Chemistry & Biology
Article

Biosynthesis of Thuggacins in Myxobacteria: Comparative Cluster Analysis Reveals Basis for Natural Product Structural Diversity

Kathrin Buntin,[1] Herbert Irschik,[2] Kira J. Weissman,[1] Eva Luxenburger,[1] Helmut Blöcker,[3] and Rolf Müller[1,2,*]
[1]Helmholtz Institute for Pharmaceutical Research and Department of Pharmaceutical Biotechnology, Saarland University, P.O. Box 151150, 66041 Saarbrücken, Germany
[2]Microbial Drugs
[3]Department of Genome Analysis
Helmholtz Center for Infection Research, Inhoffenstrasse 7, 38124 Braunschweig, Germany
*Correspondence: rom@mx.uni-saarland.de
DOI 10.1016/j.chembiol.2010.02.013

SUMMARY

The thuggacins are macrolide antibiotics that are active against *Mycobacterium tuberculosis*, the causative agent of tuberculosis. Distinct variants of these structures are produced by the myxobacteria *Sorangium cellulosum* So ce895 and *Chondromyces crocatus* Cm c5, which differ in side chain structure and modification by hydroxylation. We report here a comparative analysis of the biosynthetic gene clusters in these strains, which reveals the mechanistic basis for this architectural diversity. Although the polyketide-nonribosomal peptide cores of the molecules are highly similar, the underlying biosynthetic machineries exhibit an unexpected degree of divergence. Furthermore, the *S. cellulosum* gene cluster contains a crotonyl-CoA reductase (CCR) homolog not present in *C. crocatus*, which likely participates in assembling the unusual hexyl side chain of the So ce895 thuggacins, whereas the distinct hydroxylation pattern may result from variable action of a conserved FMN-dependent monooxygenase. Indeed, inactivation of the monooxygenase gene in *C. crocatus* resulted in production of both mono- and di-deshydroxy thuggacin derivatives, providing direct evidence for the role of this enzyme in the pathway. Finally, integration of a Tn5-derived *npt* promotor upstream of the thuggacin cluster in *C. crocatus* led to a significant increase in thuggacin production.

INTRODUCTION

The World Health Organization has estimated that a third of humanity is latently infected with *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB) (http://www.who.int/mediacentre/factsheets/fs104/en/). TB is the second most lethal infectious disease worldwide and is the leading killer of people infected with HIV (Young et al., 2008). The demand for novel tuberculosis drugs has been fueled by the recent emergence of multidrug- and extensively drug-resistant TB strains, and the challenge of mycobacterial persistance, which together necessitate complex and lengthy treatment regimes (McKinney, 2000). Current first- and second-line antibiotics act against specific components of protein, RNA, DNA, and cell wall synthesis, but are generally ineffective against dormant bacteria, raising interest in medications with alternative modes of action (Zhang et al., 2006). The bacterial respiratory chain is a particularly promising target for new chemothearapeutics, because it appears to be essential for both replicating and nonreplicating mycobacteria (Xie et al., 2005). These considerations have focused attention on the recently discovered thuggacin macrolides, which have been shown to inhibit respiration in several Gram-positive bacteria, including clinical isolates of *M. tuberculosis* (Steinmetz et al., 2007; Irschik et al., 2007).

The thuggacins have been identified from two myxobacterial species, *Sorangium cellulosum* So ce895 and *Chondromyces crocatus* Cm c5. The most striking difference between the set of compounds in each strain is the branching functionality at C2, a methyl in the case of the *C. crocatus* Cm c5 thuggacins (hereafter referred to as Cmc-thuggacins) versus a hexyl side chain in the case of the *S. cellulosum* So ce895 thuggacins (Soce-thuggacins) (Figures 1A and 1B). To the best of our knowledge, the only other secondary metabolite known to incorporate such hexyl-branching is cinnabaramide, from *Streptomyces* strain JS360 (Stadler et al., 2007). A second disparity concerns the pattern of hydroxylation: the Soce-thuggacins incorporate a hydroxyl group at C20 that is not present in the Cmc compounds, whereas the C32 hydroxyl of the Cmc thuggacins is absent in the Soce counterparts. Both strains appear to produce ring-size variants (Soce-thuggacins B and C and Cmc-thuggacin C (compounds **2**, **3** and **6**)) (Figure 1), but these metabolites likely arise from spontaneous rearrangement during isolation, and so are not biologically relevant (Steinmetz et al., 2007). Interestingly, the structural variations do not appear to influence the compounds' inhibitory properties, because Soce-thuggacins A **1** and B **2** show similar activity (Steinmetz et al., 2007).

"Retrobiosynthetic analysis" of the thuggacins strongly suggests their origin from hybrid polyketide synthase (PKS)–nonribosomal polypeptide synthetase (NRPS) systems, which

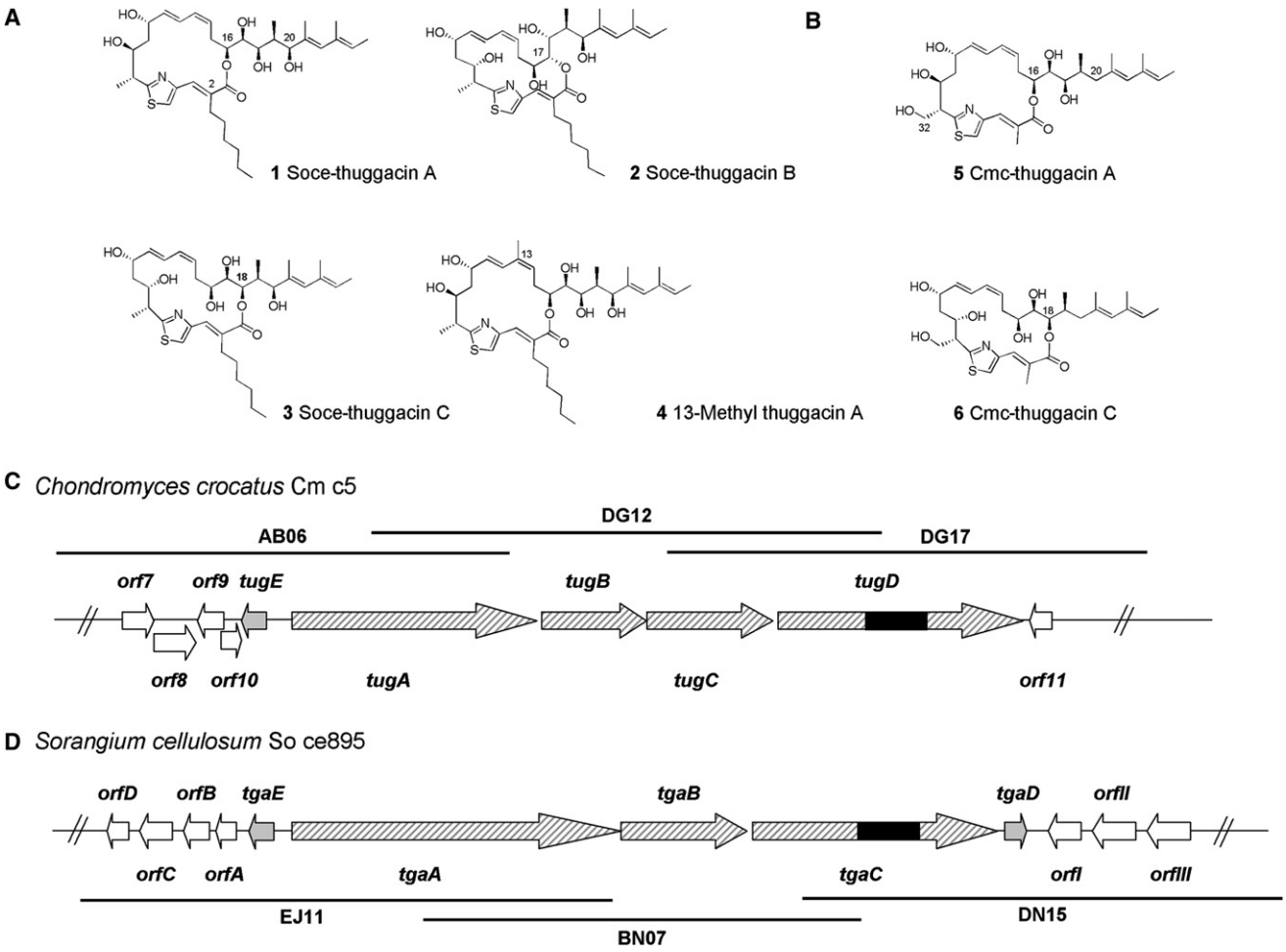

Figure 1. Structures of the Thuggacins Produced by *Sorangium cellulosum* So ce895 and *Chondromyces crocatus* Cm c5 and Organization of the Corresponding Biosynthetic Gene Clusters

(A) Thuggacins derived from *S. cellulosum* So ce895: Soce-thuggacin A (**1**), Soce-thuggacin B (**2**), Soce-thuggacin C (**3**), and the minor metabolite 13-methyl-thuggacin A (**4**). Compounds **2** and **3** are derived from spontaneous rearrangement of **1**.

(B) Thuggacins isolated from *C. crocatus* Cm c5: Cmc-thuggacin A (**5**) and Cmc-thuggacin C (**6**). **6** is derived from **5** during isolation.

(C) Schematic representation of the thuggacin biosynthetic locus in *C. crocatus* Cm c5.

(D) Schematic representation of the thuggacin biosynthetic locus in *S. cellulosum* So ce895. In (C) and (D), the gene clusters are shown in relationship to the identified cosmids. Light gray, genes involved in the biosynthetic pathways; white, genes not involved in biosynthesis; hatches, PKS genes; and black, NRPS gene. The *C. crocatus* cluster lacks a *tgaD* homolog relative to *S. cellulosum* (see also Table 1 and Figure S1).

are common in myxobacterial secondary metabolism (Wenzel and Müller, 2009b; Weissman and Müller, 2008a). PKSs and NRPSs are large multifunctional enzymes that use simple building blocks such as short acyl-CoA thioesters and amino acids to construct molecules of high complexity. These multienzymes have been likened to assembly lines, because they exhibit a modular architecture in which sets of domains or modules accomplish a single round of chain extension. In many cases, the genetic organization of the modules is colinear with the sequence of biosynthetic transformations (Weissman and Leadlay, 2005; Walsh, 2007).

A typical PKS module comprises an acyltransferase (AT) domain, which is responsible for selection of the extender unit, and a ketosynthase (KS) domain, which incorporates the building block into the growing chain via a thioclaisen-like condensation (Staunton and Weissman, 2001). Depending on the specific complement of reductive domains (ketoreductase [KR], dehydratase [DR], and enoyl reductase [ER]) that are optionally present in a module, the β-keto intermediate can undergo varying extents of reduction, resulting in either a β-hydroxyl, α,β double bond, or a fully reduced methylene functionality. Every module additionally includes an acyl carrier protein (ACP) domain to which the nascent chains are tethered in thioester linkage via a phosphopantetheine prosthetic arm, a feature that allows the intermediates to be shuttled efficiently between the active sites of the catalytic domains (Weissman and Müller, 2008b). The analogous core functions in NRPS are the adenylation (A), condensation (C) (or heterocyclization [HC]), and peptidyl carrier proteins (PCP) domains. Optional modifying activities may be also present in NRPS modules,

including oxidase (Ox), epimerase (E), and methyltransferase (MT) domains (Walsh et al., 2001). In both systems, the final product is usually released from the multienzyme complex by a C-terminal thioesterase (TE) domain through intramolecular lactonization or hydrolysis (Kohli and Walsh, 2003; Kopp and Marahiel, 2007). Maturation of the metabolite to its final bioactive form typically involves a series of regio- and stereospecific tailoring reactions such as hydroxylation, methylation, and glycosylation (Rix et al., 2002).

Given the high overall similarity between the Soce- and Cmc-thuggacins, we hypothesized that the minor structural differences could have arisen from small evolutionary adjustments of an ancestral biosynthetic pathway common to the two strains. For example, the C2 branching functionality might have been altered by recombinational replacement of a primordial AT domain with an AT of alternative building block specificity (Jenke-Kodama et al., 2006). To evaluate this idea directly, we identified and sequenced the thuggacin biosynthetic gene clusters in both *S. cellulosum* and *C. crocatus* (Bock et al., 2008) and compared the encoded proteins in detail. Surprisingly, this analysis revealed a much higher than anticipated degree of difference between the PKS/NRPS systems. Here, we interpret this observation in the context of mechanisms by which bacterial genes, and in particular PKS genes, are known to evolve (Zhang, 2003; Santoyo and Romero, 2005; Jenke-Kodama et al., 2005; Jenke-Kodama et al., 2006).

RESULTS AND DISCUSSION

Identification of the Thuggacin Biosynthetic Gene Cluster in *S. cellulosum* So ce895 and *C. crocatus* Cm c5

Because the thuggacins incorporate a thiazole ring that classically arises from oxidative cyclization of cysteine (Walsh et al., 2001), we anticipated that the biosynthetic machinery would include one NRPS module. We therefore attempted to identify the gene clusters in both strains by screening the respective cosmid libraries with probes based on amplified NRPS heterocyclization (HC) domains, coupled with inactivation of the identified HCs. To locate the thuggacin gene cluster in *C. crocatus* Cm c5 (Bock et al., 2008), we generated a 136-member sublibrary of a previously generated cosmid library (Rachid et al., 2006), whose members were positive for PKS and/or NRPS genes. After excluding cosmids that could be correlated to known gene clusters (Buntin et al., 2008; Rachid et al., 2006, 2009), the remaining cosmids were screened for the presence of HC domains, using a PCR-based approach based on degenerate HC domain primers (see Experimental Procedures). Cosmid DG17 yielded an HC amplicon whose sequence differed from that of the HC domain in the ajudazol gene cluster (Buntin et al., 2008). End sequencing of DG17 revealed PKS genes on one end and genes with no predicted function in thuggacin biosynthesis on the other, suggesting that the cosmid contained one end of the target gene cluster. Rescreening of the library with a probe generated against the PKS genes revealed cosmid DG12, which was then used to locate the remainder of the cluster on cosmid AB06. The three cosmids were sequenced on both strands (the nucleotide sequence has been deposited in GenBank under accession number GQ981380), revealing that the thuggacin gene cluster (Cmc-tug) occupies a contiguous stretch of 58.406 kbp on the *C. crocatus* chromosome.

The thuggacin biosynthetic gene cluster in *S. cellulosum* So ce895 was located within a newly generated 2300-clone cosmid library. On the basis of the genome size of the myxobacteria, this cosmid library should represent at least a 10-fold coverage of the genome (Pradella et al., 2002; Goldman et al., 2006; Schneiker et al., 2007). Degenerate HC primers were used in a PCR with *S. cellulosum* So ce895 genomic DNA. Several clones resulting from ligation of the obtained fragment into pCR2.1TOPO were sequenced, revealing an identical product. Because the translated sequence showed 76% sequence identity to the previously identified HC domain from the Cmc-tug cluster, a probe based on the sequence was used for screening the S. *cellulosum* So ce895 chromosomal library. This process resulted in cosmid DN15 with homology to PKS genes on one end, and genes unrelated to thuggacin biosynthesis on the other. As with *C. crocatus*, ''cosmid walking'' was then used to reveal the remainder of the gene cluster on overlapping cosmids BN7 and EJ11. Sequencing of the cosmids successfully identified the *S. cellulosum* thuggacin biosynthetic gene cluster (Soce-tga) on a 56.091 kbp region of the *S. cellulosum* genome. The nucleotide sequence has been deposited in GenBank under accession number GQ981381.

Insertional mutagenesis within the HC domains of both strains resulted in thuggacin-negative phenotypes, confirming the cluster identities in each case (see Figure S1 available online) (Bock et al., 2008). To define the boundaries of both gene clusters, the genes located upstream and downstream of the core PKS/NPRS genes (Table 1) were annotated and compared in silico. Only one gene (*tgaE* for Soce-tga and *tugE* for Cmc-tug) is present in the same orientation relative to the PKS/NRPS genes in both clusters (Figures 1C and 1D). The encoded proteins share 80% identity and 88% similarity, which supports the idea that they play a shared role in the biosynthesis. Additional genes were identified as potential cluster members: *orf11* of Cmc-tug (closest homolog to Orf12: putative protein phosphatase from *Sorangium cellulosum* So ce56 [CAN98414.1, 67% identity and 76% similarity]) and *tgaD* of Soce-tga (closest homolog to TgaD: alcohol dehydrogenase from *Roseosporus caulobacter* species [ZP_02382147.1; 46% identity and 61% similarity]). Because the homology between Orf11 and the putative phosphatase is high and the So ce56 genome does not contain the thuggacin genes (Schneiker et al., 2007), we considered it unlikely that Orf11 played a role in thuggacin biosynthesis. Nonetheless, to confirm this assumption, we inactivated the gene by insertional mutagenesis. As expected, biosynthesis of the thuggacins by the *orf11* mutant was unchanged (data not shown). BLAST analysis of the remaining genes in the sequenced region of Soce-tga revealed very high identity (>90% on the protein level) and synteny with house-keeping genes in the genome of the recently sequenced *Sorangium cellulosum* So ce56 (Schneiker et al., 2007). In addition, these genes do not have obvious roles to play in thuggacin biosynthesis (Table 1). Genome data are not available for an additional *Chondromyces* species to enable comparison with the Cm c5 orfs, but annotation of the remaining six genes in the upstream and downstream regions does not reveal strong candidates for thuggacin pathway members.

Table 1. Proteins Encoded in the PKS/NRPS Portion of the Thuggacin Biosynthetic Gene Clusters in *S. cellulosum* So ce895 and *C. crocatus* Cm c5 and Predicted Function of Non-PKS/NRPS Proteins Present Up- and Downstream of the *C. crocatus* and *S. cellulosum* Thuggacin Biosynthetic Gene Clusters

PKS/NRPS part of the *C. crocatus* thuggacin biosynthetic gene cluster		
Protein	aa	Proposed function (Protein domains with their positions in the sequence)
TugA	6256	PKS: ACP (6–90), KS (108–535), AT (617–943), AT (1060–1387), DH (1427–1708), KR (1728–2181), ACP (2205–2285), KS (2311–2739), AT (2840–3167); DH (3207–3488), KR (3508–3953), ACP (3977–4057), KS (4080–4508), AT (4610–4937), DH (4977–5257), ER (5475–5780), KR (5803–6057), ACP (6078–6157)
TugB	3103	PKS: KS (17–445), AT (538–868), KR (944–1406), ACP (1425–1505), KS (1526–1954), AT (2067–2398), KR (2454–2917), ACP (2935–3015)
TugC	3464	PKS: KS (28–456), AT (575–907), KR (1002–1479), ACP (1500–1580), KS (1602–2030), AT (2144–2475), DH (2515–2808), KR (2826–3278), ACP (3299–3379)
TugD	6611	PKS/NRPS: KS (17–445), AT (549–880), KR (976–1455), ACP (1483–1563), KS (1589–2017), AT (2111–2437), KR (2531–3009), ACP (3030–3110), HC (3127–3558), A (3568–4445), Ox (4111–4267), PCP (4472–4536), KS (4555–4976), AT (5068–5391), DH (5429–5718), KR (5738–6201), ACP (6221–6301), TE (6337–6603)
PKS/NRPS part of the *S. cellulosum* thuggacin biosynthetic gene cluster		
Protein	aa	Proposed function (Protein domains with their positions in the sequence)
TgaA	8552	PKS: ACP (7–91), KS (106–533), AT (615–942), AT (1045–1372), DH (1412–1699), KR (1719–2198), ACP (2219–2299), KS (2325–2753), AT (2862–3189), DH (3229–3525), ACP (3585–3665), KS (3686–4114), AT (4216–4543), DH (4583–4857), KR (4877–5343), ACP (5363–5443), KS (5469–5900), AT (5993–6329), KR (6380–6845), ACP (6882–6962), KS (6983–7411), AT (7504–7848), KR (7890–8366), ACP (8384–8464)
TgaB	3428	PKS: KS (17–445), AT (538–864), KR (960–1438), ACP (1461–1539), KS (1565–1994), AT (2087–2436), DH (2476–2770), KR (2788–3242), ACP (3263–3343)
TgaC	6687	PKS/NRPS: KS (17–445), AT (557–906), KR (1004–1513), ACP (1539–1619), KS (1645–2076), AT (2169–2495), KR (2589–3090), ACP (3111–3191), HC (3208–3639), A (3650–4521), Ox (4118–4310), PCP (4548–4612), KS (4636–5051), AT (5143–5470), DH (5509–5799), KR (5819–6268), ACP (6288–6368), TE (6410–6677)

Proteins encoded upstream of the *C. crocatus* thuggacin biosynthetic gene cluster					
Protein	aa	Proposed function of the homologous protein	Source of the homologous protein	Identity/ similarity, %	Accession number (GenBank)
Orf7	420	Hypothetical protein	*Giardia intestinalis* ATCC 50581	26/41	EET02599.1
Orf8	544	Hypothetical protein	*Gloeobacter violaceus* PCC 7421	29/44	BAC87994.1
Orf9	420	Hypothetical protein	*Sorangium cellulosum* So ce 56	75/85	CAN96487.1
Orf10	202	Hypothetical protein	*Pseudomonas syringae* pv. tomato	29/44	AAO57225.1
TugE	544	Alkanal monooxygenase	*Parvibaculum lavamentivorans* DS-1	52/69	ABS63393.1
Non-PKS/NRPS proteins downstream of the *C. crocatus* thuggacin biosynthetic gene cluster					
Orf11	237	Putative protein phosphatase	*Sorangium cellulosum* So ce 56	67/76	CAN98414.1
Non-PKS/NRPS proteins upstream of the *S. cellulosum* thuggacin biosynthetic gene cluster					
Protein	aa	Proposed function of the homologous protein	Source of the homologous protein	Identity/ similarity, %	Accession number (GenBank)
TgaE	358	Alkanal monooxygenase	*Parvibaculum lavamentivorans* DS-1	52/71	ABS63393.1
OrfA	287	Fumarylacetoacetase family protein	*Sorangium cellulosum* So ce56	90/93	CAN98413.1
OrfB	457	Aldehyde dehydrogenase	*Sorangium cellulosum* So ce 56	93/96	CAN98412.1
OrfC	613	Hypothetical protein	*Sorangium cellulosum* So ce 56	93/97	CAN98411.1
OrfD	406	Ribose ABC transporter periplasmic-binding protein	*Sorangium cellulosum* So ce 56	97/99	CAN98410.1
Non-PKS/NRPS proteins upstream of the *S. cellulosum* thuggacin biosynthetic gene cluster					
TgaD	388	Alcohol dehydrogenase BadC	*Streptomyces roseosporus*	46/61	ZP_02382147.1 (NCBI-reference number)
OrfI	251	DoxX family protein	*Flavobacterium johnsoniae* UW101	60/74	ABQ06016.1
OrfII	445	Hypothetical protein	*Sorangium cellulosum* So ce 56	92/95	CAN98438.1
OrfIII	588	Aspartyl-tRNA synthetase	*Sorangium cellulosum* So ce 56	87/92	CAN98438.1

Indeed, most of the gene products only show homology to hypothetical proteins. Overall, this analysis suggested that the thuggacin cluster in Cm c5 occupies the chromosomal region between *tugE* and *tugD*, whereas the cluster in So ce895 spans the region between *tgaE* and *tgaD* (Figures 1C and 1D).

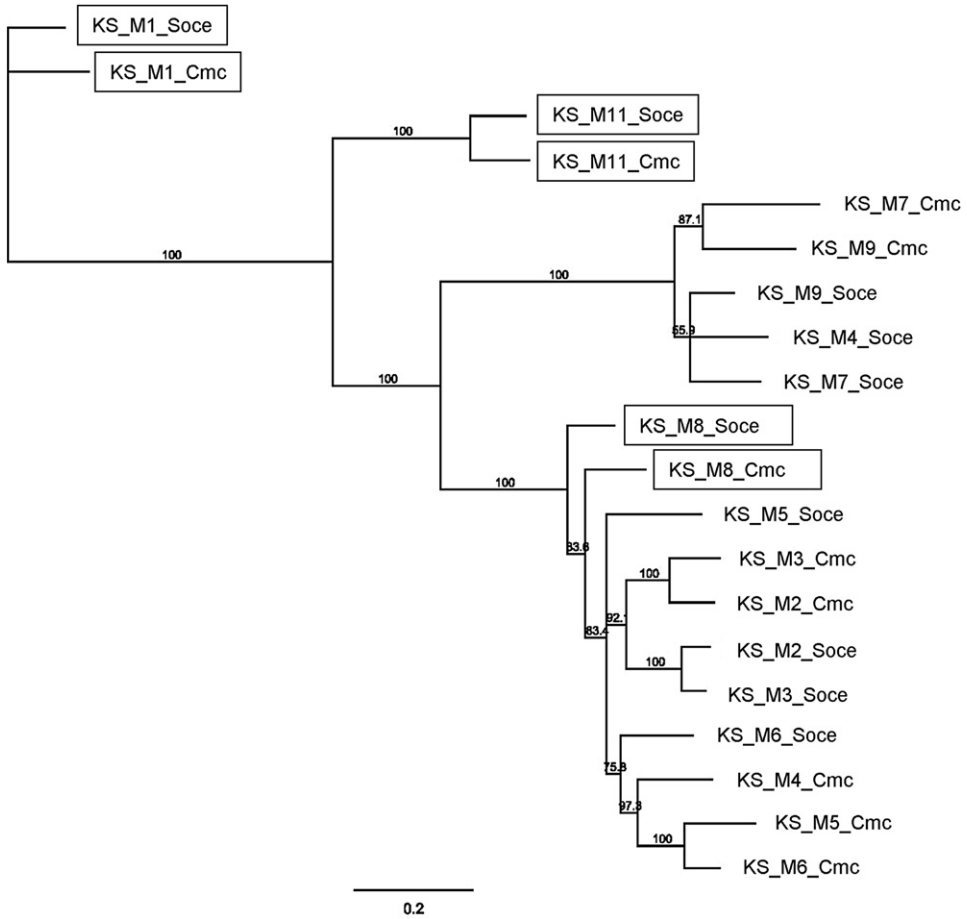

Figure 2. Phylogenetic Tree Analysis of KS Domains from Soce-tga and Cmc-tug

Phylogeny of KS domains inferred from Bayesian estimation with bootstrapping. Branch length indicates the number of inferred amino acid changes per position. KSs whose closest homolog is the corresponding KS in the other gene cluster are boxed. In each case, the source strain and module are indicated. Further sequence analysis is provided in Figure S2.

Sequence and Evolutionary Analysis of the Thuggacin Biosynthetic Gene Clusters

On the basis of phylogenetic analysis of specific domain types, Dittmann and co-workers have developed a model for the evolution of multimodular PKS (Jenke-Kodama et al., 2005, 2006). Their data suggest that PKSs evolve from one or a few ancestral modules by repeated cycles of gene duplication. Sequences differences then accumulate within the modular products as a result of gene conversion (Santoyo and Romero, 2005), a recombinational process in which information is transferred in nonreciprocal fashion between homologous sequences, as well as recombinational exchange of domains both within and between gene clusters in a given organism (Jenke-Kodama et al., 2006). Sequence evidence also suggests that whole pathways can be passed between bacterial species by horizontal gene transfer (HGT) and that such clusters can subsequently advance by gene duplication.

Given that the So ce895 and Cm c5 thuggacin core structures differ solely in the structure of the branching unit at C2 and the presence of a hydroxyl functionality at C20 in the So ce895 compounds (Figures 1A and 1B), we anticipated that the two strains may have acquired a shared pathway essentially intact, either by vertical gene transfer or HGT. We were therefore surprised to find several significant differences between the two clusters. For example, although the gene sets exhibit the high GC content that is characteristic of the myxobacteria, the precise percentage GC differs (Cmc-tug, 70%; Soce-tga, 75.8%), and although both systems incorporate a total of twelve modules (one loading module and eleven chain extension modules), the Cmc-tug PKS and NRPS modules are encoded by four genes (*tugA–tugD*), whereas the biosynthetic machinery within Soce-tga is located on three (*tgaA–tgaC*). Similarly, there are several discrepancies in the domain composition of modules 2 and 3 of Cmc TugA and Soce TgaA: module 2 of Soce TgaA lacks the KR domain present in module 2 of Cmc TugA, whereas module 3 is missing ER functionality present in the corresponding module.

To try to discover the origin of these differences, we performed phylogenetic analysis of the KS domains in both pathways (Figure 2); because the KSs are the most conserved domains within modular PKS and are not subject to recombinational swapping (Jenke-Kodama et al., 2006), they most clearly reflect

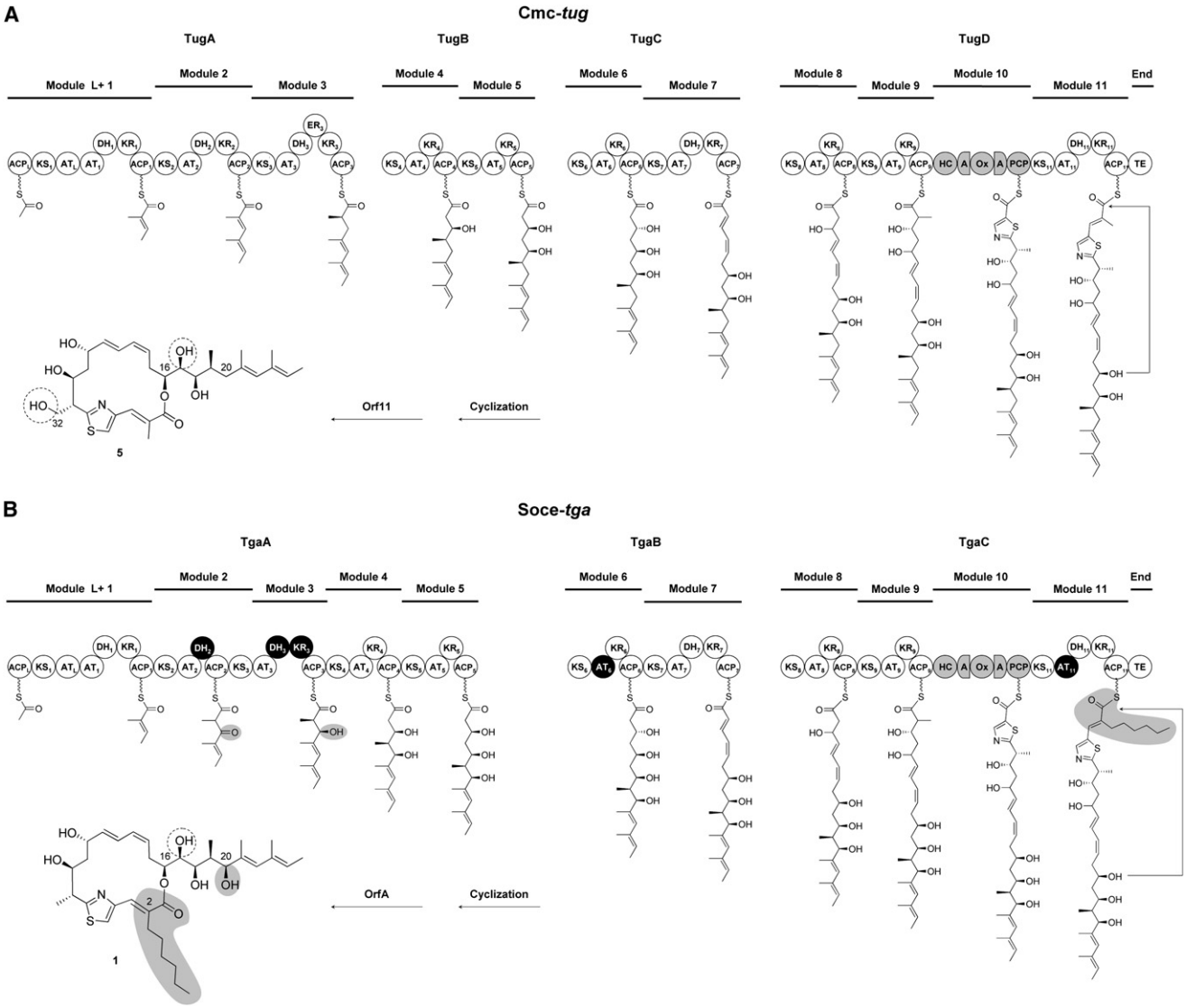

Figure 3. Models for Thuggacin Biosynthesis

(A) Schematic representation of thuggacin biosynthesis in *C. crocatus* Cm c5.

(B) Schematic representation of thuggacin biosynthesis in *S. cellulosum* Soce895. In PKS portion (white): KS, ketosynthase; AT, acyl transferase; KR, ketoreductase; DH, dehydratase; ER, enoyl reductase; ACP, acyl carrier protein; and TE, thioesterase.

In NRPS portion (gray): HC, heterocyclization; A, adenylation; Ox, oxidation; and PCP, peptidyl carrier protein. Differences between the domain compositions of the biosynthetic machineries are indicated in black, whereas the resulting structural discrepancies are shown in gray. The dashed lines indicate the hydroxyl functionalities added following release from the assembly lines by the conserved FMN-linked monooxygenase. In cases where hydroxyl functionalities generated during chain extension are subsequently lost by dehydration, the stereochemistry of the original groups has been predicted on the basis of KR motifs correlating with the direction of ketoreduction (Figure S3) (Caffrey, 2003; Reid et al., 2003).

the evolutionary history of a specific system (Jenke-Kodama et al., 2005). The closest homolog to the KS domains of modules 1, 8, and 11 is the corresponding KS in the other cluster. In contrast, the domains most similar to the remaining KS are found within the same cluster. Interestingly, the KS domains of module 4 show the greatest evolutionary divergence from each other (58% sequence identity, whereas the intercluster pairwise identity between the other KS domains is $\geq$76%). In the Cmc-tug system, KS_4 is located at an intersubunit interface and so may play a role in protein-protein recognition, whereas the corresponding KS in Soce-tga is present in the middle of a subunit (Figure 3). The different functional requirements imposed by these alternative locations likely explain the significant differences between the KS domains.

On the basis of established modes of bacterial evolution (Zhang, 2003; Santoyo and Romero, 2005), two alternative models can be proposed which account for the observed phylogenetic relationships between the KS domains. In the first, both Cm C5 and So ce895 (or their ancestors), members of the myxobacterial suborder Sorangiineae, independently acquired an identical cluster by HGT. Intracluster homogenization of KS sequences may then have occurred by gene conversion

(Santoyo and Romero, 2005). Assuming that the So ce895 cluster has diverged further from the original because it is the less colinear of the two systems, the remaining discrepancies in architecture could have arisen by the deletion of reductive activities and the mutation/exchange of AT domains (Jenke-Kodama et al., 2006), as well as a gene fusion event evolving modules 3 and 4. These modifications would in turn have influenced the direction of KS evolution (i.e., KS_4), and required a substantial change in programming in modules 2 and 3 (*vida infra*).

Alternatively, and more speculatively, both myxobacteria once contained an identical smaller cluster (perhaps consisting of modules 1, 8, and 11). This shared gene set may then have been differentially expanded by module duplication to generate the present-day pathways; the original KSs still exhibit highest cross-cluster homology, whereas those derived from internal gene duplication most resemble other KSs within the same system. In this model, although the evolutionary histories, and therefore the detailed architectures of the two clusters diverged, the pathways ultimately converged on a closely similar metabolite structure, a candidate example for divergent/convergent evolution.

It is also interesting that many of the closest homologs in the public database to the thuggacin KS domains are *Streptomyces* in origin. In contrast, performing the same analysis with KS domains from other myxobacterial systems—for example, those of the stigmatellin (Gaitatzis et al., 2002) and myxothiazol (Silakowski et al., 1999) gene clusters—returns most significant hits to cyanobacterial or other myxobacterial sequences (Figure S2). This observation suggests that the progenitor sequence of the thuggacin cluster may have been acquired from a streptomycete by HGT. Analysis of docking domain sequences from the two clusters further supports this idea. Docking domains are sequence elements at the extreme termini of PKS and NRPS subunits, which help to mediate protein-protein recognition at the multienzyme interfaces (Broadhurst et al., 2003; Richter et al., 2007; Buchholz et al., 2009). PKS docking domains fall into three distinct evolutionary classes, two of which predominate in *Streptomyces*, whereas the third is characteristic of myxobacteria and cyanobacteria (Thattai et al., 2007). The docking domain sequences within the thuggacin systems are clearly of the *Streptomyces* type (Figure S2), consistent with our HGT model. Taken together, these data represent the first direct evidence for an HGT event between streptomycetes and myxobacteria (Jenke-Kodama et al., 2005).

Biosynthesis of the Thuggacin Backbones

Inspection of the domain complement within the multienzymes allows us to make a detailed proposal for assembly of the thuggacin core structures. Both of the pathways include the required number of modules to incorporate 10 polyketide building blocks and a single amino acid (Figure 3). Thus, unlike many other myxobacterial systems (Wenzel and Müller, 2007), the thuggacin mixed PKS/NRPSs follow the colinear paradigm for assembly line biosynthesis. In an increasingly recognized variant on standard domain organization (Wenzel and Müller, 2007), biosynthesis is initiated by a mixed loading/chain extension module. In this case, the loading acyltransferase (AT_L) is anticipated to transfer the starter unit acetate onto the first acyl carrier protein (ACP_L), whereas AT_1 is likely to load ACP_1 with the extender unit methylmalonate (Wilkinson et al., 2001). In support of these functional assignments, analysis of the AT domains in silico shows that both AT_L domains lack the conserved active site arginine residue, which has been implicated in selection of dicarboxylic acid extender units (Figure S3) (Rangan and Smith, 1997; Liou et al., 2003; Long et al., 2002). In contrast, the AT_1s each contain the critical arginine and are predicted to show specificity for methylmalonate (Del Vecchio et al., 2003; Yadav et al., 2003; Haydock et al., 1995).

The first condensation is catalyzed by the sole KS (KS_1) located within the mixed module. The second and third rounds of chain extension are then carried out by the subsequent two modules, but the detailed mechanisms must differ between the two systems, because the modules do not house the same complement of catalytic domains. Module 2 of Cmc TugA incorporates the optional reducing domains KR, DH, whereas module 3 contains a full reductive loop consisting of KR, DH, and ER activities. In both cases, the domain composition is precisely that required to establish the observed structure of Cmc-thuggacin at the appropriate positions. In contrast, module 2 of Soce TgaA is missing the KR domain, although an apparently functional DH is still present (as judged on the basis of its conserved active site motif; Donadio and Katz, 1992) (Figure S3). Because the DH of module 2 cannot operate on the acyl chain in the absence of ketoreduction, the nascent keto group must not be processed during the chain extension cycle within this module (Figure 3B). To establish the required double bond functionality, the unreduced intermediate could be transferred to the ACP domain of module 3, without undergoing chain extension (Tang et al., 2004b). Reduction by KR_3 followed by DH_3-catalyzed dehydration, would then yield the required *trans* double bond. The need for domains to act out of sequence has been noted for several other myxobacterial PKS systems (Tang et al., 2004a; Buntin et al., 2008; Frank et al., 2007; Gaitatzis et al., 2002), but this is, to our knowledge, the first report of both a KR and a DH domain operating in this way. The intermediate would then be passed upstream to the module 3 KS for chain extension. At this stage, only KR_3 would act on the resulting β-keto group to give the C20 hydroxyl characteristic of the Soce-thuggacins, although the basis for programming DH_3 to operate on the triketide and not on the tetraketide is unclear. Alternatively, both the KR_3 and DH_3 domains are active in the third round of chain extension, but the resulting double bond is rehydrated at a later stage.

Inspection of the thuggacin structures predicts that biosynthesis in both systems should continue with three rounds of chain extension using malonate as extender unit. However, analysis of AT sequence motifs correlated with substrate specificity (Haydock et al., 1995; Yadav et al., 2003) suggests that AT_6 of Soce TgaB should instead be specific for methylmalonate (Figure S3). In fact, *S. cellulosum* does produce minor amounts of a thuggacin derivative incorporating a 13-methyl group (**4**, Figure 1A), but AT_6 nonetheless appears to show a preference for selection of malonate. This observation adds weight to evidence indicating that AT residues in addition to the identified sequence motifs are involved in determining substrate specificity (Del Vecchio et al., 2003; Petković et al., 2008). An alternative explanation, however, is that intermediates incorporating methylmalonate at this position stall on the assembly line because

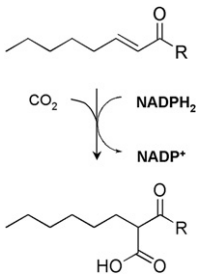

Figure 4. Model for the Generation of the 2-Carboxy-Octanoyl Extender Unit

TgaD catalyzes reductive carboxylation of either the fatty acid oxidation intermediate 2-octenoyl-CoA (R = SCoA) to yield the extender unit 2-carboxy-octanoyl-CoA, or of the fatty acid biosynthesis intermediate 2-octenoyl-ACP (R = SACP), generating 2-carboxy-octanoyl-ACP (see also Figure S4).

of inefficient processing by downstream modules, causing an overall bias in the product mixtures toward metabolites derived from malonate.

Modules 6 in both systems lack the DH activity required to generate the observed C13–C14 *cis* double bond. Thus, like DH_3, DH_7 may also act iteratively: on the octaketide generated in its own module, as well as the heptaketide from the previous module. As observed for epothilone biosynthesis, the first dehydration results in a *cis* double bond, and the second, a *trans* double bond (Tang et al., 2004b). Correspondingly, the precursor hydroxyl groups generated by modules 6 and 7 are predicted to exhibit the appropriate A-type or B-type stereochemistry (Bock et al., 2008) required for direct generation of a *cis* or a *trans* double bond via *syn* dehydration of the β-hydroxy thioester (Figure S3) (Caffrey, 2003). Despite the strong precedent for this proposal, it is impossible at this stage to exclude the participation of a DH that is not a domain within the PKS, in establishing the *cis* double bond.

Chain extension continues with two PKS modules, followed by the lone NRPS module within the assembly line. The NRPS catalyzes the incorporation of cysteine, which is further processed to a thiazole ring by the combined action of the HC and Ox domains. As has been observed previously in the tubulysin (Sandmann et al., 2004) and myxothiazol (Silakowski et al., 1999) systems, the Ox domain is located between the conserved A8 and A9 motifs of the A domain. The final PKS module, module 11, is responsible for the most notable structural difference between the Soce- and the Cmc-thuggacins, the branching functionality at C2. Consistent with the presence of a C2 methyl in the Cmc-thuggacins, the AT_{11} (Cmc TugD) incorporates recognition motifs that correlate with specificity for methylmalonate as extender unit (Figure S3). In contrast, no clear prediction can be made for AT_{11} of Soce TgaC on the basis of these residues. Although the KS phylogenetic analysis implies that module 11 was present in the shared ancestor of both thuggacin gene clusters, the overall homology between the present-day AT_{11}s is only 64% (versus 84% for the KS_{11} domains within the same module). This observation suggests that one of the two AT domains was introduced into the ancestral module by homologous recombination with another AT sequence within the strain (Jenke-Kodama et al., 2006). Taken together, these data are consistent with the idea that Soce AT_{11} accepts an uncommon building block as substrate, ultimately giving rise to the hexyl side chain of the Soce-thuggacins.

To ascertain the origin of the hexyl functionality, we considered known pathways for the biosynthesis of several unusual PKS extender units. Ethylmalonyl-CoA, for example, has been shown to derive from crotonyl-CoA by NADPH-dependent reductive carboxylation, a discovery that prompted functional revision of a family of previously identified crotonyl-CoA reductase (CCR) enzymes as dual function catalysts (Erb et al., 2007). On the other hand, hydroxymalonate and its methylated analog, methoxymalonate, are biosynthesized from 1,3-biphosphoglycerate (Wenzel et al., 2006; Dorrestein et al., 2006) in a multistep reaction performed on an ACP-bound species (Wu et al., 2000). Thus, the proximal donor of the mature extender units to the AT domain is an ACP. This information led us to consider two possible sources for the hexyl side chain of the Soce-thuggacins, 2-carboxy-octanoyl-CoA and 2-carboxy-octanoyl-ACP (Figure 4). Consistent with this proposal, inspection of the Soce AT_{11} sequence shows that it incorporates the conserved Arg residue correlated with recognition of carboxylated substrates. Notably, the octenoyl-CoA and octenoyl-ACP precursors to these proposed building blocks are intermediates respectively in fatty acid degradation and biosynthesis, and so would be readily available in the cell. Labeling of this portion of the molecule in feeding experiments with ^{13}C-labeled precursors is also consistent with a fatty acid origin for the extender unit (Steinmetz et al., 2007).

To discriminate between the two possible building blocks, we focused our attention on *tgaD*, the only gene in the cluster that might reasonably participate in extender unit biosynthesis. Although its protein product exhibits highest homology to alcohol dehydrogenases (e.g., 46% I, 61% S to BadC from *Streptomyces roseosporus*), it also shows similarity to crotonyl-CoA carboxylase/reductase enzymes (e.g., 23% I, 40% S to the bonafide CCR of *Caulobacter* species K31) (Figure S4). We also detected a clear NADPH-binding motif. On the basis of this information, we hypothesized that TgaD generates 2-carboxy-octanoyl-CoA from reductive carboxylation of octenoyl-CoA (Figure 4). However, as the sequence of TgaD appears to be undergoing an evolutionary transition from a dehydrogenase to a reductase/carboxylase, it may also carry out dehydrogenation of octanoyl-CoA to furnish its own octenoyl-CoA substrate.

To provide support for the involvement of TgaD in the pathway, we inactivated *tgaD* by insertional mutagenesis. Because *tgaD* is located at the boundary of the gene cluster and no other gene is cotranscribed with it (Figure 1D), we could rule out any polar effects on the remaining thuggacin biosynthetic genes. The resulting Soce-*tgaD*$^-$ mutant exhibited a Soce-thuggacin-negative phenotype, supporting its proposed role in the biosynthesis (Figure S4), although we cannot yet exclude other explanations for this result, such as a role for TgaD in regulation. We also screened the extracts for Soce-thuggacin analogs, incorporating alternative functionality at C2 (e.g., methylene, methyl), which might arise if Soce AT_{11} showed promiscuous activity toward alternative extender units in the absence of its preferred substrate. However, no such derivatives were

detected. Thus, AT_{11} appears to be a highly specific domain, consistent with the absence of natural C2 variants in the native host strain. Taken together, these data strongly suggest that TgaD is the newest member of a growing family of CCR enzymes that furnish unusual carboxylated extender units to PKS systems. This family includes the recently characterized SalG, which generates both propylmalonyl-CoA and chloroethylmalonyl-CoA from the respective unsaturated precursors in the salinosporamide A pathway of the marine bacterium *Salinospora tropica* (Eustáquio et al., 2009; Liu et al., 2009).

The last step in both pathways is chain termination catalyzed by the C-terminal TE domains of Cmc TugD and Soce TgaC. Release in each case occurs by attack of the internal C16 hydroxy nucleophile on the TE-bound ester (Figure 3).

Post–Assembly Line Modifications

According to classic rules of assembly-line biosynthesis, at least one of the hydroxyl functionalities in each of the thuggacins is likely to derive from post–assembly line processing. This is predicted to be the case for the C17 hydroxyl of the Soce-thuggacins (**1**–**4**), and both the C17 and C32 hydroxyls of the Cmc-thuggacins (**5**, **6**) (Figures 1A and 1B). Despite the differing numbers of oxidations taking place in the two pathways, the gene clusters contain a single candidate for performing these transformations—*tugE* of Cmc-tug and *tgaE* of Soce-tug—located upstream of the core PKS/NRPS genes (Figures 1C and 1D). The closest homologs to the translated sequences that show high mutual homology (80% identity and 88% similarity) are an FMN-dependent alkanal monooxygenase from *Parvibaculum lavamentivorans* DS-1 (ABS63411.1; 52% identity and 71% similarity) and a luciferase-like monooxygenase from *Burkholderia xenovorans* LB400 (ABE36828.1; 46% identity and 65% similarity), consistent with a shared function as thuggacin hydroxylases. In common with other myxobacterial pathways (Rachid et al., 2009), however, neither cluster contains a gene for a NAD(P)H:flavin oxidoreductase (van Berkel et al., 2006), and so this partner enzyme is presumably encoded elsewhere in the genome.

To provide direct evidence for the putative function of the monooxygenases, we aimed to inactive the *tugE* gene in *C. crocatus*. The strain is genetically more tractable than *S. cellulosum*, and targeting *tugE* also allowed us to probe the role of the lone monooxygenase in catalyzing the two observed oxidation reactions. As expected, analysis by high-performance liquid chromatography-mass spectrometry (HPLC-MS) of Cmc-*tugE*$^{-}$ relative to wild-type *C. crocatus*, demonstrated that the mutant no longer produced Cmc-thuggacin A (**5**) and Cmc-thuggacin C (**6**). Furthermore, reanalysis of the mutant extracts at high resolution using LC-coupled Fourier transform-Orbitrap mass spectrometry revealed a new compound, with an accurate mass of *m/z* $[M+H]^+$ = 530.29359 (retention time [r.t.] = 6.9 min) (Figures 5A and 5B). This mass is consistent with the molecular formula $C_{30}H_{44}NO_5S$ (calcd. $[M+H]^+$ = 530.29347, Δ = 0.2 ppm), which corresponds to a di-deshydroxy derivative (**7**) of Cmc-thuggacin, lacking the OH functionality at both C17 and C32. We also detected small amounts (~10-fold lower) of a second novel metabolite **8** with exact mass *m/z* $[M+H]^+$ = 546.28876 (r.t. = 6.92 min) (Figure 5C). The predicted molecular formula for this second compound, $C_{30}H_{44}NO_6S$ (calcd. $[M+H]^+$ = 546.28839, Δ = 0.7 ppm), tallies with one of two possible mono-deshydroxy derivative of Cmc-thuggacin (i.e., missing the OH at either C17 or C32) (Figure 5A). Because these molecules were anticipated to be intermediates on the pathway to the mature thuggacins, we next reanalyzed the wild-type Cm c5 extract using high-resolution MS for the presence of the same metabolites. We indeed detected both the di-deshydroxy compound **7** (Figure 5D) and metabolite **8** (Figure 5E), at approximately half the yields of those in the mutant. We also discovered a second compound **9** with essentially the same mass as **8** (*m/z* $[M+H]^+$ = 546.28842), but with a distinct retention time (7.6 min) (Figure 5E). Thus, we hypothesized that **9** would be the alternative mono-deshydroxy Cmc-thuggacin. Comparison of the relative yields of all thuggacins from the wild-type (**5**–**9**) and the mutant (**7**, **8**), shows that overall production by the mutant was reduced by ~10-fold. One possible explanation for this unexpected observation is that integration of the Cmc *tgaE* disruption plasmid may have altered the efficiency of transcription of the adjacent Cmc *tgaA*–*tgaD* structural genes, but this mechanism remains to be demonstrated directly.

Because the low production yields of the new derivatives were not sufficient to enable full structural proof by NMR, we aimed to assign the compound structures using high-resolution, tandem mass spectrometry (MS^2 and MS^3). On the basis of molecular formulae derived from accurate masses, we were able to propose structures for the majority of fragment ions arising from **7** (Figure 5F). The fragmentation pattern of **7** from both wild-type and mutant extracts was identical and consistent with the predicted absence of hydroxyls at both C17 and C32. Despite the fact that compounds **8** and **9** give rise to unique fragmentation patterns, we were unable to definitively assign the site of the monohydroxylation in each case based on analysis of possible fragment structures (Figure S5). Nonetheless, a strong clue as to the location of the hydroxyls is provided by consideration of the post-PKS chemistry in the *S. cellulosum* strain.

The Soce-thuggacins incorporate a single hydroxyl group at C17 (Steinmetz et al., 2007). The cluster contains a gene with high homology to Cmc *tugE*, Soce *tgaE*, whose translated sequence reveals no evidence for inactivating mutations. Given that Cmc TugE is a functional hydroxylase, we predict that Soce TgaE is also active and responsible for the C17 hydroxylation. Indeed, if Soce TgaE were instead an exclusive C32 hydroxylase, then it would be completely inactive toward the deshydroxy Soce-thuggacin precursor. Taken together, these data suggest that Cmc TugE must also act, at minimum, at C17.

In light of these considerations, two mechanisms can be proposed that account for the observed pattern of hydroxylation in the wild-type and mutant Cm c5 strains. One possibility is that TugE hydroxylates **7** at position C17 to generate **9** and then performs a second hydroxylation at C32 to yield the mature thuggacins. These reactions do not go to completion, explaining the presence of both **7** and **9** in the wild-type strain. This order of transformations is consistent with the finding that approximately equal amounts of **8** are present in the Cmc-*tugE*$^{-}$ and wild-type strains (i.e., if Cmc TugE were capable of generating **8** by C32 hydroxylation of **7**, followed by its C17 hydroxylation, then **8** would not accumulate in the wild-type).

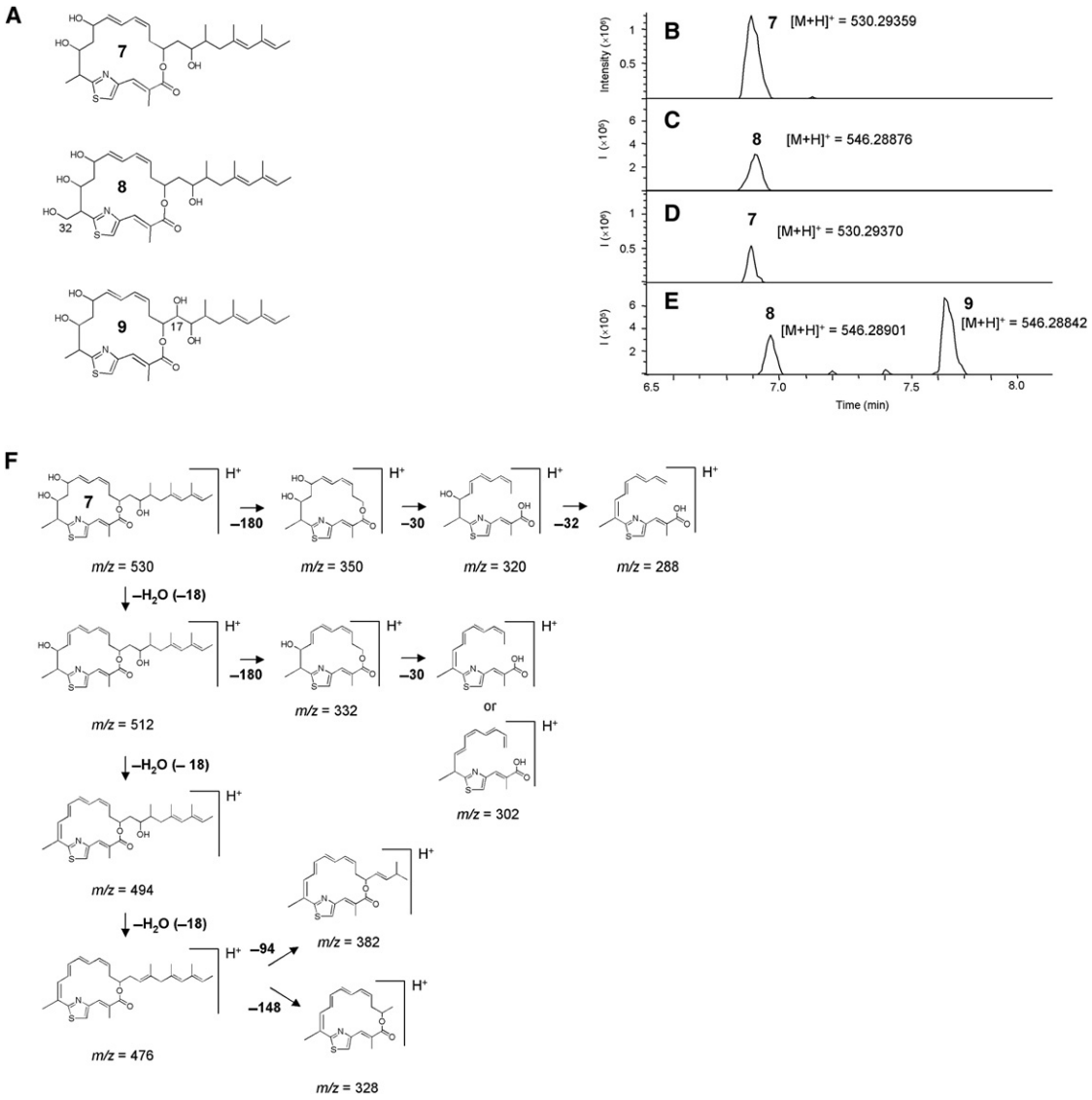

Figure 5. Identification of Mono-deshydroxy and Di-deshydroxy Cmc-thuggacins in *C. crocatus* Cm c5

(A) Proposed structures of the novel metabolites, di-deshydroxy Cmc-thuggacin **7**, and the mono-deshydroxy Cmc-thuggacins **8** and **9** (see main text). **8** lacks the C17 hydroxyl of the native Cmc-thuggacins, whereas the C32 hydroxyl is missing in metabolite **9**.

(B) HPLC-MS analysis of Cmc-*tugE*$^{-}$. Shown is the extracted ion chromatogram (EIC) of *m/z* $[M+H]^+$ = 530.3). The extract contains a compound (r.t. 6.9 min) with a mass consistent with di-deshydroxy-Cmc-thuggacin A **7** (calcd. *m/z* $[M+H]^+$ for $C_{30}H_{44}NO_5S$ = 530.29347; found: 530.29359).

(C) HPLC-MS analysis of Cmc-*tugE*$^{-}$. Shown is the EIC of *m/z* $[M+H]^+$ = 546.3. A metabolite was detected (r.t. 6.92 min) with a mass consistent with mono-deshydroxy-Cmc-thuggacin **8** (calcd. *m/z* $[M+H]^+$ for $C_{30}H_{44}NO_6S$ = 546.28839; found: 546.28876).

(D) HPLC-MS analysis of *C. crocatus* wild-type. Shown is the EIC of *m/z* $[M+H]^+$ = 530.3. Compound **7** was detected.

(E) HPLC-MS analysis of *C. crocatus* wild-type. Shown is the EIC of *m/z* $[M+H]^+$ = 546.3. Compound **8** was detected, as well as a candidate for the second mono-deshydroxy Cmc-thuggacin **9** (r.t. 7.6 min; calcd. *m/z* $[M+H]^+$ for $C_{30}H_{44}NO_6S$ = 546.28839; found: 546.28842).

(F) Proposed fragmentation pattern for di-deshydroxy-Cmc-thuggacin **7**, based on fragment accurate masses observed during MS^n analysis. In some cases, the order of fragmentation has been hypothesized on the basis of chemical logic. The data shown in (F) were obtained on a Thermo LTQ Orbitrap Hybrid FT mass spectrometer. Proposed fragmentation patterns for compounds **5**, **8** and **9** are shown in Figure S5.

Thus, in this model a second, promiscuous oxygenase is responsible for production of **8**. Alternatively, Cmc TugE catalyzes only the C17 hydroxylation, whereas the C32 modification is performed by a dedicated enzyme encoded elsewhere in the genome. Such split cluster organization has precedent in other myxobacterial pathways (Sandmann et al., 2004; Kopp et al., 2005; Perlova et al., 2006), although it remains relatively rare. As in the previous mechanism, the second enzyme would normally act on **9**, but would also show moderate activity toward **7**, generating **8**.

Nonetheless, it remains unclear why the Soce-thuggacins are not hydroxylated at C32. On the basis of the high sequence identity between Soce TgaE and Cmc TugE, if Cmc TugE is also a C32 hydroxylase, Soce TgaE would be expected to exhibit

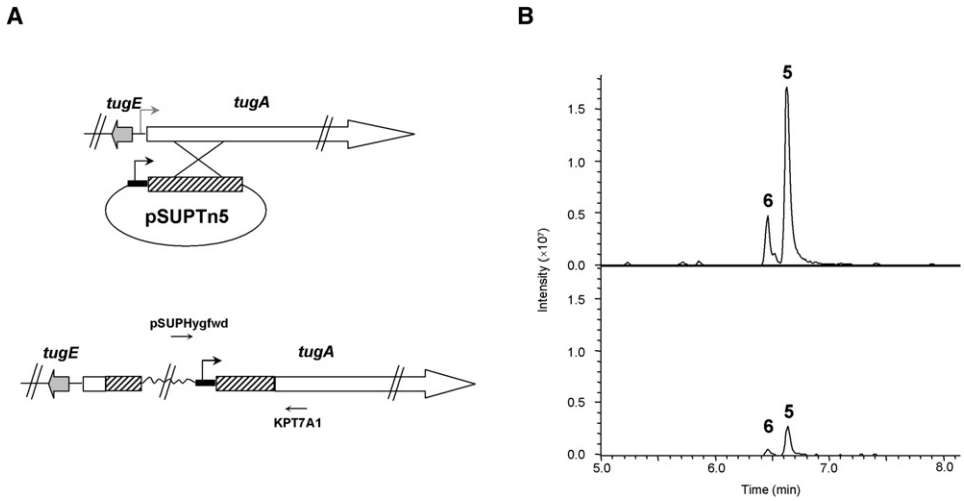

Figure 6. Insertion of the *npt* Promoter into *C. crocatus* Cm c5

(A) Schematic of the promoter insertion experiment. The promoter replacement construct contained the strong *npt* promoter (black), followed by a 1 kbp internal fragment of Cmc *tugA*. Correct integration of the plasmid resulted in insertion of the promoter upstream of *tugA*. The vector backbone sequence is indicated by a wiggly line, and the PCR primer binding sites used for genetic verification are shown.

(B) Analysis of extracts of the promoter insertion mutant (upper panel) and the *C. crocatus* wild-type (lower panel). In both cases, the extracted ion chromatograms (EIC) of *m/z* $[M+H]^+$ = 562.2830 are shown. Peaks corresponding to Cmc-thuggacin A (**5**) (r.t. 6.6 min) and Cmc-thuggacin C (**6**) (r.t. 6.4 min) are indicated.

the same regiospecificity. However, subtle differences within the Cmc TugE and Soce TgaE active sites may only allow the Soce-thuggacins to adopt an orientation that permits C17 hydroxylation. Alternatively, a single orientation may be imposed by the presence of the hexyl side chain and or C20 hydroxyl group which are unique to the Soce-thuggacins. If the C32 hydroxylation in *C. crocatus* is performed instead by a second enzyme, the lack of this reaction may be explained by the absence of a corresponding gene in *S. cellulosum*, or alternatively, by the homolog's inactivity toward the variant substrate. In principle, studies in vitro with recombinant Cmc TugE and Soce TgaE could resolve the ambiguities evident in this interpretation, but the minute quantities of the required biosynthetic substrates presently preclude such experiments. Nonetheless, our data show conclusively that Cmc TugE is a thuggacin hydroxlase, likely responsible for installing the functionality at C17.

Promoter Insertion into Cmc-tug

The thuggacins are produced at relatively low yields in comparison to the other secondary metabolites of *C. crocatus*. Because all aspects of natural product development depend on obtaining sufficient amounts of the target compounds, we intended to apply molecular methods for increasing thuggacin production titers. For this, we aimed to replace the natural promoter directly upstream of Cmc-tug with a strong constitutive promoter. The overexpression of single genes (Rachid et al., 2007; Meiser and Müller, 2008) or multigene transcriptional units by promoter exchange has already been performed successfully in other myxobacterial strains leading to increased secondary metabolite formation (Fu et al., 2008; S. C. Wenzel and R.M., unpublished results). Because the Tn5-derived *npt* promotor was shown to efficiently drive expression of a 30 kb lipopeptide biosynthetic operon in *M. xanthus* (S. C. Wenzel and R.M., unpublished results), we inserted this promoter upstream of the Cmc-tug operon. The promoter replacement construct was engineered by overlap PCR to contain the *npt* promoter and its associated ribosome-binding site, followed by a 1 kbp internal fragment of *tugA* (Figure 6A). Correct insertion of the *npt* promoter within the genome was confirmed by PCR using primers designed against both insert and flanking genome regions, followed by sequencing of the obtained products.

The *C. crocatus* wild-type and mutant strains were cultivated in triplicate and analyzed by high-resolution mass spectrometry. Relative yields of the thuggacins in each case were compared to that of ajudazol A, the major metabolite of the strain, on the assumption that its biosynthesis would not be altered by modification of the thuggacin gene cluster. Based on this internal control, integration of the *npt* promoter resulted in a 5–7-fold increase in thuggacin production (Figure 6B). Thus, this experiment provides further demonstration of the potential for using promoter exchange to influence secondary metabolite yields in native myxobacterial hosts (S.C. Wenzel, unpublished results). Such an approach may ultimately prove fruitful for up-regulating biosynthesis from some of the many natural product gene clusters in myxobacteria, for which no product is currently known (Krug et al., 2008; Wenzel and Müller, 2009a).

SIGNIFICANCE

The multienzyme polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) exhibit an assembly line architecture in which sets of enzymatic domains are organized into functional modules. The direct correspondence between the complement of domains in a particular module and the structure of its chain-extension product has encouraged attempts to produce metabolites of differing but predictable structure, by targeted alteration of specific domains. Our analysis of thuggacin biosynthesis

in the myxobacteria *Sorangium cellulosum* So ce895 and *Chondromyces crocatus* Cm c5, reveals that several enzymatic configurations can achieve a common structural end. Variations in the detailed evolutionary pathways to the modern gene clusters, as well as the unequal acquisition of accessory enzymatic functions, appear to underlie the small structural variations between the So ce895 and Cm c5 thuggacins. For example, only the *S. cellulosum* gene cluster incorporates a novel crotonyl-CoA reductase/carboxylase (CCR) homolog, which likely participates in assembling the unusual hexyl side chain of its thuggacin variants. This finding is encouraging for attempts to genetically engineer the biosynthesis of a specific natural product analog, because it suggests that more than one enzymatic solution is possible. We further demonstrate that introduction of a strong, constitutive promoter upstream of the thuggacin gene cluster in *C. crocatus*, results in a 5–7-fold up-regulation of thuggacin titers.

EXPERIMENTAL PROCEDURES

Strains and Culture Conditions

C. crocatus Cm c5 was grown at 30°C and 170 rpm in liquid Pol 03 medium (Kunze et al., 2004). Conjugation was performed as described elsewhere (Rachid et al., 2006), and mutants were cultivated in Pol03 containing 100 μg/mL hygromycin B. *S. cellulosum* So ce895 was grown at 30°C and 170 rpm in liquid HS medium and on solid PM plates as described elsewhere (Irschik et al., 2007). For secondary metabolite analysis, the strain was grown in SG medium. *S. cellulosum* So ce895 mutants were grown in HS or SG production medium, containing 100 μg/mL hygromycin B. *E. coli* DH10B, *E. coli* ET12567/pUB307, *E. coli* ET12567/pUZ8002, and *E. coli* SURE were grown in liquid LB medium at 37°C. When used, antibiotics were present at the following concentrations: kanamycin sulfate (Kan) (50 μg/mL), chloramphenicol (Cm) (20 μg/mL), and ampicillin (Amp) (20 μg/mL).

General Molecular Biology Methods

After gently crushing the cell clumps with a glass homogenizer, *C. crocatus* Cm c5 and *S. cellulosum* So ce895 chromosomal DNA was prepared as described elsewhere (Neumann et al., 1992). Plasmid DNA was isolated using the GeneJET™ Plasmid Miniprep Kit (Fermentas). DNA fragments were purified from agarose gels with the NucleoSpin Extract gel extraction kit (Macherey-Nagel). Polymerase chain reactions (PCR) were performed using *Taq* DNA polymerase (MBI Fermentas) or Phusion Polymerase (Invitrogen). DMSO was added to the reaction mixtures to a final concentration of 5%. Conditions for amplification with an Eppendorf Mastercycler Gradient Thermal Cycler were as follows: denaturation, 15–30 s at 95/98°C; annealing, 20 s at 50–62°C; extension, 15–60 s at 72°C (30 cycles); final extension at 72°C for 10 min. Constructs generated for gene inactivation were verified by sequencing. All other DNA manipulations were performed according to standard protocols (Sambrook et al., 1989). In silico DNA and amino acid sequence analysis were performed with the VectorNTI software package and ClustalW (Thompson et al., 1994), and BLAST was used for comparison with GenBank data.

Construction of the Cosmid Library for *S. cellulosum* So ce895

Chromosomal DNA of *Sorangium cellulosum* So ce895 was partially digested with *Sau*3AI, dephosphorylated, and subsequently ligated into SuperCos 1 (Stratagene), which was pretreated with *Xba*I, dephosphorylated, and additionally digested with *BamH*I. Packaging of the ligation mixture with Gigapack III Gold packaging extract (Stratagene) and transduction of the resulting phages into *E.coli* SURE (Stratagene), lead to a genomic library of 2304 clones. Single colonies were transferred into six 384-well microtiter plates using a Qbot robot (Genetix) and grown overnight in 2 YT medium. For storage at −80°C, 50 μl of freezing solution (0.076% $MgSO_4 \cdot 7\ H_2O$, 0.45% Na-citrate·2 H_2O, 0.9% $NHSO_4$, 44% glycerol, 4.7% K_2HPO_4, and 1.8% KH_2PO_4) was added to each well. High-density colony filters were generated as described elsewhere (Rachid et al., 2006).

Screening of the *C. crocatus* Cm c5 Cosmid Library for the Thuggacin Gene Cluster

A sublibrary of 136 cosmids from a previously generated cosmid library (Rachid et al., 2006) was created on the basis of positive hybridization to PKS and NRPS probes. After exclusion of cosmids that could be correlated to known gene clusters (Buntin et al., 2008; Rachid et al., 2006, 2009), the remaining cosmids were screened for the presence of HC domains, using a PCR approach based on the degenerate HC domain primers KKCYCfor (5′-GCGCCACGAGCCGTTYCCNYTNAC-3′) and KKCYCrev (5′-GCCCAGCA TGGAGGTGAASACNACNGG-3′). Sequencing of a 1 kbp PCR product amplified from cosmid DG17 revealed a new HC domain sequence. Cosmid DG17 was sequenced at both its T3 and T7 ends, revealing PKS genes at its T3 end and genes with no obvious function in thuggacin biosynthesis at its T7 end. Therefore, the complete cosmid library was rescreened with a probe based on the T3 end of DG17. Among the identified cosmids, DG12 showed the smallest extent of overlap with DG17 based on PCR analysis and restriction digest. End sequencing showed that both ends of DG12 contain PKS genes, with only the T3 end overlapping cosmid DG17. Therefore, the cosmid library was rescreened for the remainder of the cluster with a probe based on the T7 end of DG12. This experiment identified overlapping cosmid AB6. Cosmids DG17, DG12, and AB6 were shot-gun sequenced on both strands, as described elsewhere (Silakowski et al., 1999).

Screening of the *S. cellulosum* So ce895 Cosmid Library for the Thuggacin Gene Cluster

Using the degenerate HC domain primers degHCBBfwd (5′-AGCTGATCGA GCGCCAYGAYATGYT-3′) and degHCBJrev (5′-TGGTCCAGCCACACCTGN GGNGTYTG-3′), a 1 kbp fragment was amplified from *S. cellulosum* So ce895 genomic DNA. This fragment was subcloned into pCR2.1TOPO (Invitrogen), and restriction analysis was performed. As all analyzed clones showed identical restriction pattern one clone was sent for sequencing. The obtained HC-domain sequence was used for generating an appropriate probe for screening the cosmid library. The initial hybridization revealed five cosmids (DN15, EG23, FF19, AE22, and FN11); among these cosmids, AE22 and FN11 were identical according to the restriction pattern upon digestion with *BamH*I. T7 and T3 end sequencing showed that cosmid DN15 contained PKS genes at its T7 end and a citrate lyase gene at its T3 end. A portion of the PKS region was amplified from cosmid DN15 and used to rescreen the cosmid library for overlapping sequence. This analysis yielded cosmid BN7 with PKS genes at both ends. PCR analysis revealed that BN7's T3 end showed no overlap with DN15, so the next probe was designed on the basis of this end sequence. The final screening of the cosmid library identified cosmid EJ11. EJ11 showed overlap at its T3 end with BN7 and its T7 end sequence revealed genes with no obvious function in thuggacin biosynthesis. Therefore, the cosmids DN15, BN7 and EJ11 were expected to cover the entire thuggacin biosynthetic gene cluster. Sequencing of the three cosmids was performed by a shotgun approach, as described elsewhere (Silakowski et al., 1999).

Phylogenetic Analysis

The sequences of the KS domains were aligned using ClustalW (Thompson et al., 1994). Phylogenetic trees were created using neighborhood joining (NJ) as a distance method, as well as Bayesian estimation; both programs were accessed via Geneious Pro 4.8.3 (Drummond et al., 2010). For NJ, Jukes-Cantor was chosen as genetic distance model, and bootstrapping was performed with 1000 replications. The Bayesian inference method employed the JTT-amino replacement model and gamma distribution with four categories. Markov chain Monte Carlo analysis (MCMC) was performed with 1.1 million generations, four independent chains, and the Markov chain was sampled every 200 generations. Both methods gave similar tree topologies.

Conjugation of *S. cellulosum* So ce 895

Conjugation of *S. cellulosum* So ce 895 was performed in either liquid or plate format. For conjugation in liquid medium, *S. cellulosum* So ce 895 cells were grown to a cell density of approximately 2×10^8 cells/mL. The cells were

harvested, and the pellet was resuspended in KON medium to a density of 2 × 10^9 cells/mL; 1 × 10^9 cells were then used for each conjugation. A culture of *E. coli* ET12567/pUZ8002 containing the inactivation construct was grown with the antibiotics Hyg (100 μg/mL), Kan (50 μg/mL), and Cm (20 μg /mL) to an OD_{600} of 0.6–0.8. The cell density was determined and either 1 × 10^6 or 1 × 10^7 cells were used per conjugation. So ce895 cells and *E. coli* were mixed, centrifuged, and resuspended in 250 μl KON medium. Cells were incubated in a water bath at 32°C with gentle shaking for 5 hr. The mating was stopped by addition of 4 ml of HS medium containing tobramycin (Tob) (90 μg/mL) followed by incubation at 30°C incubation and 170 rpm. After 15 hr, 20 ml HS medium containing Tob (70 μg/mL) and hygromycin (150 μg/mL) were added, and the cells were incubated for an additional 8 hr at 30°C and 170 rpm, before centrifugation and plating on PM selection plates (Hyg 150 μg/mL, Tob 70 μg/mL). Following 10–14 days of incubation, visible colonies were transferred onto new selection plates.

For plate conjugation, *S. cellulosum* So ce 895 cells and *E. coli* ET12567/pUZ8002 containing the inactivation construct were used in the same amounts as for liquid conjugation. After mixing, centrifugation, and resuspension in 250 μl KON medium, the cells were incubated for 2 hr at 37°C with gentle shaking. The cells were subsequently spotted onto KON plates and incubated overnight at 37°C. The following day, the plates were transferred to 30°C and incubated for an additional 24 hr. On the third day, exconjugants were removed from the plates, resuspended in 1 ml HS, and plated on PM selection plates (Hyg 100 μg/mL, Tob 100 μg/mL). Following 10–14 days of incubation, visible colonies were transferred onto new selection plates.

Gene Inactivation in *S. cellulosum* So ce895

Constructs for gene inactivation were based on pSUPHyg containing a homologous fragment for recombination (see Supplemental Experimental Procedures) and were conjugated into *S. cellulosum* So ce895, as described earlier. HC mutants were analyzed by Southern Blot. For this, the genomic DNA of the mutants was digested with *Nco*I and hybridized with a DIG-labeled internal fragment of the HC domain. One distinct band (7.2 kbp) was detected in the case of the wild-type strain, whereas the mutant yielded two bands (7.8 and 8.7 kbp) resulting from the integrated pSUPHygHC. *tgaD* mutants were analyzed by PCR using appropriate control primers (see Supplemental Experimental Procedures). One primer was designed against the integrated vector backbone, whereas the second primer was designed against the genome sequence either upstream or downstream of the integration site. PCR of the mutant yielded distinct amplicons, whereas no products were detected from the wild-type.

Gene Inactivation and Promotor Insertion in *C. crocatus* Cm c5

Constructs for gene inactivation and promoter insertion were based on pSUPHyg containing a homologous fragment for recombination (see Supplemental Experimental Procedures) and were conjugated using *E. coli* ET12567/pUB307, as described elsewhere (Rachid et al., 2006). HC mutants were analyzed by Southern Blot. For this analysis, chromosomal DNA of wild-type and mutants was digested with *Nco*I and hybridized with a DIG-labeled internal fragment of the HC domain. In the case of the wild-type, one distinct band (3.8 kbp) was detected whereas the mutant yielded two bands of 5 and 8 kbp, resulting from the integrated pSUPHygHC. *tugE*, *orf11*, and *npt* promotor mutants were verified by PCR. For this, chromosomal DNA of the wild-type and mutants was tested with different primer combinations (one directed against the vector backbone and the second within the genomic region outside the inactivated gene), for correct insertion of the gene disruption construct (see Supplemental Experimental Procedures). Using this method, the mutants yielded PCR products of a specific size, whereas no amplicon was obtained from the wild-type DNA.

Analysis of Secondary Metabolite Production

C. crocatus Cm c5 and mutants were grown in Pol03 medium containing 1% XAD adsorber resin (Rohmer and Haas) for 3 days. The resin and cell clumps were harvested and extracted for 2 × 20 min by continuous stirring with methanol. The solvent was removed in vacuo, and the residue was redissolved in methanol, resulting in a 100-fold concentration of the original culture volume. Standard analysis of crude extracts was performed on a HPLC-DAD system (Agilent 1100) coupled to an HCTultra ESI-MS ion trap instrument (Brucker Daltonics) operating in positive ionization mode. Separation was achieved using a Luna RP-C_{18} column (Phenomenex; 125 × 2 mm, 2.5 μm particle size, flow 0.4 mL/min), with a solvent system consisting of water and acetonitrile, both containing 0.1% formic acid. The following gradient was applied: 5%–95% acetonitrile over 20 min. Compounds were detected by diode array and ESI-MS analysis. High-resolution measurements were performed on an Accela UPLC-system (Thermo-Fisher) coupled to an LTQ-Orbitrap (linear trap-FT-Orbitrap combination) operating in positive ionization mode. Compounds were separated on a BEH RP-C_{18} column (Waters; 50 × 2 mm, 1.7 μm particle size, flow 0.6 mL/min), using a solvent system consisting of water and acetonitrile, both containing 0.1% formic acid. The following gradient was applied: 5%–95% acetonitrile over 9 min. The UPLC-system was coupled to the Orbitrap by a Triversa Nanomate (Advion), a chip-based nano-ESI interface. To enable quantification, *C. crocatus* Cm c5 WT and mutant strains were grown in triplicate. Because *Chondromyces* does not grow homogenously in liquid media, each culture flask was inoculated with 2 g of wet cell mass. The analysis was performed using Xcalibur Quan Browser (Thermo).

ACCESSION NUMBERS

The thuggacin gene clusters of *Sorangium cellulosum* So ce895 and *Chondromyces crocatus* Cm c5 have been deposited in GenBank with accession numbers GQ981380 and GQ981381, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, three tables, and Supplemental References and can be found with this article online at doi:10.1016/j.chembiol.2010.02.013.

ACKNOWLEDGMENTS

We thank Shwan Rachid for providing the *Chondromyces crocatus* Cm c5 cosmid library and helpful advice during the course of this project. Elke Dittmann and an anonymous reviewer are thanked for help with the interpretation of the phylogenetic analysis. Research in R.M.'s laboratory was funded by the Bundesministerium für Bildung und Forschung and the Deutsche Forschungsgemeinschaft.

Received: October 13, 2009
Revised: January 22, 2010
Accepted: February 18, 2010
Published: April 22, 2010

REFERENCES

Bock, M., Buntin, K., Müller, R., and Kirschning, A. (2008). Stereochemical determination of thuggacins A–C, highly active antibiotics from the myxobacterium *Sorangium cellulosum*. Angew. Chem. Int. Ed. Engl. *47*, 2308–2311.

Broadhurst, R.W., Nietlispach, D., Wheatcroft, M.P., Leadlay, P.F., and Weissman, K.J. (2003). The structure of docking domains in modular polyketide synthases. Chem. Biol. *10*, 723–731.

Buchholz, T.J., Geders, T.W., Bartley, F.E., III, Reynolds, K.A., Smith, J.L., and Sherman, D.H. (2009). Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. ACS Chem. Biol. *4*, 41–52.

Buntin, K., Rachid, S., Scharfe, M., Blöcker, H., Weissman, K.J., and Müller, R. (2008). Production of the antifungal isochromanone ajudazols A and B in *Chondromyces crocatus* Cm c5: biosynthetic machinery and cytochrome P450 modifications. Angew. Chem. Int. Ed. Engl. *47*, 4595–4599.

Caffrey, P. (2003). Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. ChemBioChem *4*, 654–657.

Del Vecchio, F., Petkovic, H., Kendrew, S.G., Low, L., Wilkinson, B., Lill, R., Cortés, J., Rudd, B.A., Staunton, J., and Leadlay, P.F. (2003). Active-site

residue, domain and module swaps in modular polyketide synthases. J. Ind. Microbiol. Biotechnol. *30*, 489–494.

Donadio, S., and Katz, L. (1992). Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. Gene *111*, 51–60.

Dorrestein, P.C., Van Lanen, S.G., Li, W., Zhao, C., Deng, Z., Shen, B., and Kelleher, N.L. (2006). The bifunctional glyceryl transferase/phosphatase OzmB belonging to the HAD superfamily that diverts 1,3-bisphosphoglycerate into polyketide biosynthesis. J. Am. Chem. Soc. *128*, 10386–10387.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T., and Wilson, A. (2010) Geneious v4.8, Available from http://www.geneious.com.

Erb, T.J., Berg, I.A., Brecht, V., Müller, M., Fuchs, G., and Alber, B.E. (2007). Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. Proc. Natl. Acad. Sci. USA *104*, 10631–10636.

Eustáquio, A.S., McGlinchey, R.P., Liu, Y., Hazzard, C., Beer, L.L., Florova, G., Alhamadsheh, M.M., Lechner, A., Kale, A.J., Kobayashi, Y., et al. (2009). Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from *S*-adenosyl-L-methionine. Proc. Natl. Acad. Sci. USA *106*, 12295–12300.

Frank, B., Knauber, J., Steinmetz, H., Scharfe, M., Blöcker, H., Beyer, S., and Müller, R. (2007). Spiroketal polyketide formation in *Sorangium*: identification and analysis of the biosynthetic gene cluster for the highly cytotoxic spirangienes. Chem. Biol. *14*, 221–233.

Fu, J., Wenzel, S.C., Perlova, O., Wang, J., Gross, F., Tang, Z., Yin, Y., Stewart, A.F., Müller, R., and Zhang, Y. (2008). Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. Nucleic Acids Res. *36*, e113.

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.

Goldman, B.S., Nierman, W.C., Kaiser, D., Slater, S.C., Durkin, A.S., Eisen, J., Ronning, C.M., Barbazuk, W.B., Blanchard, M., Field, C., et al. (2006). Evolution of sensory complexity recorded in a myxobacterial genome. Proc. Natl. Acad. Sci. USA *103*, 15200–15205.

Haydock, S.F., Aparicio, J.F., Molnár, I., Schwecke, T., König, A., Marsden, A.F., Galloway, I.S., Staunton, J., and Leadley, P.F. (1995). Divergent structural motifs correlated with the substrate specificity of (methyl)malonyl-CoA: acyl carrier protein transacylase domains in the modular polyketide synthases. FEBS Lett. *374*, 246–248.

Irschik, H., Reichenbach, H., Höfle, G., and Jansen, R. (2007). The thuggacins, novel antibacterial macrolides from *Sorangium cellulosum* acting against selected Gram-positive bacteria. J. Antibiot. (Tokyo) *60*, 733–738.

Jenke-Kodama, H., Sandmann, A., Müller, R., and Dittmann, E. (2005). Evolutionary implications of bacterial polyketide synthases. Mol. Biol. Evol. *22*, 2027–2039.

Jenke-Kodama, H., Borner, T., and Dittmann, E. (2006). Natural biocombinatorics in the polyketide synthase genes of the actinobacterium *Streptomyces avermitilis*. PLoS Comput. Biol. *2*, 1210–1218.

Kohli, R.M., and Walsh, C.T. (2003). Enzymology of acyl chain macrocyclization in natural product biosynthesis. Chem. Commun. (Camb.), 297–307.

Kopp, F., and Marahiel, M.A. (2007). Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. Nat. Prod. Rep. *24*, 735–749.

Kopp, M., Irschik, H., Pradella, S., and Müller, R. (2005). Production of the tubulin destabilizer disorazol in *Sorangium cellulosum*: biosynthetic machinery and regulatory genes. ChemBioChem *6*, 1277–1286.

Krug, D., Zurek, G., Revermann, O., Vos, M., Velicer, G.J., and Müller, R. (2008). Discovering the hidden secondary metabolome of *Myxococcus xanthus*: a study of intraspecific diversity. Appl. Environ. Microbiol. *74*, 3058–3068.

Kunze, B., Jansen, R., Höfle, G., and Reichenbach, H. (2004). Ajudazols, new inhibitors of the mitochondrial electron transport from *Chondromyces crocatus*: production, antimicrobial activity and mechanism of action. J. Antibiot. (Tokyo) *57*, 151–155.

Liou, G.F., Lau, J., Cane, D.E., and Khosla, C. (2003). Quantitative analysis of loading and extender acyltransferases of modular polyketide synthases. Biochemistry *42*, 200–207.

Liu, Y., Hazzard, C., Eustáquio, A.S., Reynolds, K.A., and Moore, B.S. (2009). Biosynthesis of salinosporamides from α,β-unsaturated fatty acids: implications for extending polyketide synthase diversity. J. Am. Chem. Soc. *131*, 10376–10377.

Long, P.F., Wilkinson, C.J., Bisang, C.P., Cortés, J., Dunster, N., Oliynyk, M., McCormick, E., McArthur, H., Mendez, C., Salas, J.A., et al. (2002). Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. Mol. Microbiol. *43*, 1215–1225.

McKinney, J.D. (2000). *In vivo* veritas: the search for TB drug targets goes live. Nat. Med. *6*, 1330–1333.

Meiser, P., and Müller, R. (2008). Two functionally redundant Sfp-type 4′-phosphopantetheinyl transferases differentially activate biosynthetic pathways in *Myxococcus xanthus*. ChemBioChem *9*, 1549–1553.

Neumann, B., Pospiech, A., and Schairer, H.U. (1992). Rapid isolation of genomic DNA from Gram-negative bacteria. Trends Genet. *8*, 332–333.

Perlova, O., Gerth, K., Hans, A., Kaiser, O., and Müller, R. (2006). Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain *Sorangium cellulosum* So ce56. J. Biotechnol. *121*, 174–191.

Petković, H., Sandmann, A., Challis, L.R., Hecht, H.J., Silakowski, B., Low, L., Beeston, N., Kuscer, E., Garcia-Bernardo, J., Leadlay, P.F., et al. (2008). Substrate specificity of the acyl transferase domains of EpoC from the epothilone polyketide synthase. Org. Biomol. Chem. *6*, 500–506.

Pradella, S., Hans, A., Sproer, C., Reichenbach, H., Gerth, K., and Beyer, S. (2002). Characterisation, genome size and genetic manipulation of the myxobacterium *Sorangium cellulosum* So ce56. Arch. Microbiol. *178*, 484–492.

Rachid, S., Krug, D., Kochems, I., Kunze, B., Scharfe, M., Blöcker, H., Zabriski, M., and Müller, R. (2006). Molecular and biochemical studies of chondramide formation—highly cytotoxic natural products from *Chondromyces crocatus* Cm c5. Chem. Biol. *14*, 667–681.

Rachid, S., Gerth, K., Kochems, I., and Müller, R. (2007). Deciphering regulatory mechanisms for secondary metabolite production in the myxobacterium *Sorangium cellulosum* So ce56. Mol. Microbiol. *63*, 1783–1796.

Rachid, S., Scharfe, M., Blöcker, H., Weissman, K.J., and Müller, R. (2009). Unusual chemistry in the biosynthesis of the antibiotic chondrochlorens. Chem. Biol. *16*, 70–81.

Rangan, V.S., and Smith, S. (1997). Alteration of the substrate specificity of the malonyl-CoA/acetyl-CoA:acyl carrier protein *S*-acyltransferase domain of the multifunctional fatty acid synthase by mutation of a single arginine residue. J. Biol. Chem. *272*, 11975–11978.

Reid, R., Piagentini, M., Rodriguez, E., Ashley, G., Viswanathan, N., Carney, J., Santi, D.V., Hutchinson, C.R., and McDaniel, R. (2003). A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. Biochemistry *42*, 72–79.

Richter, C.D., Nietlispach, D., Broadhurst, R.W., and Weissman, K.J. (2007). Multienzyme docking in hybrid megasynthetases. Nat. Chem. Biol. *4*, 75–81.

Rix, U., Fischer, C., Remsing, L.L., and Rohr, J. (2002). Modification of post-PKS tailoring steps through combinatorial biosynthesis. Nat. Prod. Rep. *19*, 542–580.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Sandmann, A., Sasse, F., and Müller, R. (2004). Identification and analysis of the core biosynthetic machinery of tubulysin, a potent cytotoxin with potential anticancer activity. Chem. Biol. *11*, 1071–1079.

Santoyo, G., and Romero, D. (2005). Gene conversion and concerted evolution in bacterial genomes. FEMS Microbiol. Rev. *29*, 169–183.

Schneiker, S., Perlova, O., Kaiser, O., Gerth, K., Alici, A., Altmeyer, M.O., Bartels, D., Bekel, T., Beyer, S., Bode, E., et al. (2007). Complete genome sequence of the myxobacterium *Sorangium cellulosum*. Nat. Biotechnol. *25*, 1281–1289.

Silakowski, B., Schairer, H.U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blöcker, H., Höfle, G., Beyer, S., and Müller, 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.

Stadler, M., Bitzer, J., Mayer-Bartschmid, A., Müller, H., Benet-Buchholz, J., Gantner, F., Tichy, H.V., Reinemer, P., and Bacon, K.B. (2007). Cinnabaramides A–G: analogues of lactacystin and salinosporamide from a terrestrial streptomycete. J. Nat. Prod. *70*, 246–252.

Staunton, J., and Weissman, K.J. (2001). Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. *18*, 380–416.

Steinmetz, H., Irschik, H., Kunze, B., Reichenbach, H., Höfle, G., and Jansen, R. (2007). Thuggacins, macrolide antibiotics active against *Mycobacterium tuberculosis*: isolation from myxobacteria, structure elucidation, conformation analysis and biosynthesis. Chemistry *13*, 5822–5832.

Tang, G.L., Cheng, Y.Q., and Shen, B. (2004a). Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase. Chem. Biol. *11*, 33–45.

Tang, L., Ward, S., Chung, L., Carney, J.R., Li, Y., Reid, R., and Katz, L. (2004b). Elucidating the mechanism of *cis* double bond formation in epothilone biosynthesis. J. Am. Chem. Soc. *126*, 46–47.

Thattai, M., Burak, Y., and Shraiman, B.I. (2007). The origins of specificity in polyketide synthase protein interactions. PLoS Comput. Biol. *3*, 1827–1835.

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.

van Berkel, W.H., Kamerbeek, N.M., and Fraaije, M.W. (2006). Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. J. Biotechnol. *124*, 670–689.

Walsh, C.T. (2007). The chemical versatility of natural-product assembly lines. Acc. Chem. Res. *41*, 4–10.

Walsh, C.T., Chen, H.W., Keating, T.A., Hubbard, B.K., Losey, H.C., Luo, L.S., Marshall, C.G., Miller, D.A., and Patel, H.M. (2001). Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. Curr. Opin. Chem. Biol. *5*, 525–534.

Weissman, K.J., and Leadlay, P.F. (2005). Combinatorial biosynthesis of reduced polyketides. Nat. Rev. Microbiol. *3*, 925–936.

Weissman, K.J., and Müller, R. (2008a). A brief tour of myxobacterial secondary metabolism. Bioorg. Med. Chem. *17*, 2121–2136.

Weissman, K.J., and Müller, R. (2008b). Protein-protein interactions in multienzyme megasynthetases. ChemBioChem *9*, 826–848.

Wenzel, S.C., and Müller, R. (2007). Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry. Nat. Prod. Rep. *24*, 1211–1224.

Wenzel, S.C., and Müller, R. (2009a). The impact of genomics on the exploitation of the myxobacterial secondary metabolome. Nat. Prod. Rep. *26*, 1385–1407.

Wenzel, S.C., and Müller, R. (2009b). The biosynthetic potential of myxobacteria and their impact on drug discovery. Curr. Opin. Drug Discov. Devel. *12*, 220–230.

Wenzel, S.C., Williamson, R.M., Grünanger, C., Xu, J., Gerth, K., Martinez, R.A., Moss, S.J., Carroll, B.J., Grond, S., Unkefer, C.J., et al. (2006). On the biosynthetic origin of methoxymalonyl-acyl carrier protein, the substrate for incorporation of "glycolate" units into ansamitocin and soraphen A. J. Am. Chem. Soc. *128*, 14325–14336.

Wilkinson, C.J., Frost, E.J., Staunton, J., and Leadlay, P.F. (2001). Chain initiation on the soraphen-producing modular polyketide synthase from *Sorangium cellulosum*. Chem. Biol. *8*, 1197–1208.

Wu, K., Chung, L., Revill, W.P., Katz, L., and Reeves, C.D. (2000). The FK520 gene cluster of *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. Gene *251*, 81–90.

Xie, Z., Siddiqi, N., and Rubin, E.J. (2005). Differential antibiotic susceptibilities of starved *Mycobacterium tuberculosis* isolates. Antimicrob. Agents Chemother. *49*, 4778–4780.

Yadav, G., Gokhale, R.S., and Mohanty, D. (2003). Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. J. Mol. Biol. *328*, 335–363.

Young, D.B., Perkins, M.D., Duncan, K., and Barry, C.E. (2008). Confronting the scientific obstacles to global control of tuberculosis. J. Clin. Invest. *118*, 1255–1265.

Zhang, J. (2003). Evolution by gene duplication: an update. Trends Ecol. Evol. *18*, 292–298.

Zhang, Y., Post-Martens, K., and Denkin, S. (2006). New drug candidates and therapeutic targets for tuberculosis therapy. Drug Discov. Today *11*, 21–27.

