Chemistry & Biology, Vol. 11, 1071–1079, August, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.chembiol.2004.05.014

Identification and Analysis of the Core Biosynthetic Machinery of Tubulysin, a Potent Cytotoxin with Potential Anticancer Activity

Axel Sandmann,[1,2] Florenz Sasse,[1] and Rolf Müller[1,2,*]
[1]GBF-German Research Center for Biotechnology
Mascheroder Weg 1
38124 Braunschweig
[2]Pharmaceutical Biotechnology
Saarland University
P.O. Box 151150
66041 Saarbrücken
Germany

Summary

Myxobacteria are well known for their biosynthetic potential, especially for the production of cytotoxic compounds with potential anticancer activities. The tubulysins are currently in preclinical development. They are produced in very low quantities, and genetic manipulation of producing strains has never been accomplished. We report the development of a *mariner*-based transposon mutagenesis system for *Angiococcus disciformis* An d48. Extracts from a library of 1200 mutants were analyzed for the presence of tubulysin by a microscopic cell nucleus fragmentation bioassay. The transposition sites of four tubulysin-negative mutants were identified by vector recovery, which led to the identification and the sequencing of the corresponding core biosynthetic gene locus. Sequence analysis of more than 80,000 bp reveals an unusual multimodular hybrid polyketide synthase/peptide synthetase assembly line with a variety of unprecedented features.

Introduction

Natural products interacting with the cytoskeleton of the cells of higher organisms are commonly used as lead structures for the development of anticancer agents (e.g., taxol, dolastatin, epothilone). They are mainly derived from plant or bacterial secondary metabolism and are usually available in minute quantities only [1]. In order to provide the amount needed for preclinical and clinical development, research aimed at improving the supply of the compound is triggered once a natural product with promising activity is found. One possibility is chemical total synthesis, which is often hampered by complex stereochemistry and low overall yield [2]. Alternatively, natural and heterologous production hosts are genetically altered to increase productivity [3]. As a prerequisite, genetic systems must be available and the molecular basis of the compound's biosynthesis needs to be understood. In bacteria, most knowledge has been gained from studies with actinomycetes and bacilli, in which the biosynthesis of polyketides (PKs) and nonribosomally made peptides (NRPs) is currently under investigation in a variety of laboratories [4–6]. Given the fact that myxobacteria offer an alternative and prodigous source of novel bioactive compounds [7, 8], we initiated a program aimed at the better understanding of natural product biosynthesis and its regulation in these microorganisms [9–12]. Myxobacteria produce five different groups of compounds which have been shown to interact with the cytoskeleton of higher cells: chondramides [13, 14] and rhizopodins [15] act on actin fibers, while epothilones [16, 17], disorazols [18], and tubulysins (Figure 1, [19, 20]) interact with the microtubule network. Whereas epothilones stabilize filament formation, tubulysins induce their disruption (also see Figure 2). Epothilones are currently in phase III clinical trials as anticancer agents, while disorazol and the tubulysins are in preclinical development. Interestingly, all of the mentioned secondary metabolites are presumably made by polyketide synthase (PKS)/nonribosomal peptide synthetase (NRPS) hybrid systems, because they contain amino acids and typical polyketide fragments.

Genetic tools to study most of the myxobacterial genera are poorly established or completely unavailable. Nevertheless, the epothilone biosynthetic gene cluster has been identified from *Sorangium cellulosum* [21, 22] and used to heterologously produce epothilones in *Myxococcus xanthus* [23]. However, nothing is known about the genetics of natural product biosynthesis in the tubulysin producer.

This study reports the development of genetic tools for the tubulysin producer *Angiococcus disciformis* An d48, the identification of the tubulysin biosynthetic gene cluster by *mariner*-based transposon mutagenesis, and a molecular analysis of the tubulysin megasynthetase.

Results

Identification of the Tubulysin Biosynthetic Gene Cluster by *mariner*-Based Mutagenesis

Because no genetic system was available for *A. disciformis* An d48, growth and plating conditions for the strain were established and it was analyzed for its susceptibility to antibiotics [28]. As myxobacteria have been shown to contain a wide variety of secondary metabolic gene clusters of the PKS/NRPS type [7, 29], it did not seem reasonable to screen a gene library of the strain for such hybrid gene clusters by using a hybridization approach. Instead, a transposon mutagenesis approach was applied to identify the tubulysin biosynthetic gene cluster. The theoretical size of the tubulysin biosynthetic gene cluster was estimated as approximately 40 kbp, because seven biosynthetic PKS and NRPS modules and some downstream processing genes were expected (see below). If one assumes the genome size of *A. disciformis* An d48 to be 10 Mbp (similar to *M. xanthus* and *S. cellulosum* [7]), a randomly inserting transposon should hit the corresponding gene cluster in 1 out of 250 mutants. Rubin et al. [27] developed a *mariner*-based transposon for mycobacteria, which can be effi-

*Correspondence: rom@mx.uni-saarland.de

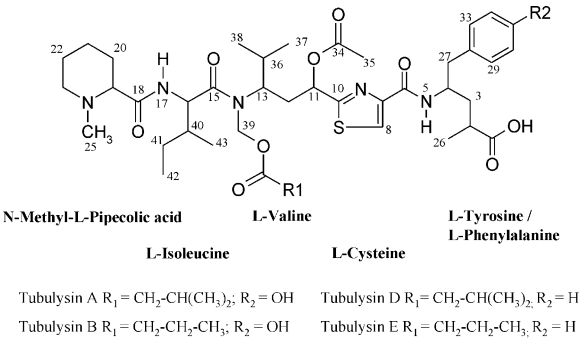

Figure 1. Structure of Tubulysins

ciently used to generate insertion mutants. Plasmid pMycoMar encodes the minitransposon *magellan*4 which contains the Tn5 kanamycin resistance gene and the oriR6K, which are flanked by *Himar*1 inverted repeats. In addition, pMycoMar encodes the *Himar*1 transposase under the transcriptional control of the T6 promotor. *Himar*1 transposition is characterized by a TA dinucleotide recognition sequence which enables random integration into the target DNA. As *A. disciformis* An d48 turned out to be kanamycin sensitive and there is no evidence so far for any plasmid replicating in any myxobacterium, we directly used pMycoMar for *A. disciformis* An d48 electroporation, which after optimization of the protocol (data not shown) led to a reasonably efficient in vivo transposition protocol.

A mutant library consisting of approximately 1200 mutants was generated and screened using the highly sensitive cell nucleus fragmentation assay (cnf) described under Experimental Procedures, because the production level of tubulysin does not allow detection of the compound by simple TLC- or HPLC-based screening methods (production is less than 1 mg/l [19]). Using the bioassay, a concentration of approximately 50 ng/ml can be detected (B. Frank, F.S., and R.M., unpublished data). After extraction, four mutants did show a tubulysin-negative phenotype in the microtiter plate based cnf assay. These mutants were independently grown in 50 ml cultures, extracted, and the resulting extracts fractionated. Each extract was used for another cnf assay, because myxothiazol that is produced in parallel can mask the activity of tubulysin. Each mutant was shown again to be tubulysin negative. In addition, HPLC-MS analysis of the extracts from the 50 ml cultures revealed the presence of tubulysin in the wild-type, whereas all of the mutants were shown to be tubulysin negative (data not shown). These mutants still produce myxothiazol.

In order to identify the transposition site in each mutant, a "transposon recovery" strategy was used. Chromosomal DNA was prepared and digested with restriction enzymes that do not cut within the transposable element. Religation and transformation into *E. coli* DH5α/λpir leads to kanamycin-resistant cells which harbor a plasmid that consists of the transposable element (with the kanamycin resistance gene and the origin of replication oriR6K) and chromosomal *A. disciformis* An d48 DNA that originally flanks the transposition site, whose sequence revealed that genes encoding PKS and NRPS with high similarities to ones from other secondary metabolite gene clusters were hit. The insertion position of the transposons within the biosynthetic gene cluster is shown in Figure 3. Restriction analysis and Southern blot analysis indicated the size of each plasmid recovered. After generation of a cosmid gene library of *A. disciformis* An d48, the biosynthetic gene cluster shown in Figure 3 was cloned and sequenced from several overlapping plasmids and cosmids using hybridization and PCR-based strategies. The average G+C content of the sequenced segment, spanning approximately 83 kbp is 68.5%.

Organization of the Tubulysin Biosynthetic Gene Cluster

Sequence motifs typical for PKS domains [30, 31] and NRPS domains [32, 33] were detected in TubA-F (Table 1 and Figure 3; also see Figure 3 for the abbreviations used). The acyl carrier protein (ACP) domains and the peptidyl carrier protein (PCP) domains of TubB-F contain the Prosite consensus signature of the putative binding site for the 4′-phosphopantetheine (Ppant) cofactor (Prosite signature numbers PS00012, R2082, and

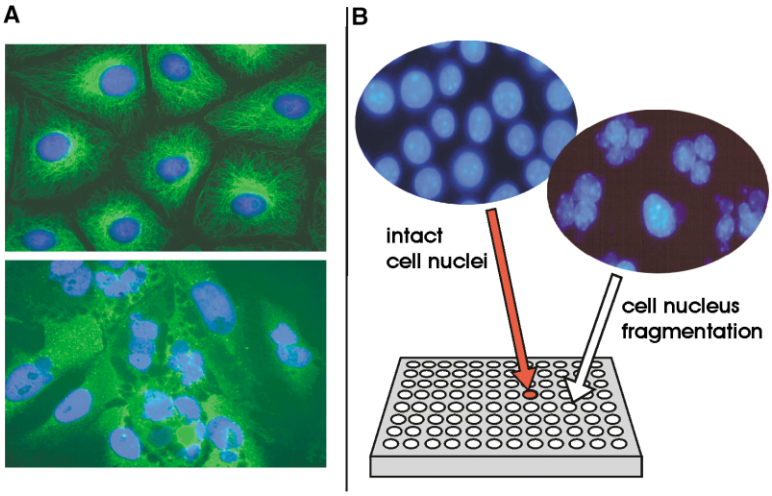

Figure 2. Activity of Tubulysin against Eukaryotic Cancer Cells and Screen for Tubulysin-Negative Mutants

(A) Influence of tubulysin A (50 ng/ml) on the microtubule cytoskeleton of Ptk_2 potoroo cells. The cells were fixed and immunostained for tubulin. In the upper picture microtubules of control cells are shown (green, microtubule network; blue, cell nuclei), the lower picture was taken after incubation with tubulysin A for 24 hr. Only a diffuse tubulin fluorescence is left. Initial cell nucleus degradation can be observed.

(B) Screen for tubulysin knockout mutants of *A. disciformis* An d48 by cnf in a bioassay. Tubulysin causes a characteristic cell nucleus fragmentation, which can be observed in extracts from producing cells (shown in white). Nonproducers do not induce cnf (shown in red).

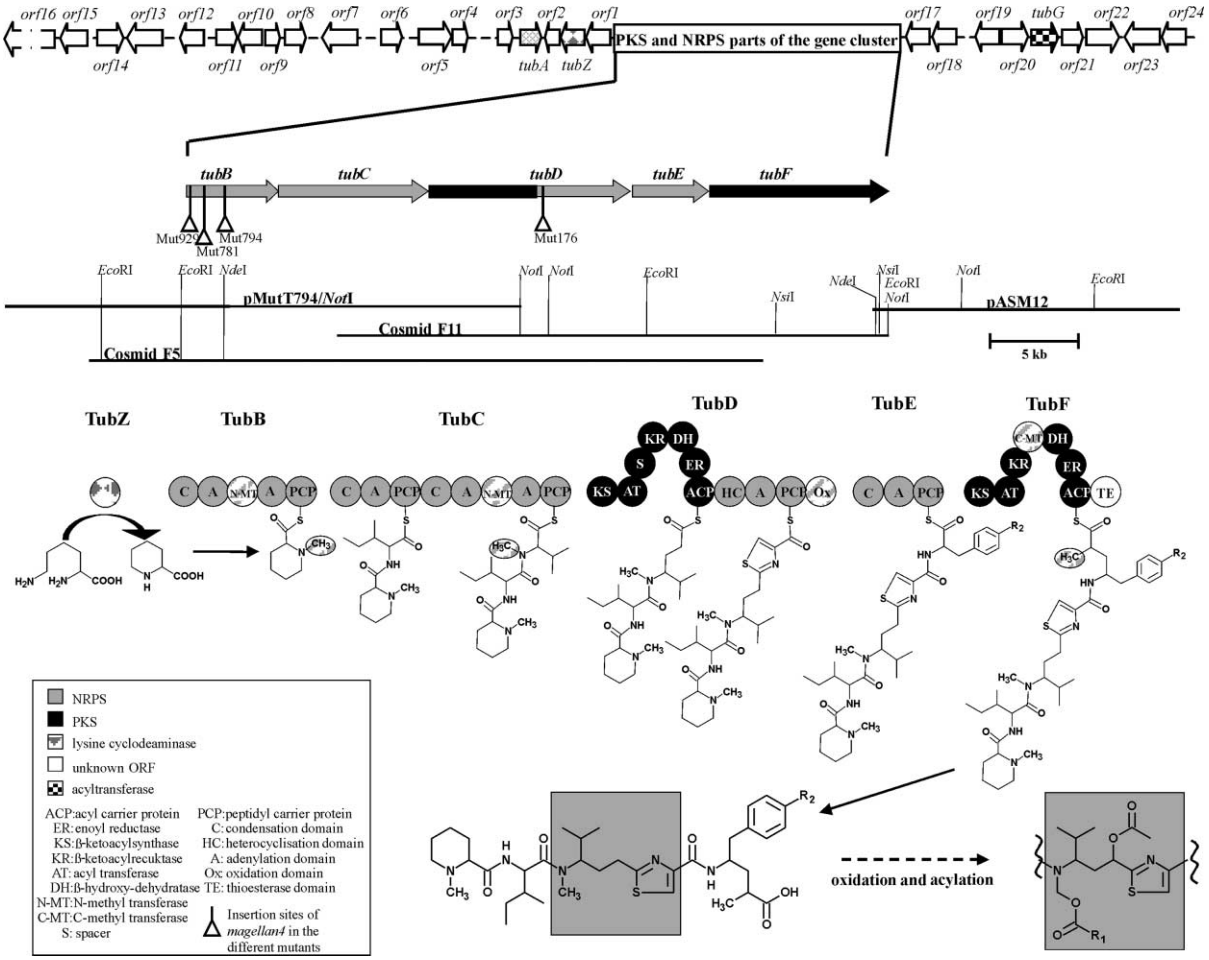

Figure 3. Model for Tubulysin Biosynthesis

In the upper section the genetic organization of the tubulysin biosynthesis gene cluster is shown. Inserts of plasmids and cosmids generated in this work are indicated by solid lines. The localization and transcriptional orientation of the genes is indicated by arrows which are not drawn to scale. Underneath, a restriction map of gene locus is presented. In the lower part, a hypothetical scheme of tubulysin biosynthesis is shown according to the domain organization of TubB-F and TubZ. Further acylation and oxidation in the presumed pretubulysin released from TubF would result in tubulysin formation. The affected part of the molecule is shown in gray. R1 and R2 are labeled according to Figure 1.

L2104). The codon bias of the genes reported is in accordance with other genes from myxobacteria [34].

The NRPS gene *tubB* most likely starts with an ATG which is located 11 bp downstream of a putative *r*ibosome *b*inding *s*ite (RBS) GGTG. *tubB* appears to be part of an operon with *tubC-F. tubC* represents another NRPS with a putative RBS (AGGA) 8 bp upstream from the start codon TTG. The following gene encodes a hybrid PKS/NRPS protein (TubD) which belongs to the covalently connected PKS and NRPS modules on one open reading frame, most of which are derived from myxobacteria (e.g., the MtaD or MelD proteins [11, 12]). *tubD* starts with ATG and a RBS (GGGA) is located 7 bp upstream. *tubE* encodes another NRPS module (a putative RBS AAGG is located 8 bp upstream of the ATG start codon) and is followed by *tubF* harboring one PKS module with a complete set of reductive loop domains in addition to the methyltransferase domain and the thioesterase domain. To our knowledge, with a MW of 309 kDa, TubF represents the largest bacterial type I PKS module known. *tubF* starts with an ATG which is located 10 bp behind its putative RBS GAGGA. The domain organization of TubB-F is shown in Figure 3 and Table 1. Table 2 displays a comparison of the core regions of the *tub* adenylation (A) domains, which are known to be important for substrate specificity [35, 36]. Except for the TubE A-domain, the comparison reveals good correlation with the "nonribosomal code" (see Discussion). The genes *tubB-E* appear to be translationally coupled. The methyl transferase (MT) domains of TubB, TubC, and TubF show the conserved S-adenosyl-methionine binding motifs [37].

Two deviations to standard NRPS domains [33] have been found in all tubulysin modules; the core motifs A2 of the A domains and the core regions of the PCP domains lack one amino acid (see Figure 4).

Upstream of *tubB* another putative operon, transcriptionally directed in the opposite direction can be found. Whereas ORF1 encodes a putative ATP-dependent anion transporter and ORF2 encodes a conserved hypothetical protein (Table 1), *tubZ* encodes a protein with similarity to lysine cyclodeaminases (see Table 1). Fur-

Table 1. Proteins Encoded in the Sequenced Region Including the Core Tubulysin Biosynthetic Gene Cluster and Their Putative Function

NRPS and PKS Part of the Gene Cluster

Protein (Gene)	Size (Da/bp)	Proposed Function (Protein Domains with Their Position in the Sequence)
TubB (*tubB*)	170.704/4.626	NRPS domains: C (76–416), A (555–1432), N-MT (950–1403), PCP (1433–1516)
TubC (*tub*C)	289.141/7.878	NRPS domains: C (77–418), A (561–1024), PCP (1025–1106), C (1137–1477), A (1613–2496), N-MT (2004–2468), PCP (2497–2575)
TubD (*tub*D)	383.778/10.548	PKS domains: KS (1–437), AT (538–858), S (861–1151), KR (1155–1433), DH (1452–1648), ER (1686–1990), ACP (2050–2150) NRPS domains: HC (2151–2620), A (2664–3134), PCP (3135–3211), Ox (3218–3516)
TubE (*tub*E)	130.337/3.546	NRPS domains: C (47–388), A (536–1049), PCP (1050–1139)
TubF (*tub*F)	309.369/8.514	PKS domains: KS (9–406), AT (537–858), KR (953–1260), C-MT (1295–1653), DH (1714–1909), ER (2049–2342), ACP (2420–2501), TE (2553–2828)

ORFs Encoded Upstream and Downstream of *tubB-tubF*

Protein (Gene)	Size (Da/bp)	Proposed Function of the Similar Protein	Sequence Similarity to Source	Similarity/Identity	Accession No. of the Similar Protein
Orf16 (*orf16*)	122.434/3.308	valyl-tRNA synthetase	*Thermotoga maritima*	69%/54%	ZP_00080134
Orf15 (*orf15*)	41.337/1.254	CheY subfamily	*Synechocystis sp.*	60%/36%	NP_440346
Orf14 (*orf14*)	52.449/1.446	response regulator	*Geobacter metallireducens*	57%/35%	ZP_00081165
Orf13 (*orf13*)	83.814/2.241	two component sensor kinase/response regulator hybrid	*Agrobacterium tumefaciens*	47%/29%	NP_535879
Orf12 (*orf12*)	58.650/1.644	60 kDa chaperonin	*Geobacter sulfurreducens*	84%/70%	NP_954380
Orf11 (*orf11*)	58.650/1.644	conserved hypothetical protein	*Streptomyces coelicolor*	53%/33%	NP_628275
Orf10 (*orf10*)	36.441/999	nucleoside-diphosphate-sugar epimerases	*Ralstonia metallidurans*	79%/65%	ZP_00026477
Orf9 (*orf9*)	23.789/645	transcriptional regulator, TetR family	*Caulobacter crescentus*	52%/30%	NP_420005
Orf8 (*orf8*)	41.614/1.137	hypothetical 119.5K protein (uvrA region)	*Myxococcus luteus*	29%/25%	JQ0405
Orf7 (*orf7*)	89.255/2.346	hypothetical protein	*Nostoc punctiforme*	46%/28%	ZP_00110488
Orf6 (*orf6*)	43.228/1.157	serine/threonine kinase Pkn14	*Myxococcus xanthus*	47%/33%	AAK64427
Orf5 (*orf5*)	68.825/1.923	protein kinase	*Stigmatella aurantiaca*	57%/42%	CAD19078
Orf4 (*orf4*)	33.507/921	adenine deaminase-related protein	*Deinococcus radiodurans*	38%/33%	NP_285591
Orf3 (orf3)	20.880/543	hypothetical protein	*Pseudomonas syringae*	46%/32%	NP_790430
TubA (*tub*A)	43.202/1.173	similarity to C domain core motivs C1-C2	*Nostoc punctiforme*	40%/24%	ZP_00109659
Orf2 (*orf2*)	25.390/660	conserved hypothetical protein	*Neurospora crassa*	42%/29%	CAD11370
TubZ (*tubZ*)	39.764/1.101	lysine cyclodeaminase	*Streptomyces hygroscopicus*	53%/39%	CAA60467
Orf1 (*orf1*)	46.030/1.218	anion transporting ATPase	*Aquifex aeolicus*	65%/40%	NP_213468
Orf17 (*orf17*)	38.371/1.056	patatin-like protein	*Anabaena sp.* 90	39%/23%	CAC01602
Orf18 (*orf18*)	38.371/1.056	patatin-like protein	*Anabaena sp.* 90	40%/24%	CAC01602
Orf19 (*orf19*)	20.040/555	hypothetical protein	*Streptomyces avermitilis*	35%/29%	NP_823780
Orf20 (*orf20*)	43.282/1.158	hypothetical protein	*Microbulbifer degradans*	62%/41%	ZP_00065421
TubG (*tubG*)	28.039/759	N-hydroxyarylamine O-acetyltransferase	*Streptomyces avermitilis*	53%/47%	NP_826733
Orf21 (*orf21*)	33.859/927	hypothetical protein	*Xanthomonas axonopodis*	39%/28%	NP_641500
Orf22 (*orf22*	32.668/882	hypothetical protein	*Rhizobium etli*	49%/37%	NP_659913
Orf23 (*orf23*)	49.133/12.63	similar to penicillin binding protein	*Gloeobacter violaceus*	67%/49%	NP_926044
Orf24 (*orf24*)	37.621/1.077	NosF	*Nostoc sp.*	44%/31%	AAF17284

ther upstream, TubA is encoded, which represents a protein with similarity to fragments of C domains from NRPSs. Sequence analysis of an additional 31.2 kbp upstream revealed the presence of several putative regulators and two component systems, which are briefly described in Table 1 (see EMBL entry). Downstream of

Table 2. Substrate Specificity Pocket of *tub* A-Domains

Position of the amino acid within the A-domain	235	236	239	278	299	301	322	330
Consensus pipecolic acid	D	Y	Q	Y	C/L	G	H	L
A-domain TubB	D	I	Q	Y	I	A	Q	V
Consensus isoleucine	D	G A	F	F	L F Y	G	V I	V T
A1-domain TubC	D	A	F	F	L	G	V	T
Consensus valine	D	A V F	F Y W	W F M	I F L	A G	A G	T V
A2-domain TubC	D	A	F	W	L	G	G	T
Consensus cysteine	D	L H	Y F E	N S	M L D	S I V	M L G	I
A-domain TubD	D	L	Y	N	M	S	L	I
Consensus tyrosine	D	G A	T L	I G	T	A G	E	V
Consensus phenylalanine	D	A L	P F W	I T G	M V I	G A	G A	T V I
A-domain TubE	A	L	Y	W	L	A	S	F

The numbering of the amino acids corresponds to the GrsA gramicidin synthetase. The GrsA phenylalanine substrate pocket has been crystallized [43], which led to the establishment of the nonribosomal code [35, 36].

tubF, 13 kbp of DNA were sequenced and analyzed. Within this region, genes encoding hypothetical ORFs were detected. In addition, a protein with similarity to acyltransferases was found (Figure 3 and Table 1).

Discussion

After the description of the epothilone biosynthetic gene cluster [21, 22], this is only the second report of a mega-synthetase involved in the biosynthesis of a tubulin-interacting compound from a myxobacterium. As tubulysin has a different mode of action from epothilone, it is expected that the modification of tubulysin biosynthesis will similarly trigger research toward the generation of improved tubulysin analogs in the future.

Analysis of 1200 mutants of *A. disciformis* An d48 and localization of the insertion positions in the genome of tubulysin nonproducers, all of which mapped to the same gene locus, led to the identification of the tubulysin gene cluster. Based on the in silico analysis of the genes involved, a hypothetical biosynthetic route to tubulysin could be established as described below.

adenylation-domains (motif A2)

```
                              L I
Konsensus              LKAGxAYLVPLD

TubC A1          ·····LQAGAAY-VPLD·····
TubC A2          ·····LKAGAAY-VPLD·····
TubB A           ·····WKAGAGF-LPLD·····
TubD A           ·····LQAAAAY-LPLD·····
TubE A           ·····VLAGVAP-LVLP·····
```

PCP-domains

```
                                 D I
Konsensus               DxFFxxLGGHSL

TubB PCP          ·····DNFFQ-LGGHSL·····
TubC PCP1         ·····DDFFD-LGGHSL·····
TubC PCP2         ·····DNFFD-VGGHSL·····
TubD PCP          ·····DSLLD-LGADSV·····
TubE PCP          ·····ASFFD-LGGNSI·····
```

Figure 4. Tubulysin Core Motifs A2 from A-Domains and Core Motifs of PCP Domains Differ from Standard NRPS Domains

Consensus core domain motifs are shown according to Konz and Marahiel [33]. Amino acids "missing" in the tubulysin domains are indicated in gray.

Proposed Biosynthetic Pathway to Tubulysin in *A. disciformis* An d48

Tubulysin biosynthesis most likely starts with the biosynthesis of pipecolic acid from lysine via the action of the putative lysine cyclodeaminase TubZ. This proposal is in accordance with the description of RapL, a protein highly similar to TubZ, which has been found in the rapamycin biosynthetic gene locus. RapL is proposed to be a lysine cyclodeaminase based on mechanistic and sequence comparisons with ornithine cyclodeaminase [38]. TubB indeed contains an A-domain with a substrate binding pocket typical for pipecolic acid (see Table 2). Differing from rapamycin biosynthesis, the pipecolic acid moiety in tubulysin is *N*-methylated, which is consistent with the finding that a MT domain is located within the A-domain of TubB. The insertion of domains within A-domains is unusual but not unprecedented; the MT domains of TubB and TubC are located between A-domain core regions A8 and A9, similar to microcystin biosynthetic protein MycA [39]. Interestingly, insertion of oxidation (Ox) domains at the same position has been reported for the myxothiazol biosynthetic proteins MtaD and MtaG (between core motifs A8 and A9 and between A4 and A5, respectively) [11]. In contrast, most of the known NRPS MT domains are located between the A- and PCP domains [33]. The biosynthetic role of the C domain at the N terminus of TubB is not clear, but it might have some function in the acylation of the product of the PKS/NRPS system (see below). Further, we propose that TubC extends the methylated pipecolic acid bound to the TubB-PCP with isoleucine and valine (which corresponds well with the nonribosomal code of the A-domains involved; see Table 2), the latter of which will be *N*-methylated by the TubC MT domain that is similar to the one in TubB.

At this point, the biosynthetic intermediate switches to TubD, a hybrid of PKS and NRPS. The first module of the protein represents a PKS containing a complete reductive loop that is used for the extension with a malonate moiety, which perfectly matches the final natural

product. Surprisingly, the order of the domains within the reductive loop is different from that of other known PKS systems. The KR domain is found C-terminal from the AT domain and followed by DH and ER domains, the latter of which is attached to the ACP domain. In contrast, typical reductive loops follow the order KS-AT-DH-ER-KR-ACP. The same atypical organization is also found in TubF (with the exception that a MT domain is inserted between the KR and DH domains). Such an unusual domain organization should influence the corresponding three-dimensional structure of PKS and especially PKS/NRPS hybrid systems and thus shows the high degree of structural flexibility of these megasynthetases, which could possibly suggest new experimental approaches to generate engineered enzymes and to better understand the currently obscure topology of the multidomain proteins during multistep catalysis.

The NRPS module of TubD contains a heterocyclization domain, a cysteine-specific A-domain (see Table 2), a PCP and an Ox domain (which is located behind the PCP). The Ox domain of EpoP was shown to oxidize the intermediate thiazoline ring after heterocyclization of cysteine during epothilone biosynthesis [40]. Next, TubE extends the growing chain with phenylalanine or tyrosine, finally resulting in the production of tubulysin D or A, respectively. The A-domain of this protein shows no significant similarity to any amino acid "specificity pocket," and we hypothesize that it is capable of loading either amino acid. We have identified a number of other atypical A-domains from myxobacteria which significantly differ from the "nonribosomal code." Sequence comparisons show that such differences exist for cyanobacterial NRPS domains as well (A.S., Silke Wenzel, and R.M., unpublished data, cp. [39]). Further examples of deviations from well-defined core regions of NRPSs [33] can be found in the tubulysin megasynthetase; both the A-domain core A2 and the PCP domain core lack one conserved amino acid (see Figure 4). The effect of these amino acid exchanges and deletions remains to be elucidated.

TubF, the last protein in the tubulysin assembly line, contains eight active domains and thus represents the largest known bacterial type I PKS module. Between the KR and DH domains a MT is found, which most likely methylates the β-keto intermediate in the α position, which eventually results in the methyl group in position 2 of tubulysin. All methyl groups correlated to the described MT domains have been shown by feeding experiments to be derived from ^{13}C-labeled methionine [20]. Action of the KR, DH, and ER domains results in the saturated and ACP-bound intermediate that is finally released by the action of TubF TE-domain as free acid.

The assumed product released from the PKS/NRPS assembly line would now have to be oxidized in positions 39 and 11, followed by acylations of the resulting OH groups (see Figure 3). However, no candidate gene(s) for these two oxidation processes could be found within the sequenced region, indicating that both reactions might be performed by enzymes encoded outside of the identified biosynthetic region. Alternatively, oxidases encoded in the sequenced region which cannot be identified as such by computer-based analysis might be involved in the biosynthesis. The same holds true for the acylation reactions. The latter could be catalyzed by the N-terminal C domain of TubB in analogy to lipopeptide biosynthesis, where N-terminal C domains are believed to couple an CoA-activated lipid to the peptide core [41]. Nevertheless, the C domain of TubB does not cluster with such C domains in phylogenetic tree analysis (data not shown). Alternatively, TubA could be involved in the acylations, because it shows similarity to fragments of NRPS C-domains (only core motifs C1 and C2 can be identified). On the other hand, it has been shown that core motif C3 is necessary for acyl transferase activity [42]. Thus, *tubG* located 7 kb downstream of *tubF*, seemed to be the most likely candidate gene for the acylation reaction(s). After establishing a preliminary and so far inefficient system for gene inactivation by homologous recombination in *A. disciformis* An d48, the gene was inactivated with no effect on tubulysin biosynthesis (A.S. and R.M., unpublished data). Therefore, it is not clear at this stage how the oxidation and acylation reactions are performed. Further experiments aimed at the identification of the post-PKS processing genes in *A. disciformis* An d48 are in progress.

The description of the tubulysin core biosynthetic genes which are most likely located between *tubA* and *orf17* (Figure 3) sets the stage for a detailed genetic and biochemical analysis of the biosynthesis of the compound. Studies aimed at increasing the tubulysin yield, as well as the heterologous expression of the core biosynthetic genes are in progress.

Significance

Microorganisms are used as production hosts for an immense variety of natural products with various biological activities. Myxobacteria are gaining attention due to their biosynthetic potential, especially for the production of cytotoxic compounds with potential anticancer activities. The epothilones and the tubulysins are examples in clinical and preclinical development, and the potential of tubulysins is demonstrated by the extremely low concentration needed to inhibit the growth of mammalian cell lines by 50% (IC_{50} value in the picomolar range). Tubulysins are produced in very low quantities, and genetic manipulation of producing strains has not been reported. This fact prompted us to develop a *mariner*-based transposon mutagenesis system for the tubulysin producer *Angiococcus disciformis* An d48 and identifiy the core biosynthetic gene cluster for tubulysin. Extracts from 1200 mutants were analyzed for their ability to produce the compound by a microscopic cell nucleus fragmentation bioassay based on sensitive mouse cells. No tubulysin activity was detected in extracts of four mutants. Each corresponding transposition site was identified by vector recovery, revealing that all of the mutants group to the same gene locus. The identified core tubulysin biosynthetic gene cluster is approximately 40,000 bp in size and represents a multimodular hybrid polyketide synthase/peptide synthetase assembly line with a variety of unprecedented features. The biosynthesis of the unusual *N*-methyl pipecolic acid starter moiety can be explained, and a novel arrangement of reductive loops

in polyketide synthases is reported. The identification of the gene cluster is relevant for potential combinatorial biosynthesis and sets the stage for the development of engineered tubulysin derivatives.

Experimental Procedures

DNA Manipulations, Analysis, Sequencing, and PCR

A. disciformis An d48 chromosomal DNA was prepared as described [24]. Southern blot analysis of genomic DNA was performed using the standard protocol for homologous probes of the DIG DNA labeling and detection kit (Roche Diagnostics, Mannheim, Germany).

Sequencing of the transposon plasmid pMutT794 and the cosmid F11 was performed by a shotgun approach as described previously [11]. Polymerase chain reaction (PCR) was carried out using *Pfu* DNA-Polymerase (Stratagene) according to the manufacturer's protocol. DMSO was added to the reaction mixture to a final concentration of 5%. Conditions for amplification with the Eppendorf Mastercycler gradient (Eppendorf, Germany) were as follows: denaturation for 30 s at 95°C, annealing for 30 s at 60°C and extension for 45 s at 72°C; 30 cycles and a final extension at 72°C for 10 min.

Primers used for the amplification of an internal fragment of *tubA* (given in the 5′ to the 3′ direction): $TubA_{up}$ 5′-CAG CAT GTC GAG GCA GAG C-3′; $TubA_{down}$ 5′-TCC CGA GCA TCC AAC GCA GAG-3′; internal fragment of *orf1*: $Orf1_{up}$ 5′-CCC TTC GGC AGC CCC AAC AT-3′, $Orf1_{down}$ 5′-GCC CGG CAA GAC AGA GC-3′; internal fragment of *tubZ*: $TubZ_{up}$ 5′-GGA GGT GGC CGT GCA GAG GAT GTC-3′, $TubZ_{down}$ 5′-CTG CAC GCG CTG ATG GAT GAG GTC-3′; internal fragment of *orf2*: $Orf2_{up}$ 5′-TCC AGG TAG GCA GAT TGA GAC-3′, $Orf2_{down}$ 5′-CAG CGG CAG TGG AGT GAA G-3′; internal fragment of *tubC*: ASTIs 2A 5′-GCT CCC GGG CCA CGT GGT TGA AGA-3′, ASTIs 2B 5′-CCG CGG GCC GTG GCA GTG GTG TA-3′; internal fragment of *tubD*: ASTIs 1A 5′-CAC CCG GAC CTG CCT GGA TTC-3′, ASTIs 1B 5′-TGC TCG GCT GGC GCT ACT CAC-3′; fragment behind *tubF*: TIs_{up} 5′-TGG CAG CCA GCC CGA GC-3′; Seq_{TE} 5′-GAC GCT GCT GCG GCC ACC TCA CG-3′.

All other DNA manipulations were performed according to standard protocols [25]. Amino acid and DNA alignments were done using the programs of the Lasergene software package (DNAstar Inc.) and ClustalW [26].

Generation of a Mutant Library of *A. disciformis* An d48 by Transposon Mutagenesis

A. disciformis An d48 was cultivated at 30°C in 50 ml tryptone medium (10 g tryptone, 2 g $MgSO_4$, 0.1% vitamin B12 [10 ng/ml], 0.2% glucose per liter medium, pH adjusted to 7.2) to a cell density of 2×10^8 cells/ml. Cells were harvested by centrifugation at 20°C (20 min, 4000 rpm) and resuspended in an equal amount of washing buffer (5 mM HEPES, 0.5 mM $CaCl_2$, pH adjusted to 7.2 using NaOH). After a second washing step in 25 ml buffer, cells were resuspended to give a final cell count of 1×10^9 cells in 40 μl. Transposon pMycoMar (1–3 μg; [27]) and 40 μl cell suspension were mixed and transferred into an ice-cold electroporation cuvette (0.1 cm). The electroporation conditions were: 200 Ω, 25 mF, and 0.85 kV/cm. Tryptone medium (1 ml) was added to the cells and the suspension was transferred into an Erlenmeyer flask with 50 ml tryptone medium. After incubation for 6 hr at 30°C, cells were harvested (20 min, 4000 rpm, 20°C) and resuspended in 1 ml tryptone medium. A dilution series of 1×10^8 to 1×10^4 cells was plated with 3 ml tryptone soft agar onto kanamycin selection tryptone medium (50 μg/ml). Mutant-colonies became visible after incubation at 30°C for 5–8 days.

A total of 1200 mutants were transferred into 96-well plates filled with 200 μl M7 medium per well (5 g probion [single-cell protein; Höchst AG], 1 g $CaCl_2$, 1 g $MgSO_4$, 1 g yeast extract, 5 g soluble starch, 10 g HEPES, 0.1% vitamin B12 [10 ng/ml] per liter medium, adjusted to pH 7.4) and incubated at 32°C. After cultivation for 10 days, 150 μl of each suspension was taken to dryness at 37°C (N_2 gas). Cell pellets were resuspended in 100 μl methanol, 10 μl of which were used for a *c*ell *n*ucleus *f*ragmentation (cnf) assay to detect the presence of tubulysin in each extract.

Toxicity Assay

L929 mouse cells (DSMZ: ACC2) were cultivated in DME medium (Invitrogen, Groningen, The Netherlands) at 37°C and harvested using a cell scraper. The resulting cell suspension was diluted with DME medium to a cell density of 50,000/ml and 120 μl of this suspension were aliquoted into each well of a 96-well plate. An aliquot (10 μl) of the extract from each *A. disciformis* An d48 mutant was added to the wells and the mixture was incubated for 5 days at 37°C. Subsequently, the L929 mouse cells were analyzed for tubulysin-induced cnf using a Zeiss Axiovert fluorescence microscope (Figure 2) [19].

For the verification of the result, mutants whose extracts failed to disrupt cell nuclei in the cnf assay were grown for 9 days in 50 ml M7 medium containing 1 ml adsorber-resin Amberlite XAD-16 (Rohm & Haas, Frankfurt/Main). Cells were harvested, extracted as described [19], and their extracts were used for a second cnf assay. In addition, extracts were fractionated by HPLC [20 μl of each sample was loaded onto the HPLC-column (Nucleosil 120-3 C_{18}, 2 × 125 mm, Macherey-Nagel, Düren) and chromotographed at a flow rate of 0.5 ml/min. Fractions were collected over 30 s, corresponding to a total volume of 250 μl per fraction and each fraction was subjected to the cnf assay. This approach effectively separates tubulysin from myxothiazol, another secondary metabolite produced by *A. disciformis* An d48 (F.S., unpublished data). Myxothiazol can mask the tubulysin activity in lower dilutions (F.S., unpublished data) as it inhibits the electron transport chain and leads to cell death leaving the cell nuclei intact while tubulysin causes apoptosis and thus cell nucleus fragmentation. Thus, the loss of tubulysin activity is evident by the presence of nonfragmented nuclei. In parallel, the integrity of the L929 mouse cell nuclei used in the cnf assay were analyzed by staining with 4′-6-*di*amino-2-*p*heny*l*indodol (DAPI) (Vollenweider, 1992). Both, the toxicity and the staining method clearly identified four tubulysin mutants: MutT176, MutT781, MutT794, and MutT929.

Identification of the *mariner*-Transposition Site by Vector Recovery

Chromosomal DNA (5 μg) prepared from each *A. disciformis* An d48 mutant grown in tryptone were digested using NotI and BamHI, phenol/chloroform purified, and ligated overnight at 16°C. The ligation mixture (1–3 μl) was electroporated into *E. coli* DH5α-λpir cells. Plasmids extracted from resulting kanamycin-resistant colonies were used for sequencing with primers K-388: 5′-TGG GAA TCA TTT GAA GGT TGG-3′ and K-390 5′-GGG TAT CGC TCT TGA AGG GAA C-3′.

Cloning and Sequencing of the Tubulysin Biosynthetic Gene Cluster

Plasmid pMutT794/NotI harbors 52,985 bp of chromosomal DNA from *A. disciformis* An d48 (see Figure 3). Together with the transposable element, 55,184 bp were sequenced. To clone the rest of the tubulysin biosynthetic gene cluster, a cosmid library of *A. disciformis* An d48 was prepared as described [11]. In colony hybridization experiments, approximately 1920 cosmid-harboring single colonies were used according to the manufacture's protocol (Roche Diagnostics, Mannheim) under stringent conditions. Screening probes were derived from *tubC* and *tubD* (oligonucleotides ASTIs 2 A/B were used to generate the probe derived from *tubC* and primer pair ASTIs 1A/B was used to generate the probe derived from *tubD*). Cosmids F4, F5, F7, F11, F13, and F16 were identified and further analyzed by PCR and restriction analyses. Additionally, these six cosmids were end sequenced and physically mapped (data not shown). The size of the overlap with known *tub* sequence was defined (see Figure 3).

Cosmid F11 was sequenced and harbors *tub*D, *tub*E, and *tub*F. Using PCR and Southern analysis, (primers TIs_{up}/Seq_{TE} were used to generate the probes) another cosmid (F7) harboring DNA located downstream of the tubulysin gene cluster was isolated. A NsiI/EcoRI fragment from F7 approximately 12 kbp in size was isolated and cloned into pUC18 (predigested with PstI/EcoRI) resulting in plasmid pASM12. This plasmid was subjected to in vitro transposition using the GPS-1 Genome Priming System (New England Biolabs, Schwalbach) according to the manufacturer's protocol. Plasmid

DNA was extracted from 192 kanamycin-resistant colonies and sequenced using primers N and S (N: 5′-ACT TTA TTG TCA TAG TTT AGA TCT ATT TTG-3′; S: 5′-ATA ATC CTT AAA AAC TCC ATT TCC ACC CCT-3′).

Acknowledgments

We would like to thank H. Reichenbach for furnishing strain *A. disciformis* An d48, H. Steinmetz for help with tubulysin chemical analysis, and B. Hinkelmann for technical assistance. We would like to acknowledge the helpful comments of H.G. Floss, B. Wilkinson, G. Höfle, and V. Simunovic regarding this manuscript.

Received: March 2, 2004
Revised: May 3, 2004
Accepted: May 18, 2004
Published: August 20, 2004

References

1. Newman, D., Cragg, G., and Snader, K. (2003). Natural products as sources of new drugs over the period 1981–2002. J. Nat. Prod. *66*, 1022–1037.
2. Nicolaou, K., Roschangar, F., and Vourloumis, D. (1998). Chemical biology of epothilones. Angew. Chem. Int. Ed. Engl. *37*, 2014–2055.
3. Donadio, S., Monciardini, P., Alduina, R., Mazza, P., Chiocchini, C., Cavaletti, L., Sosio, M., and Puglia, A.M. (2002). Microbial technologies for the discovery of novel bioactive metabolites. J. Biotechnol. *99*, 187–198.
4. Staunton, J., and Weissman, K.J. (2001). Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. *18*, 380–416.
5. Walsh, C.T. (2002). Combinatorial biosynthesis of antibiotics: challenges and opportunities. Chembiochem *3*, 125–134.
6. Cane, D., and Walsh, C. (1999). The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. Chem. Biol. *6*, R319–R325.
7. Gerth, K., Pradella, S., Perlova, O., Beyer, S., and Müller, R. (2003). Myxobacteria: proficient producers of novel natural products with various biological activities—past and future biotechnological aspects with the focus on the genus *Sorangium*. J. Biotechnol. *106*, 233–253.
8. Reichenbach, H., and Höfle, G. (1999). Myxobacteria as producers of secondary metabolites. In Drug Discovery from Nature, S. Grabley and R. Thieriecke, eds. (Berlin: Springer Verlag), pp. 149–179.
9. Gaitatzis, N., Kunze, B., and Müller, R. (2001). *In vitro* reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. Proc. Natl. Acad. Sci. USA *98*, 11136–11141.
10. Gaitatzis, N., Silakowski, B., Kunze, B., Nordsiek, G., Blöcker, H., Höfle, G., and Müller, R. (2002). The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. J. Biol. Chem. *277*, 13082–13090.
11. Silakowski, B., Schairer, H.U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blöcker, H., Höfle, G., Beyer, S., et al. (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.
12. Weinig, S., Hecht, H.-J., Mahmud, T., and Müller, R. (2003). Melithiazol biosynthesis: further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases. Chem. Biol. *10*, 939–952.
13. Kunze, B., Jansen, R., Sasse, F., Höfle, G., and Reichenbach, H. (1995). Chondramides A–D, new antifungal and cytostatic depsipeptides from *Chondromyces crocatus* (myxobacteria). Production, physico-chemical and biological properties. J. Antibiot. (Tokyo) *48*, 1262–1266.
14. Sasse, F., Kunze, B., Gronewold, T.M., and Reichenbach, H. (1998). The chondramides: cytostatic agents from myxobacteria acting on the actin cytoskeleton. J. Natl. Cancer Inst. *90*, 1559–1563.
15. Sasse, F., Steinmetz, H., Hofle, G., and Reichenbach, H. (1993). Rhizopodin, a new compound from *Myxococcus stipitatus* (myxobacteria) causes formation of rhizopodia-like structures in animal cell cultures. Production, isolation, physico-chemical and biological properties. J. Antibiot. (Tokyo) *46*, 741–748.
16. Gerth, K., Bedorf, B., Höfle, G., Irschik, H., and Reichenbach, H. (1996). Epothilons A and B: antifungal and cytotoxic compounds from *Sorangium cellulosum* (myxobacteria) production, physico-chemical and biological properties. J. Antibiot. (Tokyo) *49*, 560–563.
17. Bollag, D.M., McQueney, P.A., Zhu, J., Hensens, O., Koupal, L., Liesch, J., Goetz, M., Lazarides, E., and Woods, M. (1995). Epothilones, a new class of microtubule-stabilizing agents with Taxol-like mechanism of action. Cancer Res. *55*, 2325–2333.
18. Elnakady, Y., Sasse, F., Lünsdorf, H., and Reichenbach, H. (2004). Disorazol A_1, a highly effective antimitotic agent acting on tubulin polymerization and inducing apoptosis in mammalian cells. Biochem. Pharmacol. *49*, 560–563.
19. Sasse, F., Steinmetz, H., Heil, J., Höfle, G., and Reichenbach, H. (2000). Tubulysins, new cytostatic peptides from myxobacteria acting on microtubuli—production, isolation, physico-chemical and biological properties. J. Antibiot. (Tokyo). *53*, 879–885.
20. Steinmetz, H., Glaser, N., Herdtweck, E., Sasse, F., and Höfle, G. (2004). Isolation, crystal and solution structure, and biosynthesis of tubulysins—powerful inhibitors of tubulin polymerisation from myxobacteria. Angew. Chem. Int. Ed. Engl., in press.
21. Molnar, I., Schupp, T., Ono, M., Zirkle, R.E., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratman, A., Cyr, D.D., et al. (2000). The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. Chem. Biol. *7*, 97–109.
22. Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C., and Julien, B. (2000). Cloning and heterologous expression of the epothilone gene cluster. Science *287*, 640–642.
23. Arslanian, R.L., Tang, L., Blough, S., Ma, W., Qiu, R.G., Katz, L., and Carney, J.R. (2002). A new cytotoxic epothilone from modified polyketide synthases heterologously expressed in *Myxococcus xanthus*. J. Nat. Prod. *65*, 1061–1064.
24. Neumann, B., Pospiech, A., and Schairer, H.U. (1992). Rapid isolation of genomic DNA from Gram negative bacteria. Trends Genet. *8*, 332–333.
25. Sambrook, J., Fritsch, E.F., and Maniatis, T. eds. (1989). Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
26. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.
27. Rubin, E., Akerley, B., Novik, V., Lampe, D., Husson, R., and Mekalanos, J. (1999). *In vivo* transposition of *mariner*-based elements in enteric bacteria and mycobacteria. Proc. Natl. Acad. Sci. USA *96*, 1645–1650.
28. Sandmann, A., Sasse, F., Hinkelmann, B., and Müller, R. (2002). Search for genes involved in the production of bioactive secondary metabolites in the myxobacterium *Angiococcus disciformis* An d48 using a *mariner* based transposon mutagenesis. In Jahrestagung der Vereinigung für Allgemeine und Angewandte Mikrobiologie (VAAM), C. Schreiber, ed. (Heidelberg, Germany: Spektrum Akademischer Verlag GmbH), pp. 66.
29. Silakowski, B., Kunze, B., and Müller, R. (2001). Multiple hybrid polyketide synthase/non-ribosomal peptide synthetase gene clusters in the myxobacterium *Stigmatella aurantiaca*. Gene *275*, 233–240.
30. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., and Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. Science *252*, 675–679.
31. Schwecke, T., Aparicio, J.F., Molnar, I., König, A., Khaw, L.E., Haydock, S.F., Oliynyke, M., Caffrey, P., Cortes, J., Lester, J.B., et al. (1995). The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. Proc. Natl. Acad. Sci. USA *92*, 7839–7843.

32. Kleinkauf, H., and von Döhren, H. (1997). Products of secondary metabolism. In Biotechnology, H. Kleinkauf and H. von Döhren, eds. (Weinheim, Germany: Verlag Chemie), pp. 277–322.
33. Konz, D., and Marahiel, M.A. (1999). How do peptide synthetases generate structural diversity? Chem. Biol. *6*, R39–R48.
34. Shimkets, L. (1993). The myxobacterial genome. In Myxobacteria II, M. Dworkin and D. Kaiser, eds. (Washington, DC: American Society for Microbiology), pp. 85–108.
35. Challis, G., Ravel, J., and Townsend, C. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem. Biol. *7*, 211–224.
36. Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem. Biol. *6*, 493–505.
37. Kagan, R.M., and Clarke, S. (1994). Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggest a common structure for these enzymes. Arch. Biochem. Biophys. *310*, 417–427.
38. Khaw, L.E., Böhm, G.A., Metcalfe, S., Staunton, J., and Leadlay, P.F. (1998). Mutational biosynthesis of novel rapamycins by a strain of *Streptomyces hygroscopicus* NRRL 5491 disrupted in *rapL*, encoding a putative lysine cyclodeaminase. J. Bacteriol. *180*, 809–814.
39. Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., and Neilan, B.A. (2000). Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. Chem. Biol. *7*, 753–764.
40. Chen, H.W., O'Connor, S., Cane, D.E., and Walsh, C.T. (2001). Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. Chem. Biol. *8*, 899–912.
41. Konz, D., Doekel, S., and Marahiel, M.A. (1999). Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. J. Bacteriol. *181*, 133–140.
42. De Crecy-Lagard, V., Marliere, P., and Saurin, W. (1995). Multienzyme nonribosomal peptide biosythesis: identification of the functional domains catalysing peptide elongation and epimerisation. C. R. Acad. Sci. III *318*, 927–936.
43. Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997). Structural basis for the activation of phenylalanine in the nonribosomal biosynthesis of gramicidin S. EMBO J. *16*, 4174–4183.

Accession Numbers

The nucleotide sequence reported here has been submitted to the EMBL database under accession number AJ620477.