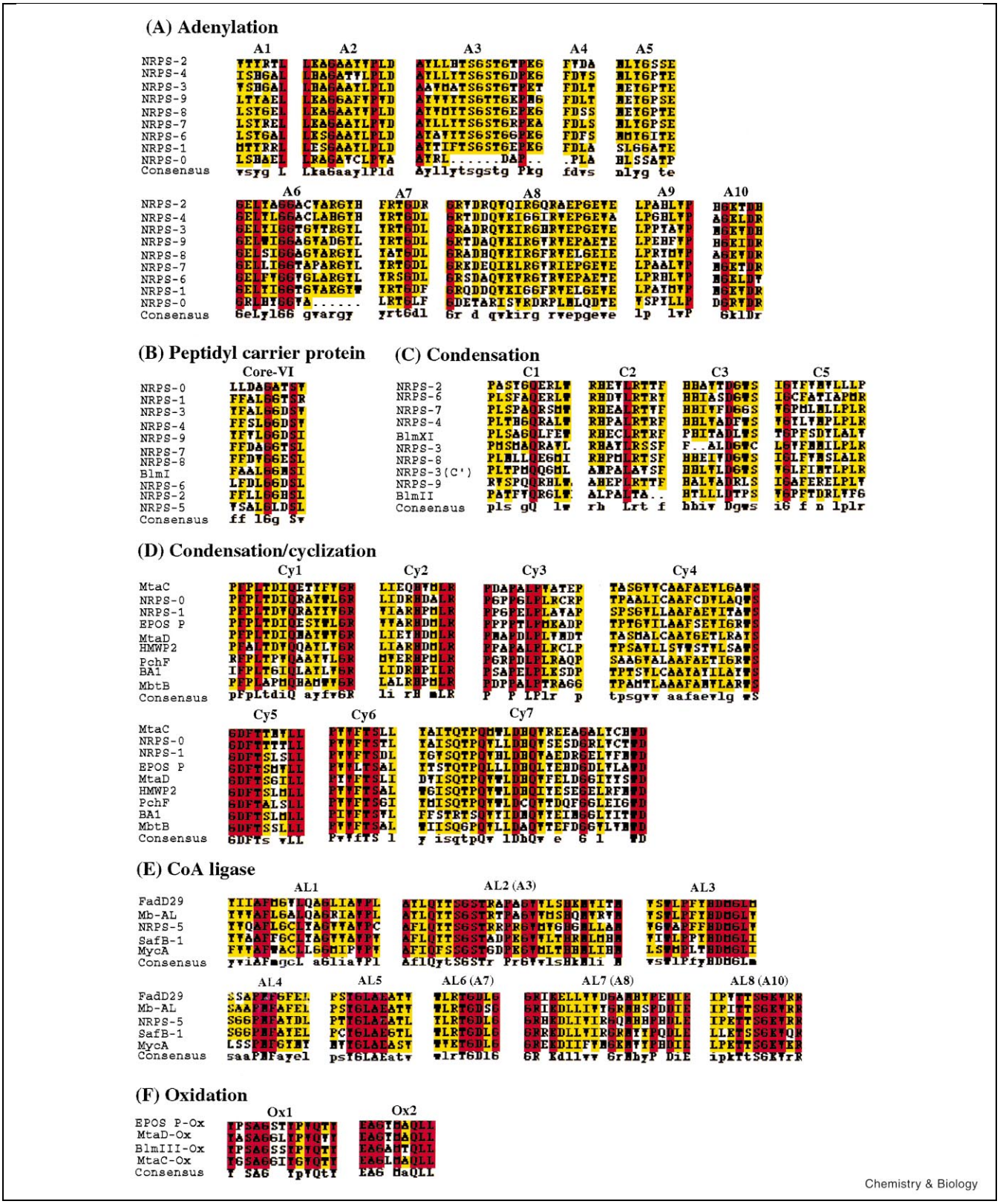

Figure 3. Conserved motifs [4,5] from Blm NRPS and alignments of the Cy, CoA ligase, and Ox domains from Blm NRPS with different Cy domains (MtaC, MtaD [10]; EPOS P [11], HMWP2 [8,47,48]; PchF [96]; BA1 [40]; and MbtB [70]), CoA ligases (FadD29 [97]; Mb-AL [41], MycA [98]; and SafB-1 [77]), and Ox domains (EPOS P-Ox [11]; MtaD- and MtaC-Ox [10]).

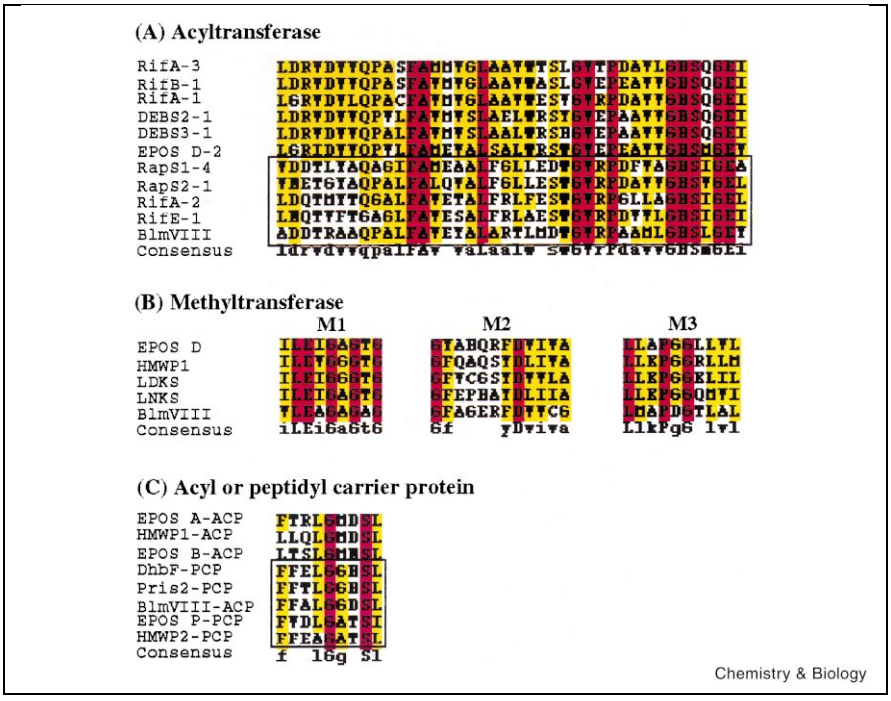

Figure 4. Alignments of the conserved motifs [1] from Blm PKS with different AT domains (RifA-1, -2, -3, RifB-1, RifE-1 [88]; DEBS2-1, DEBS3-1 [99]; EPOS D-2 [11]; RapS1-4, RapS2-1 [89]), MT domains (EPOS D [11]; HMWP1 [8,47,48]; LDKS, LNKS [85]), and ACP or PCP domains (EPOS A-ACP, EPOS B-ACP, EPOS P-PCP [11]; HMWP1-ACP, HMWP2-PCP [8,47,48]). BlmVIII-AT resembles to malonyl CoA-specific ATs (boxed) and BlmVIII-ACP resembles to PCPs (boxed).

NRPS-5 unlikely acts as an NRPS module and hence is named as a starter module, presumably involved in the biosynthesis of the β-aminoalaninamide moiety of BLM. Also noteworthy is the C domain of NRPS-3 that lacks both His residues of the conserved HHxxxDG active site for transpeptidation (C3, Figure 3C) [39].

The *blmI*, *blmII*, and *blmXI* genes encode NRPSs with an unusual architecture. In contrast to the aforementioned modular NRPSs consisting of multiple domains, BlmI (Figure 3B), BlmII, and BlmXI (Figure 3C) are discrete proteins homologous to individual domains of type I NRPSs. We have previously characterized BlmI as a type II PCP [36].

The *blm* PKS gene

The *blmVIII* gene encodes a PKS module consisting of a ketoacyl synthase (KS), an acyltransferase (AT), a methyltransferase (MT), a ketoreductase (KR), and an ACP domain [1]. The KS domain is highly homologous to all known KSs, including the active site Cys, as well as the two highly conserved His residues [42–45]. However, it displays lower similarity to the KSs from *Streptomyces* than to a particular set of KS domains from other bacteria, such as EPOS B (68% similarity and 55% identity) [11] – also known as EpoC [12] – or MtaD (68% similarity and 50% identity) [10]. Close examination of the latter family of sequences revealed that it contains all the KS domains known to interact with NRPS modules in hybrid NRPS–PKS systems [8–12,46–49]. The AT domain is predicted to be specific for malonyl CoA, on the basis of sequence comparison with ATs of known substrate specificity [50], indicative that the PKS module uses malonyl CoA as the preferred extender unit (Figure 4A). The latter is consistent with early feeding experiments that established that the 15′-CH_3 group is derived from methionine, presumably via an AdoMet-dependent MT [26,27]. A candidate for the latter enzyme activity, an MT domain, is identified between the AT and KR domain. The MT domain contains the three core motifs of AdoMet-dependent MTs [51], and is highly homologous to other MT domains of PKS, such as the one found in EPOS D (47% similarity and 32% identity) [11] – also known as EpoE [12] (Figure 4B). The KR domain shows significant homology to known KR domains, such as the one found in MtaB (43% similarity and 31% identity) [10] and EPOS C (44% similarity and 34% identity) [11] – also known as EpoD [12], with the signature NADPH binding site of GxGxxGxxxA [52]. The last domain of the PKS module is an ACP, with its signature motif around the highly conserved 4′-phosphopantethein attachment site Ser residue [53]. Intriguingly, this ACP domain exhibits the greatest similarity to PCP, such as the one found in pristinamycin I synthetase (58% similarity and 46% identity) [54], rather than ACPs, such as the one found in EPOS A (45% similarity and 32% identity) [11] – also known as EpoA [12] (Figure 4C).

Sugar biosynthesis genes

Five putative sugar biosynthesis genes, *blmC*, *blmD*, *blmE*, *blmF*, and *blmG*, are identified within the *blm* gene cluster. The deduced *blmC* gene product shows a high degree of similarity to a family of NDP-sugar synthases, such as RfaE

(50% similarity and 37% identity) [55]. The gene product of *blmD* is closely related to a family of carbamoyl transferases, such as MmcS (45% similarity an 33% identity) [56]. BlmE strongly resembles a family of glycosyl transferase, such as RfaF (45% similarity and 29% identity) [57]. The deduced product of *blmF* appears to be bifunctional. The N-terminal part of BlmF (1–250 amino acids) exhibits low similarity to several glycosyl transferases, such as Cps14I (37% similarity and 26% identity) [58]. The C-terminal part of BlmF (300–494 amino acids) shows low similarity to BlmORF1 [29] and to the catalytic domain of bovine Asp/Asn β-hydroxylase (44% similarity and 35% identity) [59]. BlmORF1 has been proposed by Calcutt and Schmidt as the putative His β-hydroxylase [29], although we were unable to demonstrate this activity in vitro using P-6m as a substrate [6]. The *blmG* gene product shows strong homology to various sugar epimerases, such as YtcB (44% similarity and 28% identity) [60].

Resistance, regulatory, and other genes

Antibiotic production genes have been found nearly in all cases to be clustered in one region of the bacterial chromosome, consisting of structural, resistance, and regulatory genes. Two resistance genes, *blmAB*, have been characterized previously, which confer BLM resistance to the producing organism by drug sequestering (BlmA) or modification (BlmB) [28–33]. In addition, BlmORF7 has also been proposed to be a member of the ATP-binding cassette (ABC) transporter family of proteins [29]. The latter proteins confer resistance to the producing organisms by transporting the drug out of the cells [61], although this function could not be proved for BlmORF7 by Calcutt and Schmidt [29]. The gene product of ORF29 is closely related to a family of transmembrane transporters, such as YchM (42% similarity and 31% identity) [62], suggesting that ORF29 could also be involved in BLM resistance by drug transporting. Multiple mechanisms for drug resistance are common for many antibiotic-producing organisms [56,63].

Antibiotic production in *Streptomyces* appears to be regulated in response to nutritional status and a variety of environmental conditions. Pathway specific regulatory genes have been identified in several of the known gene clusters [64]. At least, two genes that may be involved in regulation of BLM biosynthesis in *S. verticillus* ATCC15003 are identified. ORF15 shows a high degree of similarity to SyrP (52% similarity and 37% identity) that participates in a phosphorylation cascade controlling syringomycin production [65], and ORF30 resembles a family of metal-responsive repressors, such as SmtB (52% similarity and 40% identity) [66].

The remaining genes identified within the *blm* cluster are *orf8, 13, 18* and *28*. The gene product of *orf8* shows a high sequence homology to a family of oxidases, such as Yqe-R (48% similarity and 28% identity) [67]. The *orf18* product resembles known asparagine synthetases, such as AsnB (43% similarity and 38% identity) [68]. *orf28* encodes a protein similar to several phosphatidylserine decarboxylases, such as Mth1026 (45% similarity and 29% identity) [69]. It is not clear what role such a protein could play in BLM biosynthesis. Finally, no function could be predicted for *orf13*. The deduced product of *orf13* resembles MbtH (65% similarity and 42% identity), which is a protein with unknown function found in the mycobactin gene cluster [70].

Predictions of amino acid specificity of Blm NRPSs

Marahiel and co-workers have recently postulated the specificity-conferring codes for A domains of NRPS by comparing the sequence of the A domain of gramicidin synthetase A, PheA, with 160 other A domains of NRPS modules [18]. These codes, consisting of 10 residues that putatively form the amino acid binding pockets, provide a structural basis of substrate recognition in the A domains of NRPSs and could be used to predict from the primary sequence the specificity of biochemically uncharacterized A domains. Using these codes, we were able to predict the amino acid specificity for NRPS module 0 (Cys), 1 (Cys), 3 (Asn), 4 (Ser), 6 (Thr), and 9 (Asn) (Table 2), but not for NRPS module 2, 7, and 8, the codes of which do not match to any of those defined by Marahiel and co-workers. Since the amino acid precursors for BLM biosynthesis have been established early by feeding experiments, the remaining three amino acid residues are β-Ala, Ala, and His, whose codes were either not available (β-Ala and His) or poorly defined (Ala) in Marahiel and co-workers' original work [18]. To update the existing codes, we performed additional sequence analysis to define the code assignments for the β-Ala and His binding pockets and to refine the code assignments for the Ala binding pockets.

The same strategy used by Marahiel and co-workers was adopted in the current work (During the review of this manuscript, a similar approach to predict the substrate specificity of A domains of NRPSs was published by Townsend and co-workers [71]. Although the two groups use a slightly different set of amino acid residues to define the substrate binding pockets of A domains, both methods were based on the same X-ray structure of PheA [72] and yield consistent predictions.) The selectivity-conferring residues that constitute the codes were determined by comparing the primary amino acid sequences of the A domains of the Blm NRPS modules with those in the current GenBank database. NRPS-2 was found to be most related to the Ebony protein from *Drosophila* and the second module of FxbB from *Mycobacterium smegmatis*. Ebony is involved in the biosynthesis of β-alanyl-dopamine and has been proposed to activate β-Ala [73]. FxbB is involved in exochelin biosynthesis, also activating β-Ala [74,75]. On the basis of these three proteins, we now de-

fine a code assignment for the β-Ala binding pocket, predicting that NRPS-2 specifies for β-Ala (Table 2). It is noteworthy that the acidic residue Asp-235, which is conserved in all α-amino acid-activating codes, is changed into Val-235 in the newly defined code for β-Ala. Asp-235 and Lys-517 have been proposed to mediate key interactions with the α-amino group and the carboxyl group of the substrate, respectively, and are therefore essential for activating an α-amino acid [18]. The Asp235Val alteration in the β-Ala code is consistent with a β-amino acid substrate. Similar mutations at Asp-235 have been noted for a few other NRPSs activating substrates that do not have an α-amino group, such as AcvA for the ω-carboxyl group of α-aminoadipic acid, EntE for 2,3-dihydroxy benzoic acid, and YbtE for salicylic acid [18].

NRPS-8 was found to be closely related to the third module of BacC, which activates His [40]. Although this is the only additional domain known to activate His, we would like to put forward the code assignment for the His binding pocket, suggesting that NRPS-2 specifies for His (Table 2).

NRPS-7 was found to be most related to the second module of the McyA protein for microcystin biosynthesis from *Microcystis aeruginosa*, which was proposed to activate Ala [76], and the second module of the SafB protein, SafB-2, for saframycin biosynthesis from *Myxococcus xanthus*, which was proposed to activate Gly [77]. Since the only amino acid precursor common to BLM and saframycin biosynthesis is Ala, we questioned if the original assignments of amino acid specificity for each module of the SafAB proteins – SafB-1 for Ala, SafB-2 for Gly, SafA-1 for Tyr, and SafA-2 for Tyr – were incorrect. Close examination of the original work revealed that these assignments, in fact, were based mainly on the colinearity rule [77], rather than on rules for substrate recognition [18]. Since many exceptions to the colinearity rule have been noted recently [75,78,79], we re-examined the SafAB sequence and found that SafB-1 showed significant homology to NRPS-5 and a group of ALs, such as the AL of *M. bovis* (53% similarity and 38% identity) (Figure 3E) [41]. SafA-1 is closely related to the Gly-activating modules from Ta1 [9] and CdaPSII [80], from which a new code assignment for the Gly binding pocket is derived (Table 2). This leaves SafB-2 as the only candidate for Ala activation, despite the fact that it does not agree with the known code for Ala derived from two fungal NRPSs [18]. Taken together these data, we propose that both NRPS-7, McyA-2, and SafB-2 activate Ala, defining a second code assignment for the Ala binding pockets (Table 2). Multiple codes of A domains, analogous to codon degeneracy for protein biosynthesis, have been observed for several amino acids, such as Leu (four codes), Tyr (three codes), and Val (three codes) [18].

Table 2
Predictions of substrate specificity of Blm NRPSs based on the specificity-conferring codes of A domains (shown in bold) [18].

Domain	235	236	239	278	299	301	322	330	331	517	Similarity (%)[a]
L-Cys(2)	**D**	**L**	**Y**	**N**	**L**	**S**	**L**	**I**	**W**	**K**	
NRPS-0	P	L	Y	H	L	G	L	P	W	R	67
NRPS-1	D	L	Y	N	L	S	L	I	W	K	100
β-Ala[b]	**V**	**D**	**X**[b]	**V**	**I**	**S**	**X**[b]	**G**	**D**	**K**	
NRPS-2	V	D	W	V	I	S	L	A	D	K	70
Ebony	V	D	A	V	V	S	F	G	D	K	80
FxbB-2	V	D	T	L	I	S	S	G	D	K	70
L-Asn	**D**	**L**	**T**	**K**	**L**	**G**	**E**	**V**	**G**	**K**	
NRPS-3	D	L	T	K	V	G	E	V	G	K	90
NRPS-9	D	L	T	K	V	G	E	V	G	K	90
L-Ser	**D**	**V**	**W**	**H**	**L**	**S**	**L**	**I**	**D**	**K**	
NRPS-4	D	V	W	H	V	S	L	V	D	K	90
L-Thr	**D**	**F**	**W**	**N**	**I**	**G**	**M**	**V**	**H**	**K**	
NRPS-6	D	F	W	S	V	G	M	I	H	K	90
L-Ala[b]	**D**	**L**	**F**	**N**	**N**	**A**	**L**	**T**	**Y**	**K**	
NRPS-7	D	L	F	N	N	A	L	T	Y	K	100
McyA-2	D	L	F	N	N	A	L	T	Y	K	100
SafB-2	D	L	F	N	N	A	L	T	Y	K	100
L-His[b]	**D**	**S**	**X**[b]	**L**	**X**[b]	**A**	**E**	**V**	**X**[b]	**K**	
NRPS-8	D	S	A	L	I	A	E	V	W	K	70
BacC-3	D	S	E	L	T	A	E	V	C	K	70
Gly[b]	**D**	**I**	**L**	**Q**	**L**	**G**	**L**	**I**	**W**	**K**	
SafA-1	D	I	L	Q	L	G	L	V	W	K	100
Ta1	D	I	L	Q	L	G	M	I	W	K	100
CdaPSII-2	D	I	L	Q	L	G	L	I	W	K	100

[a]Similarity is calculated by BestFit in the GCG program [93].
[b]Amino acid specificity conferring codes of A domains defined in this work, and X indicates a variable amino acid within the determined code.

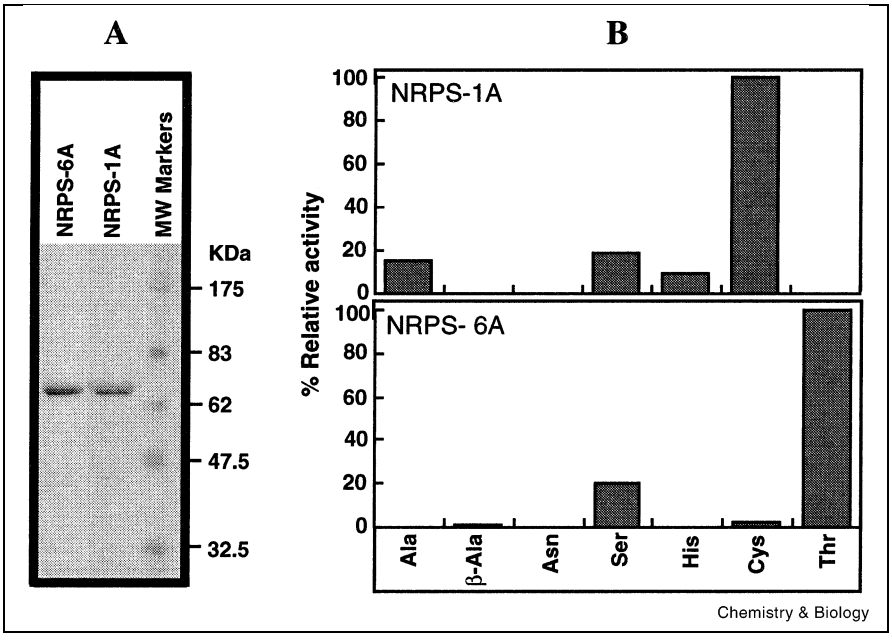

Figure 5. (A) Purified NRPS-1A and NRPS-6A after overexpression in *E. coli* as analyzed by electrophoresis on a 10% SDS-polyacrylamide gel (the calculated molecular weights for NRPS-1A and NRPS-6A are 64 212 and 61 899, respectively). **(B)** Substrate specificities as determined by the ATP–PPi exchange reaction with the amino acids known to incorporate into BLM as substrates (100% relative activity corresponds to 103 000 cpm for NRPS-1A and 256 000 cpm for NRPS-6A).

Specific adenylation of Cys by NRPS-1 and Thr by NRPS-6

While sequence analysis based on the specificity conferring codes of A domains in NRPSs allowed us to predict the substrates for all the Blm NRPS modules, accounting for all nine amino acids required for BLM biosynthesis, we set to verify these predictions by biochemical characterization of the NRPS modules. The latter strategy has been hampered by either the lack of expression of the desired genes or the poor solubility of the resultant proteins in heterologous hosts such as *E. coli* and *Streptomyces lividans*. However, we were successful in overexpressing the A domains from both the NRPS-1 and NRPS-6 modules in *E. coli*. The recombinant proteins were purified (Figure 5A), and their substrate specificities were examined according to the amino acid-dependent ATP–PPi assay [36]. NRPS-1A and NRPS-6A activate specifically L-Cys and L-Thr, respectively, among the amino acids tested (Figure 5B), providing experimental evidence for the sequence-based predictions.

Bioconversion of BLM aglycone into BLMs in *S. lividans*

S. verticillus ATCC15003 remains refractory to all means of introducing plasmid DNA into its cells, such as protoplast-mediated transformation, transformation by electroporation, and conjugation from *E. coli*, in spite of exhaustive effort by us [6] and other laboratories [29]. The lack of a genetic system for *S. verticillus* ATCC15003 has prevented us from carrying out genetic analysis of BLM biosynthesis in vivo. To provide additional experimental evidence to support that the cloned gene cluster encodes BLM biosynthesis, we turned to heterologous expression of the putative sugar biosynthesis genes in *S. lividans* to test their activities in bioconversion of the BLM aglycones into BLMs. We chose cosmid pBS9 because it harbors not only the desired sugar biosynthesis genes, *blmCDEFG*, but also the two BLM resistance genes, *blmAB*, as well as the putative ABC transporter ORF7. The latter are essential for the bioconversion experiment by providing BLM resistance to *S. lividans* [28]. In addition, we also cloned both the putative pathway regulator, *orf30*, and the transmembrane transporter, *orf29*, to yield pBS12. Co-introduction of both pBS9 and pBS12 into *S. lividans* ensures that the plasmid born genes are properly expressed in the heterologous host and that BLM will be exported out of *S. lividans* cells if produced (The other putative regulatory gene, *orf15*, is included by pBS9).

Conditions for BLM isolation, purification, and detection were first established using the wild-type organism. Fermentation of *S. verticillus* ATCC15003 under the standard conditions produces BLMs as a mixture of congeners, differing structurally at the C-terminal amines of the glycopeptides [24–27]. Figure 6B shows a typical profile of the partially purified BLMs after Amberlite XAD-16 column upon high performance liquid chromatography (HPLC) analysis. The identities of BLM A2 and B2 were confirmed by HPLC co-elution with authentic standards (Figure 6A) and electrospray ionization–mass spectrometry (ESI–MS) analysis of the purified BLMs. Upon ESI–MS analysis, BLM A2 showed a molecular ion peak at $m/z = 1414.2$, consistent with the M^+ ion for $C_{55}H_{84}N_{17}O_{21}S_3$, and BLM B2 showed a molecular ion peak at $m/z = 1425.5$, consistent with the $[M+H]^+$ ion for $C_{55}H_{84}N_{20}O_{21}S_2$. The ESI–MS results for the purified BLMs were further

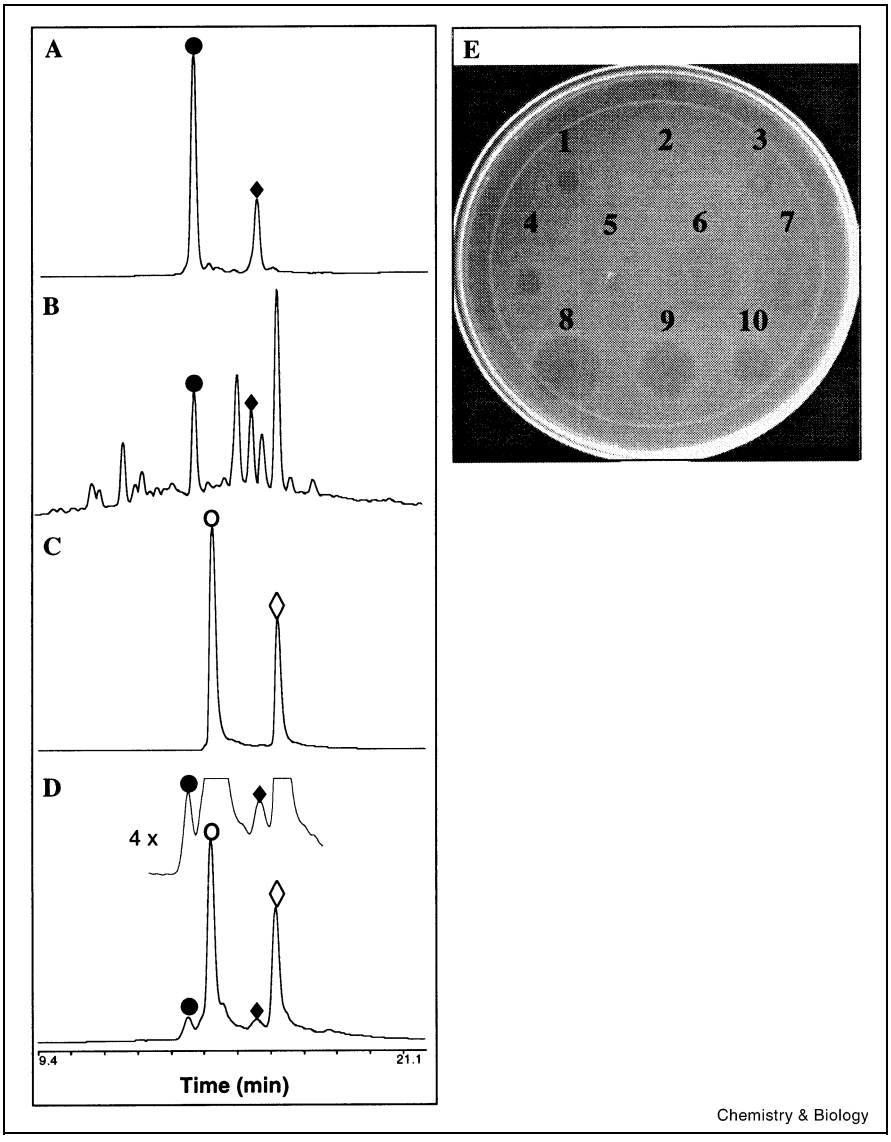

Figure 6. Bioconversion of BLM aglycones into BLMs in *S. lividans* 1326 (pBS9/pBS12) as analyzed by HPLC **(A–D)** and antibacterial activity assay **(E)**.
(A) Authentic samples of BLM A2 (●) and B2 (◆). **(B)** Partially purified BLMs isolated from *S. verticillus* ATCC15003.
(C) Authentic samples of BLM A2 aglycone (○) and B2 aglycone (◇). **(D)** BLM A2 and B2 isolated from *S. lividans* 1326 (pBS9/pBS12) cultured in the presence of exogenous BLM A2 and B2 aglycones.
(E) Bioassay using *B. subtilis* PB2 as a test organism: samples 1–3, 300, 30, 3 μg of BLM A2 and B2 aglycone standard mixture; sample 4, *S. lividans* 1326 control; sample 5, *S. lividans* 1326 (pBS9/pBS12) control cultured in the absence of exogenous BLM A2 and B2 aglycones; samples 6 and 7, BLM A2 and B2 purified, respectively, from *S. lividans* 1326 (pBS9/pBS12) cultured in the presence of exogenous BLM A2 and B2 aglycones; samples 8–10, 30, 3, 0.3 μg of BLM A2 and B2 standard mixture.

supported by assaying their antibacterial activity against *Bacillus subtilis* [31] (Figure 6E).

Both pBS9 and pBS12, as well as the vector pOJ446 as a control, were introduced into *S. lividans* 1326 by protoplast-mediated transformation [81]. The resulting apramycin and thiostrepton-resistant transformants *S. lividans* 1326 (pBS9/pBS12) were cultured in media supplemented with exogenously added aglycones of BLM A2 and B2 [82]. Transformants from the control, *S. lividans* 1326 (pOJ446), were cultured under identical conditions. Production of BLM A2 and B2 was analyzed by HPLC and bioassay. The isolated products from *S. lividans* 1326 (pBS9/pBS12) exhibited two pairs of distinctive peaks, whose identities as recovered BLM A2 and B2 aglycones and the newly synthesized BLM A2 and B2 were confirmed by HPLC co-elution with authentic standards (Figure 6D). Both the BLM A2 and B2 aglycones recovered and BLM A2 and B2 synthesized were purified by HPLC and assayed for their antibacterial activity against *B. subtilis* [28]. As shown in Figure 6E, while the BLM aglycones were totally inactive, clear inhibition zones were detected with both the purified BLM A2 and B2 samples.

Discussion

In this report, we have localized a 140-kb contiguous DNA region, flanking the known *blmAB* resistance genes, from *S. verticillus* ATCC15003, determined 85-kb of its sequence, and established that the cloned gene cluster encodes biosynthesis of the antitumor antibiotic BLMs. The latter conclusion is based on the following results: (1) The sequenced genes are clustered with two previously characterized BLM resistance genes, *blmAB*. (2) The numbers and predicted substrate specificities of NRPS and PKS

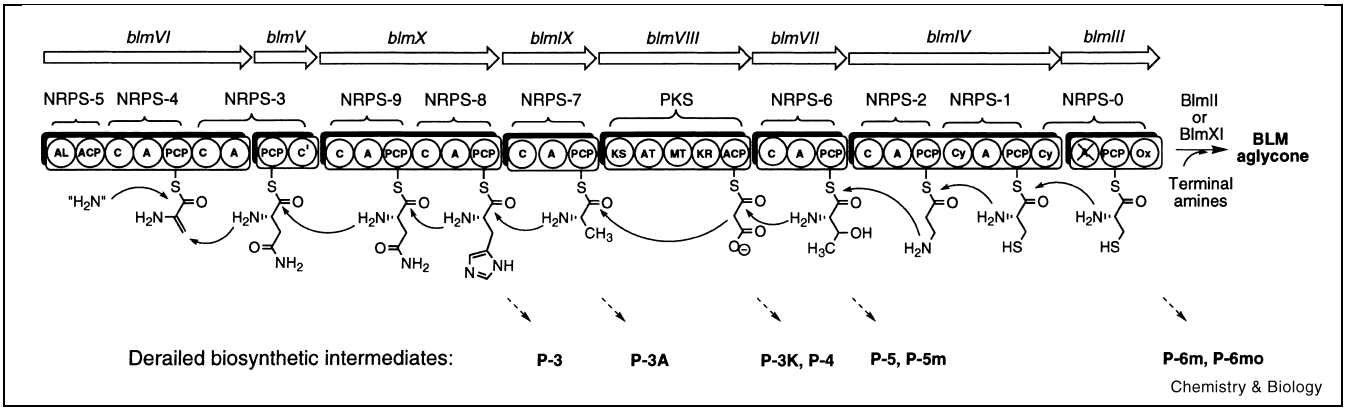

Figure 7. A linear model for the BLM megasynthetase-templated assembly of the BLM peptide/polyketide/peptide aglycone from nine amino acids and one acetate. Arrows with broken line indicate where biosynthetic intermediates were derailed. Three-letter amino acid designations were used. NRPS and PKS domains are abbreviated as follows: A, adenylation; ACP, acyl carrier protein; AL, acyl CoA ligase; AT, acyltransferase; C and C′, condensation; Cy, condensation/cyclization; KR, ketoreductase; KS, ketoacyl synthase; MT, methyltransferase; Ox, oxidation; PCP, peptidyl carrier protein.

modules agree with the structure of the BLM peptide–polyketide backbone. (3) The amino acid specificities of two NRPS modules were confirmed biochemically in vitro, and the domain organization of the BlmIII and BlmIV proteins, including the unusual Cy and Ox domains, concurs with the unique bithiazole moiety of BLMs. (4) The five identified sugar biosynthesis genes, with their deduced functions, correspond to the sugar moiety of BLMs. (5) Fermentation of an *S. lividans* strain expressing a part of the cloned gene cluster in the presence of exogenously added BLM aglycones resulted in the production of BLMs.

Most of the bacterial NRPS gene clusters characterized to date are organized in operon-type structures, encoding multimodular NRPS proteins with individual modules organized along the chromosome in a linear order that parallels the order of the amino acids in the resultant peptides, i.e., following the 'colinearity rule' for the NRPS-templated assembly of peptides from amino acids [1–5]. Inspection of the *blm* gene cluster (Figure 2B) showed that the Blm NRPS and PKS modules apparently are not organized according to the 'colinearity rule' for BLM biosynthesis (Figure 1B) (Exception to the 'colinearity rule' was also noted for the syringomycin synthetase from *Pseudomonas syringae* [78] and for the exochelin synthetase from *M. smegmatis* [75]. In fact, Grandi and co-workers have demonstrated recently in *B. subtilis* that neither the operon-type structure nor the physical linkage of individual modules is essential for proper assembly and activity of the surfactin NRPS megasynthetase [79].) Realizing that the BLM biosynthesis cannot be rationalized according to the 'colinearity rule', we predicted the substrate specificities of individual NRPS modules according to the specificity conferring codes for A domains [18] and deduced the extender unit of the PKS module according to the signature motifs of the AT domain [50]. Using the substrate specificity of individual NRPS and PKS modules as a guide, we were able to align the nine NRPS and one PKS modules to constitute the Blm megasynthetase (Figure 7) according to our hybrid NRPS–PKS–NRPS model for BLM biosynthesis (Figure 1B). Although its overall structure resembles that of known NRPSs and PKSs, the Blm megasynthetase exhibits several novel features. The latter is in fact expected since the BLM backbone contains several unusual structural features, such as the β-aminoalaninamide and the pyridimidine moieties, both of which are unprecedented in peptide biosynthesis.

On the basis of the BLM structure and the deduced functions of individual NRPS and PKS domains, we propose the following model for the Blm megasynthetase-templated synthesis of BLMs. The individual modules are first primed with amino acid or carboxylic acid precursors, the substrate specificities of which are determined by the A domains of the NRPS or the AT domain of PKS modules, respectively (Figure 7). We propose that the NRPS-4-activated Ser is further dehydrated into dehydroalanine (Dha) before its incorporation into the peptide product. Although dehydroamino acids, such as Dha, are enamines and are unlikely to be sufficiently stable, their stability could be improved when they are linked covalently, via a thioester bond, to the PCP domains and hence are sequestered by the protein. Dha or dehydrothreonine (Dht), also known as 2,3-dehydroaminobutyric acid, are present in many peptides, such as thiostrepton [83], microcystin [76], and syringomycin [78]. Although it is yet to be determined if the dehydration requires an additional enzyme, sequence analyses of both the NRPS-4 module and the NRPS modules for Dha in the microcystin cluster [76] and for Dht in the

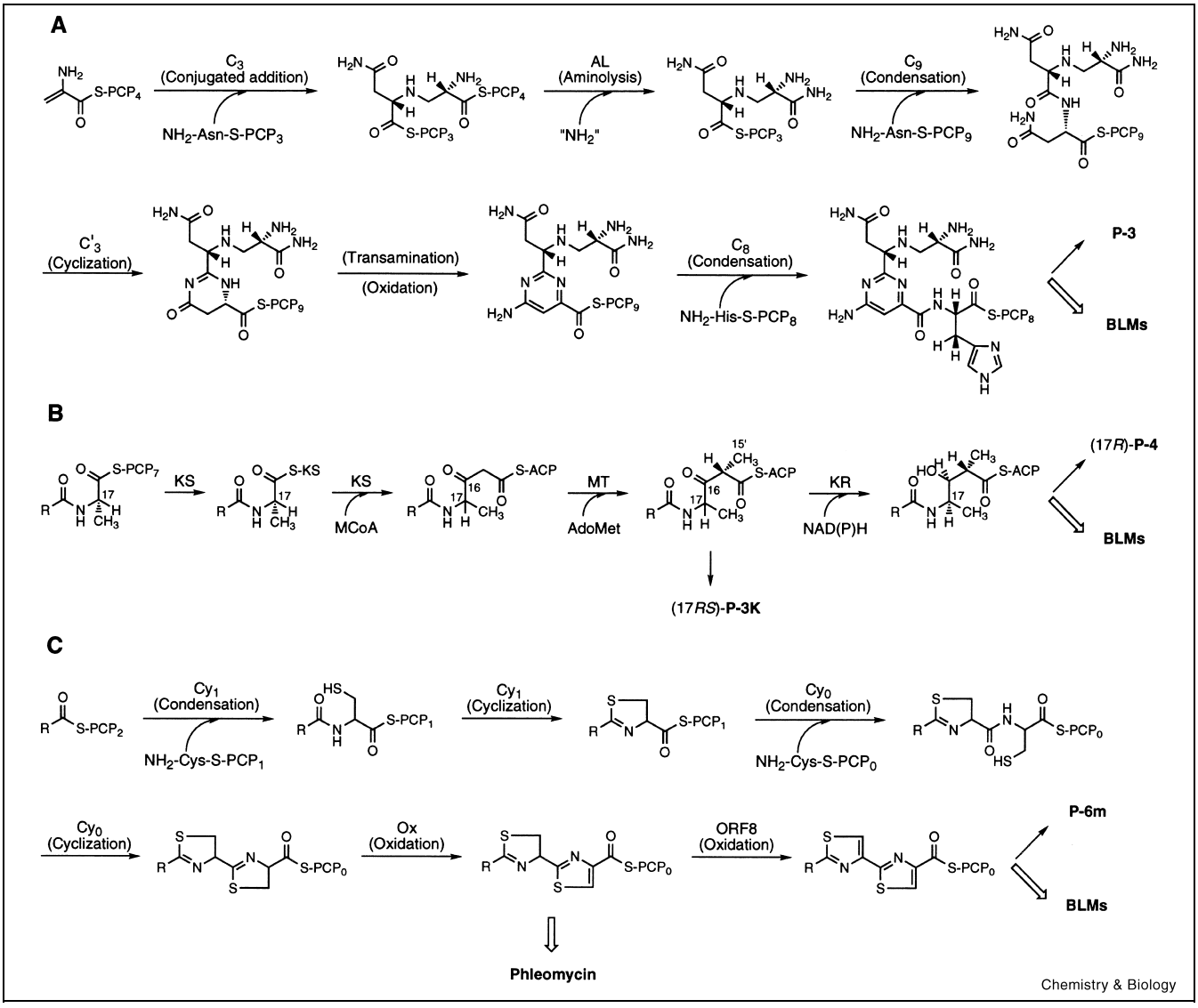

Figure 8. Proposed pathways for BLM megasynthetase-templated assembly of the BLM aglycone. **(A)** Early steps of BLM biosynthesis, supported by the isolation of P-3. **(B)** Steps for the biosynthesis of the polyketide moiety of BLM, supported by the isolation of (17*RS*)-P-3K and (17*R*)-P-4. **(C)** Steps for the biosynthesis of the bithiazole moiety of BLM, supported by the isolation of P-6m. Three-letter amino acid designations were used. NRPS and PKS domains are abbreviated as follows: A, adenylation; ACP, acyl carrier protein; AL, acyl CoA ligase; AT, acyl transferase; C and C′, condensation; Cy, condensation/cyclization; KR, ketoreductase; KS, ketoacyl synthase; MT, methyltransferase; Ox, oxidation; PCP, peptidyl carrier protein. A subscript number designates the Blm NRPS module, to which the domain belongs. MCoA, malonyl CoA; 'NH_2', unspecified ammonia donor.

syringomycin cluster [78] showed no differences to any other typical NRPS modules.

The synthesis of the BLM peptide–polyketide backbone could start by a conjugate addition between the Ser-derived Dha on NRPS-4 module and Asn on NRPS-3 module, followed by an aminolysis to cleave the Ser–Asn adduct off the Blm megasynthetase to furnish the unusual β-aminoalaninamide moiety (Figure 8A) (A conjugated addition to a Dha has also been proposed for the biosynthesis of the pyridine moiety of the peptide antibiotic GE2270A [83].) The former reaction could be catalyzed by the C domain of NRPS-3 that apparently is nonfunctional for normal transpeptidation due to the lack of both His residues in the conserved HHxxxDG active site (Figure 3C) [39], and the latter reaction could be catalyzed by the AL-like domain of the NRPS-5 starter module in a process that resembles the AL-catalyzed synthesis of acyl CoA from carboxylic acid but in the reverse direction in the presence of an amino donor (Figure 3E) [41]. Elongation resumes by

NRPS-9 and NRPS-8 modules sequentially to incorporate the second Asn residue and a His residue into BLM. The isolation of P-3 from fermentation broth suggests that modifications of the side chain of the second Asn residue to form the pyrimidine ring occur before the next elongation step, requiring minimally three additional steps – cyclization, transamination, and desaturation. While the extra C domain, C′ of NRPS-3, could serve as a candidate for the cyclase activity [39,46,70], both transamination and desaturation reactions are likely catalyzed by discrete enzymes that interact with the BLM megasynthetase cooperatively. On the basis of their predicted functions, ORF18 could be a candidate for catalyzing the transamination, whereas either ORF8 or BlmF could be a candidate for catalyzing the desaturation. Although it is unclear how such enzymes would fit into the current models for either NRPS or PKS, similar juxtaposition between either an NRPS or a PKS and discrete modifying enzymes have been suggested for syringomycin [78] and rifamycin biosyntheses [84], respectively.

Following the formation of the pyrimidine and incorporation of the His residue by NRPS-8 module, the peptide intermediate is further elongated by the NRPS-7 module to incorporate Ala into BLM, as evidenced by the isolation of P-3A. At this stage, the NRPS-7-bound growing peptide intermediate is transferred onto the KS domain of the BlmVIII PKS module, and elongation continues according to the chemistry of polyketide biosynthesis. Interestingly, the BlmVIII PKS module apparently furnishes the 'propionate' unit in two steps by evolving a malonyl CoA-specifying AT domain coupled with an AdoMet-requiring MT domain (Figure 8B). This domain organization of the BlmVIII PKS module agrees well with early feeding experiments that had clearly established that the polyketide moiety of BLM was derived from an acetate and a methionine [26,27]. Mechanistically C-methylation requires the formation of a carbanion at the α-carbon, followed by nucleophilic attack to the methyl group of AdoMet. Since the α-H is much more readily to be deprotonated in the β-ketoacyl-*S*-ACP than in the β-hydroxyacyl-*S*-ACP intermediate, one would predict that C-methylation occurs prior to β-keto reduction. The latter prediction is supported by the isolation of P-3K. It has been assumed that fungal PKSs in general contain MT domains for the introduction of a methyl branch into the polyketide products, as it has been shown recently in lovastatin biosynthesis [85]. In contrast, bacterial PKSs in general incorporate the methyl branch into the polyketide product by selecting a methyl malonyl-CoA extender unit. All PKSs from actinomycetes examined to date incorporate the alkyl branches into the resultant polyketides by selecting various alkyl malonates as the extending units that are determined by the AT domains [50]. However, a growing number of PKS genes has been discovered recently that derive the methyl branch of the polyketide products from AdoMet by a MT domain, such as the EPOS PKS for epothilone biosynthesis from *Sorangium cellulosum* [11,12] and the HMWP1 PKS for yersiniabactin biosynthesis from *Yersinia enterocolitica* and *Yersinia pestis* [8,46,47]. Intriguingly, the stereochemistry at C-17 appears to be epimerized during the PKS-mediated elongation cycle, as indicated by the *S*-configuration in P-3A and *R*-configuration in P-4 (Figure 1B). It could be imagined that the epimerization occurs with the β-ketoacyl-*S*-ACP intermediate, as evidenced by the isolation of P-3K as a pair of (17*RS*)-diastereomers. The KR domain then selectively reduces the (17*R*)-β-ketoacyl-*S*-ACP, yielding (17*R*)-β-hydroxyacyl-*S*-ACP, consistent with the isolation of (17*R*)-P-4 exclusively. In fact, similar stereoselective reduction of a racemic β-ketoacyl intermediate has been proposed for the KR domains of both the first and second modules of the 6-deoxyerythronolide B synthase [86,87].

The BlmVIII PKS bound growing intermediate resumes peptide elongation by the NRPS-6 module to incorporate the Thr residue into the BLM backbone, consistent with the isolation of P-5. Before additional peptide elongation by NRPS-2, NRPS-1, and NRPS-0 modules to finally yield the full length BLM peptide/polyketide/peptide backbone, the NRPS-6-bound growing intermediate apparently is modified by a C-methylation at the pyrimidine moiety, as evidenced by the isolation of P-5m. Results from early feeding experiments clearly established that this methyl group is derived from methionine, suggesting an AdoMet-dependent MT for this C-methylation reaction [26,27]. Since no additional AdoMet-dependent MT could be identified within the sequenced *blm* gene cluster, the MT domain of the BlmVIII PKS could play a role in the C-methylation of the pyrimidine ring. The latter proposal deviates from the 'processive' model known for both NRPS and PKS [1–5]. However, analogous juxtaposition of PKS activities for peptide modification has been proposed for yersiniabactin biosynthesis – reduction of the thiazoline ring synthesized by the HMWP2 NRPS was proposed to be catalyzed by the KR domain of the HMWP1 PKS [46,47].

The NRPS-2, NRPS-1, and NRPS-0 modules incorporate the β-Ala, Cys, and Cys residues, respectively, completing the assembly of the BLM backbone. In addition to typical NRPS domains, such as C, A, and PCP, for peptide elongation, these NRPS modules are characterized by the Cy (NRPS-1 and NRPS-0) [40] and Ox (NRPS-0) [11] domains as well as an inactive A domain (NRPS-0). The identification of the Cy and Ox domain in NRPS-1 and NRPS-0 modules agrees well with the unique bithiazole structure of BLMs. While thiazoline is the direct product of the Cy domain [40,46,47], the thiazoline-to-thiazole conversion requires an additional oxidation step [83]. Very recently, two biosynthesis gene clusters for thiazole-containing hybrid peptide–polyketide metabolites –

the epothilone gene cluster from *S. cellulosum* [11,12] and the myxothiazol gene cluster from *Stigmatella aurantiaca* [10] – were characterized, and an Ox domain was proposed by Molnar and co-workers for the biosynthesis of the thiazole moiety of epothilone [11]. The A domain of the NRPS-0 module exhibits poor consensus sequence at A1, A2, A6, A7, and A10 motifs and lacks completely the A3, A4, A5, and A8 motifs (Figure 3A). In addition, the two invariable residues, D in A4 and K in A10 that are proposed to be essential to interact with the α-amino group and the carboxyl group for activating an α-amino acid [18], are mutated into P in A4 and R in A10 for the NRPS-0 module, respectively. These features strongly suggests that the NRPS-0 A domain is inactive and that the loading of the NRPS-0 PCP domain is possibly catalyzed by the A domain of the upstream NRPS-1 module. Inactive domains within PKS modules are known, such as the enoylreductase domain of EPOS A and EPOS E [11] – also known as EpoA and EpoF [12]. However, we are not aware of any inactive domain within an NRPS module with the possible exception of a whole module in the exochelin gene cluster [74,75]. Loading of multiple PCPs by a single A domain has been proved for yersiniabactin biosynthesis [46,47].

Taken together the aforementioned structural features of the NRPS-2, -1 and -0 modules and precedence of thiazole biosynthesis in the literature, a biosynthetic pathway for the bithiazole moiety of BLMs could be proposed as shown in Figure 8C. While it cannot be excluded in the absence of any experimental data that the single Ox domain in NRPS-0 module is responsible for the oxidation of both thiazole rings, we prefer the proposed processive mechanism of Cys incorporation, cyclization, and oxidation for bithiazole biosynthesis, consistent with the lack of an Ox domain in the NRPS-1 module. The latter hypothesis is supported by the occurrence of two Ox domains for bithiazole biosynthesis observed from the myxothiazol gene cluster [10] and calls for an additional oxidase for the formation of the first thiazole ring in BLM biosynthesis. A possible candidate for this activity would be ORF8 that is highly homologous to coproporphyrinogen III oxidases [67]. In fact, a cell-free preparation from *S. verticillus* ATCC15003 has been reported to catalyze the conversion of phleomycins to BLMs in the presence of NAD^+ [27], supporting the presence of a discrete oxidase. We could imagine that the phleomycin producer results from a simple mutation at the latter oxidase gene, losing its ability to oxidize the first thiazoline ring.

Before the final release of the BLM peptide–polyketide backbone from the BLM megasynthetase, the β-position of the His residue is hydroxylated, providing the glycosidic linkage for the two sugars found in BLMs. Both ORF1, suggested early by Calcutt and Schmidt in characterizing the *blmAB* resistance locus [29] and BlmF show significant sequence homology to amino acid β-hydroxylases, serving as candidates for catalyzing this hydroxylation. The isolation of both P-6m and P-6mo in the form of free carboxylic acid suggests that β-hydroxylation of the His residue takes place after the complete assembly of the BLM peptide–polyketide backbone but before its release from the BLM megasynthetase upon nucleophilic attack by the terminal amines.

Most of the NRPSs and PKSs studied to date release the full length peptide or polyketide intermediate from enzyme complexes by the catalysis of a thioesterase (TE) domain that is usually found at the distal C-terminus of the NRPS or PKS protein [1–3]. The BLMs are characterized by an unique terminal amide moiety, differing structurally from other well characterized peptide or polyketide metabolites. Myxothiazol [10] and rifamycin [88] are the only two other peptide and polyketide metabolites with a terminal amide structure, whose biosynthetic gene clusters have been characterized. Although a TE domain is identified at the distal C-terminus of the proposed Mta NRPS–PKS hybrid enzyme complex, the amino group of myxothiazol terminal amide is proposed to be derived from an amino acid by the action of MtaG, an NRPS module with an unprecedented monooxygenase domain [10]. The latter domain catalyzes the oxidative cleavage at the α-position of the last amino acid of the full length peptide–polyketide intermediate to afford the terminal amide of myxothiazol, and this mechanism is proposed to be a general strategy for the formation of terminal amides in other peptide or polyketide metabolites [10]. In contrast, the Rif PKS gene clearly lacks a TE domain. Associated with the Rif PKS, instead, is a discrete protein RifF, an amide synthase [88]. The latter enzyme was proposed to catalyze the release of the complete polyketide from the PKS by intramolecular amide formation to generate the macrocyclic lactam structure in rifamycin biosynthesis, representing another mechanism for the amide formation in peptide and polyketide biosynthesis [88]. The formation of the amide moiety of BLMs is apparently distinct from both the typical peptide or polyketide biosynthesis and myxothiazol or rifamycin biosynthesis. Neither TE domain, unusual NRPS module with a monooxygenase domain, nor discrete amide synthase has been identified within the sequenced *blm* gene cluster. These findings, in fact, are consistent with early feeding experiments that clearly established that various amines can be incorporated into the C-terminal amine moiety of BLMs [24–27], excluding the amino acid origin of the terminal amino group, as has been proposed for myxothiazol biosynthesis [10]. In addition, over 40 different amines have been incorporated into BLMs by directed biosynthesis [24–27], suggesting that the enzyme activity responsible for the amide formation must have exhibited a very broad substrate specificity for the amine. Interestingly, flanking the modular *blm* NRPS and PKS genes are two *orfs*, *blmXI* and *blmII*, whose deduced products are highly

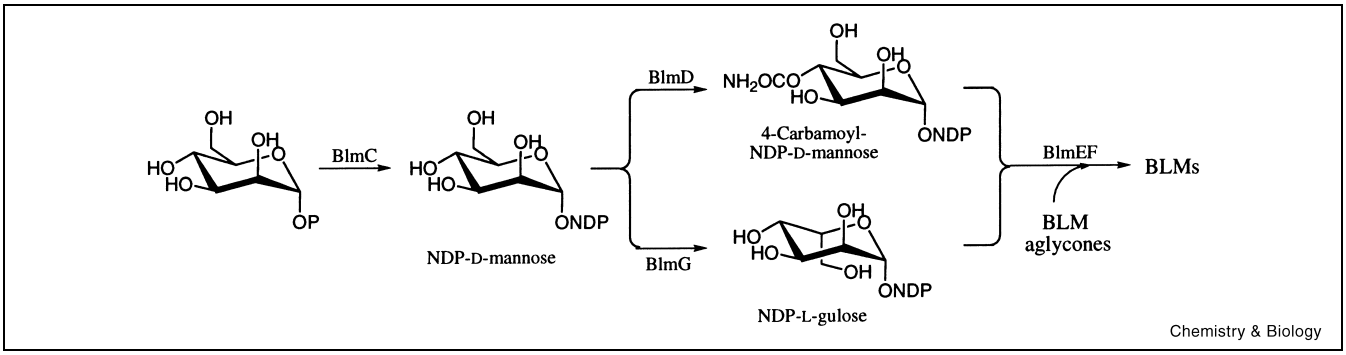

Figure 9. Proposed biosynthetic pathway for the two sugars of BLM from D-mannose.

homologous to the C domain of modular NRPS (Figure 3C). On the basis of the analogous chemistry, these discrete condensation enzymes could serve as candidates catalyzing the release of the full length BLM peptide–polyketide backbone from the BLM megasynthetase complex by nucleophilic attack of the RCO-*S*-PCP species with an amine to generate the BLM amide structure, providing yet another mechanism for amide formation in peptide and polyketide biosynthesis. C domain-catalyzed release of either a RCO-*S*-PCP from NRPS or RCO-*S*-ACP from PKS has, in fact, been proposed for rapamycin [7] and lovastatin [85] biosynthesis, respectively.

The attachment of the two unusual sugars – L-gulose and 3-*O*-carbamoyl-D-mannose – to the BLM aglycones finally completes the biosynthesis of BLMs. On the basis of the sequenced sugar biosynthesis genes and their deduced functions, the following pathway could be proposed for the biosyntheses of both sugars from D-mannose-phosphate (Figure 9). The exact order for individual reactions to make these modified sugars and for glycosylations to attach the sugars to the BLM aglycones has to be determined by future experiments.

Among all the NRPSs, PKSs, and hybrid NRPS–PKS systems examined so far, the Blm megasynthetase consists of the largest number of individual proteins. The precise interactions among all the Blm NRPS and Blm PKS proteins to constitute the Blm megasynthetase complex, as well as other discrete proteins juxtaposed with the BLM megasynthetase complex to modify the growing BLM intermediates, therefore, reflect a remarkable power of protein–protein recognition. Although we are yet to provide direct evidence supporting the specific protein–protein interactions between the neighboring proteins, it is striking to note that all the biosynthetic intermediates isolated are derailed from either NRPS or PKS modules at the junctions between the interacting proteins (Figure 7). Since it is not difficult to imagine that an intermediate is more likely to fall off the enzyme complex when it is subjected to interpeptide transfer than to intrapeptide transfer, we view the latter observation as a strong evidence supporting the current model of the Blm megasynthetase.

The biosynthesis of natural products of hybrid peptide–polyketide origin provides an excellent opportunity to investigate intermodular communication between NRPS and PKS modules and to elucidate the mechanism of constructing hybrid NRPS–PKS systems for expanding the size and diversity of 'unnatural' natural product libraries [2,3,6]. In addition to the *blm* cluster described in this report, a growing number of gene clusters for hybrid peptide–polyketide biosynthesis have been characterized recently, which includes the EPOS synthetase [11] – also known as Epo synthetase [12] – for epothilones from *S. cellulosum*, the Mta synthetase for myxothiazol from *S. aurantiaca* [10], the Rap synthetase from *Streptomyces hygroscopicus* [7,89], the Ta1 synthetase for antibiotic TA from *M. xanthus* [9], and the HMWPs for yersiniabactin from *Y. enterocolitica* and *Y. pestis* [8,46,47]. Detailed comparison of individual domains or modules between hybrid and non-hybrid systems could shed light on the mechanism for the altered catalytic activities and intermodular communication between the interacting NRPS and PKS modules to constitute hybrid enzymes.

The BlmIX/BlmVIII/BlmVII system combines the features of both hybrid NRPS/PKS and PKS/NRPS systems. The fact that both the BlmIX and BlmVII NRPS modules and the BlmVIII PKS module are three separate proteins with a typical domain organization for NRPS and PKS enzymes greatly simplifies the mechanistic analysis of the hybrid NRPS/PKS/NRPS complex. We have found that the KS domain of BlmVIII is more similar to the KSs known to interact with NRPS modules in hybrid NRPS/PKS systems than to KSs of type I PKSs. We attribute these subtle differences to their unique reactivity that catalyzes the transfer of the peptidyl intermediate from the PCP to the KS domain, which presumably takes place prior to chain elongation. Subsequent condensation catalyzed by

the KS domain between the peptidyl intermediate and malonyl-*S*-ACP results in the elongation of the growing peptide with a carboxylic acid. Equally striking are the discoveries that the ACP domain of BlmVIII is more similar to a PCP than to an ACP and that the C domain of BlmVII has an additional N-terminal segment of about 50 amino acids that is rich in arginine, aspartic acid, and glutamic acid. The latter feature is analogous to the N-terminal interpolypeptide linker for type I PKS, which has recently been demonstrated to play a critical role in intermodular communication [20]. These unique features of the ACP domain from the BlmVIII PKS module and the C domain from the BlmVII NRPS module could provide the molecular basis for the C domain to recognize the acyl-*S*-ACP as a substrate. Subsequent condensation catalyzed by the C domain between acyl-*S*-ACP and amino acyl-*S*-PCP results in the elongation of the growing polyketide (as far as this condensation is concerned) with an amino acid. Although answers to many of these questions may ultimately have to come from in vitro biochemical investigations, the availability of the *blm* gene cluster has now set the stage for such future studies.

Significance

The anticancer antibiotic BLM has been incorporated into chemotherapy of several malignancies under the trade name of Blenoxane® since the late 1970s. However, wide application of this drug has been prevented by early development of drug resistance and dose-limiting pulmonary toxicity. Attempts at analog development to circumvent these limitations have been hampered by the structural complexity of the naturally available BLMs. Complementary to making microbial metabolites and their structural analogs by chemical synthesis, genetic manipulation of genes governing secondary metabolism offers a promising alternative to prepare these compounds biosynthetically. This work represents a complete characterization of the gene cluster for BLM production in *S. verticillus* ATCC15003, which consists of 10 NRPS genes, one PKS genes, five sugar biosynthesis genes, as well as genes encoding other biosynthesis, resistance, and regulatory proteins. The availability of the *blm* gene cluster should facilitate future attempts, using combinatorial biosynthesis and rational metabolic pathway engineering, to produce novel BLM analogs with improved therapeutic efficacy and lower toxicity as antitumor agents.

The structural and catalytic similarities between modular NRPS and PKS inspired us to search for a hybrid NRPS–PKS system. The fact that individual domains and modules of both NRPS and PKS are considerably tolerant towards genetic engineering supports the wisdom of combining individual NRPS and PKS modules for combinatorial biosynthesis. A great challenge in constructing a hybrid NRPS–PKS system lies at revealing the molecular basis for intermodular communication between NRPS and PKS. This issue can now be investigated in detail through experiments using the *blm* NRPS and PKS genes since they constitute a natural hybrid NRPS–PKS. In addition, characterization of the BLM megasynthetase should also enable us to define the mechanism for thiazole biosynthesis, potentially leading to the making of novel thiazole-containing molecules by engineering peptide biosynthesis.

Materials and methods

General procedures

E. coli DH5α [90], *E. coli* XL 1-Blue MR (Stratagene, La Jolla, CA, USA), *E. coli* BL21(DE-3) (Novagen, Madison, WI, USA), *B. subtilis* PB2 (Prof. Chet Price, University of California, Davis, CA, USA), *S. lividans* 1326 [81], and *S. verticillus* ATCC15003 (American Type Culture Collection, Rockville, MD, USA) were used in this work. pWHM3 [91] and pOJ446 [34] were described previously, and pQE60 (Qiagen, Santa Clarita, CA, USA), pET28a and pET29a (Novagen) were from commercial sources. BLM A2 and B2 were gifts from Prof. Claude F. Meares, University of California, Davis. The BLM A2 and B2 aglycones were prepared from BLM A2 and B2 according to a literature procedure [82].

Plasmid preparation was carried out by using commercial kits (Qiagen). Total *S. verticillus* ATCC15003 DNA was isolated according to literature protocols [81,92]. Restriction enzymes and other molecular biology reagents were from commercial sources, and digestion and ligation followed standard methods [90]. For Southern analysis, digoxigenin labelling of DNA probes, hybridization, and detection were performed according to the protocols provided by the manufacturer (Boehringer Mannheim Biochemicals, Indianapolis, IN, USA).

Automated DNA sequencing was carried out on an ABI Prism 377 DNA sequencer (Perkin-Elmer/ABI, Foster City, CA, USA), and this service was provided by either the DBS Automated DNA Sequencing Facility, University of California, Davis, or Davis Sequencing Inc. (Davis, CA, USA). Data were analyzed by the ABI Prism Sequencing 2.1.1 software and the Genetics Computer Group (GCG) program (Madison, WI, USA) [93].

Cloning and sequencing of the blm gene cluster

A genomic library of *S. verticillus* ATCC15003 was constructed in pOJ446 according to literature procedures [92] and screened with probes made from both ends of the *blmAB* locus [29], leading to the localization of 140-kb contiguous DNA, of which 100 kb is upstream and 40 kb is downstream of the *blmAB* genes (Figure 2A). Heterologous NRPS probes, amplified from *S. verticillus* ATCC15003 by polymerase chain reaction (PCR) according to literature procedures, were described previously and were used to screen the entire 140-kb DNA by Southern analysis under various hybridization conditions [6]. The DNA sequence reported in this work was determined from pBS5, pBS6, and pBS8.

Overproduction and biochemical characterization of the NRPS-1A and NRPS-6A proteins

Heterologous expression of the A domains in *E. coli* were performed according to literature procedures [36,94]. NRPS-1A (forward primer 5′-AACCCATGGCTGCTTCCCTGACCCGCCTGGCC-3′ (the *Nco*I site is underlined) and reverse primer 5′-CCTAGATCTACGGGCAGGTGGGGCGGT-3′ (the *Bgl*II site is underlined)) and NRPS-6A (forward primer 5′-GGGAATTCCATATGATCCTCACGTCCTTCCAC-3′ (the *Nde*I site is underlined) and reverse primer 5′-GGCAAGCTTGGGTGAGGGTCCGTTCGGT-3′ (the *Hin*dIII site is underlined)) were amplified by PCR from *S. verticillus* ATCC15003 cosmid clones. The resulting 1.6-kb fragment of NRPS-1A was first cloned into the *Nco*I–*Bgl*II sites of pQE60 and then moved as an *Nco*I–*Hin*dIII fragment into the similar sites of pET29a

to yield pBS13, and the resulting 1.6-kb fragment of NRPS-6A was directly cloned into the *Nde*I–*Hin*dIII sites of pET28a to yield pBS14. Introduction of pBS13 and pBS14 into *E. coli* BL21(DE-3) under standard expression conditions resulted in production of NRPS-1A (with an N-terminal S-tag and a C-terminal His_6-tag) and NRPS-6A (with an N-terminal His_6-tag), respectively. The soluble fractions of fusion proteins were subjected sequentially to an affinity chromatography on Ni–NTA resin and an anion exchange chromatography on a Hyper-D column (PerSeptive Biosystem, Framingham, MA, USA), resulting in NRPS-1A and NRPS-6A with near homogeneity. Amino acid-dependent ATP–PPi assays were performed essentially according to the literature procedures [36].

Production and determination of BLM production

Spore stock of *S. verticillus* ATCC15003 was inoculated to the seed medium (25 ml), consisting of 1% glucose, 1% starch, 0.75% peptone, 0.75% beef extract, and 0.3% NaCl, pH 7.0, and incubated at 30°C, 300 rpm for 38 h. The seed inoculum (5 ml) was then added to the production medium (50 ml), consisting of 6.4% milet jelly, 0.5% glucose, 3.5% soybean meal, 0.75% cornsteep liquor, 0.2% $NaNO_3$, 0.3% NaCl, 0.1% K_2HPO_4, 0.05% $ZnSO_4{\cdot}7H_2O$, 0.01% $CuSO_4{\cdot}5H_2O$, and incubated at 30°C, 300 rpm for 6 days [24–27].

The fermentation broth was extracted with EtOAc, and the aqueous phase, which contains BLMs, was loaded to an Amberlite XD-16 column. After extensive wash with H_2O, BLMs were eluted with MeOH–0.01 N HCl (4:1, vol/vol). The elute was concentrated in vacuo, re-dissolved in 50% aqueous MeOH, and subjected to HPLC analysis on an Altima C-18 column (5 μm, 250 × 4.6 mm, Alltech Associates, Deerfield, IL, USA). The column was equilibrated with 1% NH_4OAc buffer, pH 5.7, with 20% MeOH (vol/vol) and developed with a 20-min linear gradient from 20 to 70% MeOH in the same NH_4OAc buffer, at a flow rate of 0.8 ml/min with UV–Vis detection at 300 nm on a Dynamax gradient HPLC system (Rainin, Walnut Creek, CA, USA).

ESI–MS determination was performed on an LCQ-Decca mass spectrometer (ThermoQuest, San Jose, CA, USA) at the Cancer Center, University of California, Davis. To assay the antibacterial activity of BLMs, *B. subtilis* PB2 was used as the test organism [28]. *B. subtilis* PB2 was grown overnight at 37°C in TSB [81], and the resulting culture was diluted (1:100) in 50% glycerol and stored in aliquots at −20°C. For bioassay, liquid samples (150 μl) were added to stainless steel cylinders placed on agar plates of TSB at half nutrient concentration pre-seeded with the *B. subtilis* PB2 glycerol stock (0.15% vol/vol) [95]. The plates were incubated at 37°C overnight to determine sizes of the inhibition zones.

Bioconversion of BLM aglycones into BLMs in S. lividans

A 3.4-kb *Bam*HI fragment, containing the putative pathway regulator *orf30* and transmembrane transporter *orf29*, was cloned into the same sites of pWHM3 to yield pBS12. Both pBS9 and pBS12 were introduced into *S. lividans* 1326 by protoplast-mediated co-transformation, with both apramycin (25 μg/ml) and thiostrepton (5 μg/ml) selection [81]. pOJ446 was similarly introduced into *S. lividans* 1326 as a control, selected with apramycin. The resultant transformants, *S. lividans* 1326 (pBS9/pBS12) and *S. lividans* 1326 (pOJ446), were grown in R2YE medium [81] containing appropriate antibiotics, respectively, and incubated at 30°C, 300 rpm for 3 days. A sterile solution of the BLM aglycones was then added (final concentration 33–160 μg/ml), and the resulting culture was incubated for additional 7 days. Isolation, purification, and identification of both the BLM aglycones and BLMs from *S. lividans* 1326 (pBS9/pBS12) and *S. lividans* 1326 (pOJ446) were carried out similarly as described for *S. verticillus* ATCC15003.

Accession numbers

The entire sequence of the gene cluster for BLM production in *S. verticillus* ATCC15003 has been deposited into GenBank under accession numbers AF149091 and AF210249.

Acknowledgements

We thank Dr. M. Calcutt and Prof. F. Schmidt, University of Missouri–Columbia, for clones of the *blmAB* locus, Dr. T. O'Hare, Oregon State University, for advice on constructing genomic library in pOJ446, Dr. J. Wu at the Cancer Center, University of California, Davis, for EIS–MS analysis, and Prof. C.R. Hutchinson, University of Wisconsin, Madison, for helpful discussion and constant encouragement. This work was supported in part by an Institutional Research Grant from the American Cancer Society and the School of Medicine, University of California, Davis; the National Institutes of Health Grant AI40475; and the Searle Scholars Program/The Chicago Community Trust. C.S. was supported in part by a postdoctoral fellowship from the Spanish Ministry of Education and Culture.

References

1. Cane, D.E. (1997). A special thematic issue on polyketide and nonribosomal polypeptide biosynthesis. *Chem. Rev.* **97**, 2463–2706.
2. Cane, D.E., Walsh, C.T. & Khosla, C. (1998). Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**, 63–68.
3. Cane, D.E. & Walsh, C.T. (1999). The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chem. Biol.* **6**, R319–R325.
4. Konz, D. & Marahiel, M.A. (1999). How do peptide synthetases generate structural diversity? *Chem. Biol.* **6**, R39–R48.
5. von Döhren, H., Dieckmann, R. & Pavela-Vrancic, M. (1999). The nonribosomal code. *Chem. Biol.* **6**, R273–R279.
6. Shen, B., Du, L., Sanchez, C., Chen, M. & Edwards, D.J. (1999). Bleomycin biosynthesis in *Streptomyces verticillus* ATCC15003: a model of hybrid peptide and polyketide biosynthesis. *Bioorg. Chem.* **27**, 155–171.
7. König, A., et al., & Leadlay, P.F. (1997). The pipecolate-incorporating enzyme for the biosynthesis of the immunosuppressant rapamycin-nucleotide sequence analysis, disruption and heterologous expression of *rapP* from *Streptomyces hygroscopicus*. *Eur. J. Biochem.* **247**, 526–534.
8. Pelludat, C., Rakin, A., Jacobi, C.A., Schubert, S. & Heesemann, J. (1998). The yersiniabactin biosynthetic gene cluster of *Yersinia enterocolitica*: organization and siderophore-dependent regulation. *J. Bacteriol.* **180**, 538–546.
9. Paitan, Y., Alon, G., Orr, E., Ron, E.Z. & Rosenberg, E. (1999). The first gene in the biosynthesis of the polyketide antibiotic TA of *Myxococcus xanthus* codes for a unique PKS module coupled to a peptide synthetase. *J. Mol. Biol.* **286**, 465–474.
10. Silakowski, B., et al., & Muller, R. (1999). New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.* **274**, 37391–37399.
11. Molnar, I., et al., & Ligon, J.M. (1999). The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium celluosum* So ce90. *Chem. Biol.* **7**, 97–109.
12. Tang, L., et al., & Julien, B. (2000). Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640–642.
13. Stachelhaus, T., Schneider, A. & Marahiel, M.A. (1995). Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science* **269**, 69–72.
14. de Ferra, F., Rodriguez, F., Tortora, O., Tosi, C. & Grandi, G. (1997). Engineering of peptide synthetases. *J. Biol. Chem.* **272**, 25304–25309.
15. Elsner, A., Engert, H., Saenger, W., Hamoen, L., Venema, G. & Bernhard, F. (1997). Substrate specificity of hybrid modules from peptide synthetases. *J. Biol. Chem.* **272**, 4814–4819.
16. Stachelhaus, T., Schneider, A. & Marahiel, M.A. (1998). Targeted alteration of the substrate specificity of peptide synthetase by rational module swapping. *Mol. Gen. Genet.* **257**, 308–318.
17. Belshaw, P.J., Walsh, C.T. & Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**, 486–489.
18. Stachelhaus, T., Mootz, H.D. & Marahiel, M.A. (1999). The spec-

ificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505.
19. McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M. & Ashley, G. (1999). Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel 'unnatural' natural products. *Proc. Natl. Acad. Sci. USA* **96**, 1846–1851.
20. Gokhale, R.S., Tsuji, S.Y., Cane, D.E. & Khosla, C. (1999). Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482–485.
21. Xue, Q., Ashley, G., Hutchinson, C.R. & Santi, D.V. (1999). A multiplasmid approach to preparing large libraries of polyketides. *Proc. Natl. Acad. Sci. USA* **96**, 11740–11745.
22. Aso, M., Kondo, M., Suemune, H. & Hecht, S.M. (1999). Chemistry of the bleomycin-induced alkali-labile DNA lesion. *J. Am. Chem. Soc.* **121**, 9023–9033, and references cited therein.
23. Sikic, B.I., Rosenzweig, M. & Carter, S.K. (1985) *Bleomycin chemotherapy*, Academic Press, New York.
24. Fujii, A. (1979). Biosynthetic aspects of bleomycin–phleomycin group antibiotics. In *Bleomycin: Chemical, Biochemical, and Biological Aspects*. (Hecht, S.M., Ed.), pp. 75–91, Springer-Verlag, New York.
25. Umezawa H. (1980). Recent progress in bleomycin studies. In *Anticancer agents based on natural product models*. (Cassady, J.N. and Douros, J.D., Eds.), pp. 147–166, Academic Press, New York.
26. Takita, T. (1984). The bleomycins: properties, biosynthesis, and fermentation. *Drugs Pharm. Sci.* **22**, 595–603.
27. Takita, T. & Muroka, Y. (1990). Biosynthesis and chemical synthesis of bleomycin. In *Biochemistry of Peptide Antibiotics: Recent Advances in the Biotechnology of β-Lactams and Microbial Peptides*. (Kleinkauf, H. and von Dohren, H., Eds.), pp. 289–309, W. de Gruyter, New York.
28. Sugiyama, M., et al., & Davies, J. (1994). Characterization by molecular cloning of two genes from *Streptomyces verticillus* encoding resistance to bleomycin. *Gene* **151**, 11–16.
29. Calcutt, M.J. & Schmidt, F.J. (1994). Gene organization in the bleomycin-resistance region of the producer organism *Streptomyces verticillus*. *Gene* **151**, 17–21.
30. Kumagai, T., Hibino, R., Kawano, Y. & Sugiyama, M. (1999). Mutation of the *N*-terminal proline 9 of BLMA from *Streptomyces verticillus* abolishes the binding affinity for bleomycin. *FEBS Lett.* **450**, 227–230.
31. Kawano, Y., Kumagai, T., Muta, K., Matoba, Y., Davies, J. & Sugiyama, M. (2000). The 1.5 Å crystal structure of a bleomycin resistance determinant from bleomycin-producing *Streptomyces verticillus*. *J. Mol. Biol.* **295**, 915–925.
32. Sugiyama, M., Kumagai, T., Shionoya, M., Kimura, E. & Davies, J. (1994). Inactivation of bleomycin by an *N*-acetyltransferase in the bleomycin-producing strain *Streptomyces verticillus*. *FEMS Microbiol. Lett.* **121**, 81–86.
33. Matsuo, H., Mochizuki, H., Davies, J. & Sugiyama, M. (1997). Production of bleomycin *N*-acetyltransferase in *Escherichia coli* and *Streptomyces verticillus*. *FEMS Microbiol. Lett.* **153**, 83–88; erratum, (1998). *FEMS Microbiol. Lett.* **158**, 153.
34. Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Rao, R.N. & Schoner, B.E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli to Streptomyces spp*. *Gene* **116**, 43–49.
35. Wright, F. & Bibb, M.J. (1992). Codon usage in the G+C-rich *Streptomyces* genome. *Gene* **113**, 55–65.
36. Du, L. & Shen, B. (1999). Identification and characterization of a type II peptidyl carrier protein from the bleomycin producer *Streptomyces verticillus* ATCC15003. *Chem. Biol.* **6**, 507–517.
37. Stachelhaus, T. & Marahiel, M.A. (1995). Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *J. Biol. Chem.* **270**, 6163–6169.
38. Stachelhaus, T., Hüser, A. & Marahiel, M.A. (1996). Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem. Biol.* **3**, 913–921.
39. Stachelhaus, T., Mootz, H.D., Bergendahl, V. & Marahiel, M.A. (1998). Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J. Biol. Chem.* **273**, 22773–22781.
40. Konz, D., Klens, A., Schörgendörfer, K. & Marahiel, M.A. (1997). The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem. Biol.* **4**, 927–937.
41. Fitzmaurice, A.M. & Kolattukudy, P.E. (1997). Open reading frame 3, which is adjacent to the mycocerosic acid synthase gene, is expressed as an acyl coenzyme A synthase in *Mycobacterium bovis* BCG. *J. Bacteriol.* **179**, 2608–2615.
42. Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G. & Lindqvist, Y. (1998). Crystal structure of β-ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzymes. *EMBO J.* **17**, 1183–1191.
43. Moche, M., Schneider, G., Edwards, P., Dehesh, K. & Lindqvist, Y. (1999). Structure of the complex between the antibiotic cerulenin and its target, β-ketoacyl-acyl carrier protein synthase. *J. Biol. Chem.* **274**, 6031–6034.
44. Olsen, J.G., Kadziola, A., von Wettstein-Knwles, P., Siggaard-Andersen, M., Lindquist, Y. & Larsen, S. (1999). The X-ray structure of β-ketoacyl [acyl carrier protein] synthase I. *FEBS Lett.* **460**, 46–52.
45. Qiu, X., et al., & Abdel-Meguid, S.S. (1999). Crystal structure of β-ketoacyl-acyl carrier protein synthase III. *J. Biol. Chem.* **274**, 36465–36471.
46. Gehring, A.M., Mori, I., Perry, R.D. & Walsh, C.T. (1998). The nonribosomal peptide synthetase HMWP2 forms a thiozoline ring during biogensis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry* **37**, 11637–11650.
47. Gehring, A.M., et al., & Perry, R.D. (1998). Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem. Biol.* **5**, 573–586.
48. Albertini, A.M., Caramori, T., Scoffone, F., Scotti, C. & Galizzi, A. (1995). Sequence around the 159 degree region of the *Bacillus subtilis* genome: the *pksX* locus spans 33.6 kb. *Microbiology* **141**, 299–309.
49. Kunst, F. et al. (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256.
50. Haydock, S.F., et al., & Leadlay, P.F. (1995). Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthase. *FEBS Lett.* **374**, 246–248.
51. Kagan, R.M. & Clarke, S. (1994). Widespread occurance of three sequence motifs in diverse *S*-adenosylmethione-dependent methyltransferases suggests a common structure for these enzymes. *Arch. Biochem. Biophys.* **310**, 417–427.
52. Scrutton, N.S., Berry, A. & Perham, R.N. (1990). Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* **343**, 38–42.
53. Walsh, C.T., Gehring, A.M., Weinreb, P.H., Quadri, L.E.N. & Flugel, R.S. (1997). Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr. Opin. Chem. Biol.* **1**, 309–315.
54. de Crécy-Lagard, V., et al., & Thibaut, D. (1997). Pristinamycin I biosynthesis in *Streptomyces pristinaespiralis*: molecular characterization of the first two structural peptide synthetase genes. *J. Bacteriol.* **179**, 705–713.
55. Valvano, M.A., Marolda, C.L., Bittner, M., Glaskin-Clay, M., Simon, T.L. & Klena, J.D. (2000). The *rfaE* gene from *Escherichia coli* encodes a bifunctional protein involved in the biosynthesis of the lipopolysaccharide core precursor ADP-L-glycero-D-manno-heptose. *J. Bacteriol.* **182**, 488–497.
56. Mao, Y.Q., Varoglu, M. & Sherman, D.H. (1999). Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from *Streptomyces lavendulae* NRRL2564. *Chem. Biol.* **6**, 251–263.
57. de Kievit, T.R. & Lam, J.S. (1997). Isolation and characterization of two genes, *waaC* (*rfaC*) and *waaF* (*rfaF*), involved in *Pseudomonas aeruginosa* serotype O5 inner-core biosynthesis. *J. Bacteriol.* **179**, 3451–3457.
58. Kolkman, M.A., Wakarchuk, W., Nuijten, P.J. & van der Zeijst, B.A. (1997). Capsular polysaccharide synthesis in *Streptococcus pneumoniae* serotype 14: molecular analysis of the complete *cps* locus and identification of genes encoding glycosyltransferases required for the biosynthesis of the tetrasaccharide subunit. *Mol. Microbiol.* **26**, 197–208.
59. Jia, S., et al., & Friedman, P.A. (1992). cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase. *J. Biol. Chem.* **267**, 14322–14327.
60. Kaneko, T., et al., & Tabata, S. (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109–136.

61. Méndez, C. & Salas, J.A. (1998). ABC transporters in antibiotic-producing actinomycetes. *FEMS Microbiol. Lett.* **158**, 1–8.
62. Hoffmann, B., Strauch, E., Gewinner, C., Nattermann, H. & Appel, B. (1998). Characterization of plasmid regions of foodborne *Yersinia enterocolitica* biogroup 1A strains hybridizing to the *Yersinia enterocolitica* virulence plasmid. *Syst. Appl. Microbiol.* **21**, 201–211.
63. Wilson, V.T. & Cundliffe, E. (1999). Molecular analysis of *tlrB*, an antibiotic-resistance gene from tylosin-producing *Streptomyces fradiae*, and discovery of a novel resistance mechanism. *J. Antibiot.* **52**, 288–296.
64. Chater, K.F. & Bibb, M.J. (1997). Regulation of bacterial antibiotic production. In *Products of secondary metabolism*. Biotechnology, vol. 6. (Kleinkauf, H., von Dohren, H., Eds.), pp. 57–105, VCH, Weinheim.
65. Zhang, J.-H., Quigley, N.B. & Gross, D.C. (1997). Analysis of the *syrP* gene, which regulates syrignomycin synthesis by *Pseudomonas syrignae* pv. *syringae. Appl. Environ. Microbiol.* **63**, 2771–2778.
66. Thelwell, C., Robinson, N.J. & Turner-Cavet, J.S. (1998). An SmtB-like repressor from *Synechocystis* PCC 6803 regulates a zinc exporter. *Proc. Natl. Acad. Sci. USA* **95**, 10728–10733.
67. Mizuno, M., et al., & Kobayashi, Y. (1996). Systematic sequencing of the 283 kb 210–232 degrees region of the *Bacillus subtilis* genome containing the skin element and many sporulation genes. *Microbiology* **142**, 3103–3111.
68. Scofield, M.A., Lewis, W.S. & Schuster, S.M. (1990). Nucleotide sequence of *Escherichia coli asnB* and deduced amino acid sequence of asparagine synthetase B. *J. Biol. Chem.* **265**, 12895–12902.
69. Smith, D.R., et al., & Reeve, J.N. (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* delta H: functional analysis and comparative genomics. *J. Bacteriol.* **179**, 7135–7155.
70. Quadri, L.E., Sello, J., Keating, T.A., Weinreb, P.H. & Walsh, C.T. (1998). Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem. Biol.* **5**, 631–645.
71. Challis, G.L., Ravel, J. & Townsend, C.A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**, 211–224.
72. Conti, E., Stachelhaus, T., Marahiel, M.A. & Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174–4183.
73. Hovemann, B.T., Ryseck, R.-P., Walldorf, U., Störtkuhl, K.F., Dietzel, I.D. & Dessen, E. (1998). The *Drosophila ebony* gene is closely related to microbial peptide synthetases and shows specific cuticle and nervous system expression. *Gene* **221**, 1–9.
74. Zhu, W., Arceneaux, J.E., Beggs, M.L., Byers, B.R., Eisenach, K.D. & Lundrigan, M.D. (1998). Exochelin genes in *Mycobacterium smegmatis*: identification of an ABC transporter and two non-ribosomal peptide synthetase genes. *Mol. Microbiol.* **29**, 629–639.
75. Yu, S., Fiss, E. & Jacobs Jr., W.R. (1998). Analysis of the exochelin locus in *Mycobacterium smegmatis*: biosynthesis genes have homology with genes of the peptide synthetase family. *J. Bacteriol.* **180**, 4676–4685.
76. Nishizawa, T., Asayama, M., Fujii, K., Harada, K. & Shirai, M. (1999). Genetic analysis of the peptide synthetase genes for a cyclic heptapeptide microcystin in *Microcystis spp. J. Biochem.* **126**, 520–529.
77. Pospiech, A., Bietenhander, J. & 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*. *Microbiology* **142**, 741–746.
78. Guenzi, E., Galli, G., Grgurina, I., Gross, D.C. & Grandi, G. (1998). Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.* **273**, 32857–32863.
79. Guenzi, E., Galli, G., Grgurina, I., Pace, E., Ferranti, P. & Grandi, G. (1998). Coordinate transcription and physical linkage of domains in surfactin synthetase are not essential for proper assembly and activity of the multienzyme complex. *J. Biol. Chem.* **273**, 14403–14410.
80. Bentley, S.D., Parkhill, J., Barrell, B.G. & Rajandream, M.A. (1999) Direct submission to GenBank with an accession number CAB38517.
81. Hopwood, D.A., et al. & Schrempf, H. (1985). *Genetic Manipulation of Streptomyces: a Laboratory Manual*. John Innes Foundation, Norwich.
82. Kenani, A., Lamblin, G. & Henichart, J.-P. (1988). A convenient method for cleavage of the D-mannosyl-L-gulose disaccharide from bleomycin A2. *Carbohydr. Res.* **177**, 81–89.
83. Roy, R.S., Gehring, A.M., Milne, J.C., Belshaw, P.J. & Walsh, C.T. (1999). Thiazole and oxazole peptides: biosynthesis and modular machinery. *Nat. Prod. Rep.* **16**, 249–263.
84. Yu, T.-W., et al., & Floss, H,G. (1999). Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. *Proc. Natl. Acad. Sci. USA* **96**, 9051–9056.
85. Kennedy, J., Auclair, K., Kendrew, S.G., Park, C., Vederas, J.C. & Hutchinson, C.R. (1999). Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science* **284**, 1368–1372.
86. Holzbaur, I.E., et al., & Leadlay, P.F. (1999). Molecular basis of Celmer's rules: the role of two ketoreductase domains in the control of chirality by the erythromycin modular polyketide synthase. *Chem. Biol.* **6**, 189–195.
87. Böhm, I., Holzbaur, I.E., Hanefeld, U., Cortés, J., Staunton, J. & Leadlay, P.F. (1998). Engineering of a minimal modular polyketide synthase, and targeted alteration of the stereospecificity of polyketide chain extension. *Chem. Biol.* **5**, 407–412.
88. August, P.R., et al., & Floss, H.G. (1998). Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. *Chem. Biol.* **5**, 69–79.
89. Schwecke, T., et al., & Leadley, P.F. (1995). The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. USA* **92**, 7839–7843.
90. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning, a Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
91. Vara, J., Lewandowska-Sharbek, M., Wang, Y.-G., Donadio, S. & Hutchinson, C.R. (1989). Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). *J. Bacteriol.* **171**, 5872–5881.
92. Rao, R.N., Richardson, M.A. & Kuhstoss, S. (1987). Cosmid shuttle vectors for cloning and analysis of *Streptomyces* DNA. *Methods Enzymol.* **153**, 166–198.
93. Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for VAX. *Nucleic Acids Res.* **12**, 387–395.
94. Mootz, H.D. & Marahiel, M.A. (1997). The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843–6850.
95. Olano, C., Rodriguez, A.M., Michel, J.M., Mendez, C., Raynal, M.C. & Salas, J.A. (1998). Analysis of a *Streptomyces antibioticus* chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring. *Mol. Gen. Genet.* **259**, 299–308.
96. Rowland, B.M. & Taber, H.W. (1996). Duplicate isochorismate synthase genes of *Bacillus subtilis*: regulation and involvement in the biosyntheses of menaquinone and 2,3-dihydroxybenzoate. *J. Bacteriol.* **178**, 854–861.
97. Cole, S.T., et al., & Barrell, B.G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544.
98. Duitman, E.H., et al., & Vater, J. (1999). The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc. Natl. Acad. Sci. USA* **96**, 13294–13299.
99. Donadio, S. & Katz, L. (1992). Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. *Gene* **111**, 51–60.