

Genome Mining for Novel Natural Product Discovery

Gregory L. Challis[†]

Department of Chemistry, University of Warwick, Coventry CV4 7AL, U.K.

Received August 1, 2007

Introduction

Genomics has resulted in the deposition of a huge quantity of DNA sequence data from a wide variety of organisms in publicly accessible databases. Such data can be exploited to generate new knowledge in several areas relevant to medicinal chemistry including the characterization of human physiological processes, the identification and validation of new drug targets in human pathogens, and the discovery of new chemical entities (NCEs^a) from natural sources, which may form the basis for new drug leads. The term “genome mining” has been used in various fields to describe the exploitation of genomic information for the discovery of new processes, targets, and products. This Miniperspective will focus on the development of genome mining approaches for the discovery of new natural products. It will also discuss future prospects for the application of genome mining technology to NCE discovery and lead generation.

Natural products and their derivatives form the basis of many important drugs that have found widespread use in the clinic, e.g., as antibacterial (penicillin **1**, vancomycin **2**, erythromycin **3**, daptomycin **4**), antifungal (amphotericin B **5**), immunosuppressant (cyclosporin A **6**, tacrolimus **7**), antitumor (doxorubicin **8**, paclitaxel **9**, bleomycin A₂ **10**, calicheamicin **11**), and cholesterol-lowering agents (mevastatin **12**) (Figure 1).^{1,2} Despite this, most large pharmaceutical companies are no longer seriously engaged in the search for new drug leads from natural sources. The advent of combinatorial chemistry is partly responsible for this decline, together with the high frequency at which known natural products are rediscovered in activity screens of natural extracts.³ However, combinatorial chemistry has failed to deliver leads that form the basis for development of successful new drugs and new possibilities in natural product drug discovery have been opened up by the genomic age.^{4,5} Thus, it is likely that we will soon witness a resurgence of interest in natural products for new drug discovery.

The concept of exploiting genomic sequence data for the discovery of new natural products has grown out of the rapid expansion in knowledge of the genetic and biochemical basis for secondary metabolite biosynthesis, particularly in microorganisms, in the 1980s and 1990s.⁶ As large quantities of genomic sequence data began to accumulate in public databases at the turn of the century, it quickly became apparent that many genomes, in particular those of plants and microorganisms, contain numerous genes encoding proteins likely to participate

in the assembly of structurally complex bioactive natural products but not associated with the production of known metabolites. In the microbial arena, this phenomenon was first recognized during analysis of the complete genome sequences of the model actinomycete *Streptomyces coelicolor* A3(2) and the industrial actinomycete *Streptomyces avermitilis*.^{7–9} Similar observations have since been made for other microbial genomes, e.g., *Pseudomonas fluorescens* Pf-5, *Saccharopolyspora erythraea* NRRL2338, and *Aspergillus* species, as well as some plant genomes.^{10–15}

Progress in understanding the biochemical programming and molecular basis of substrate specificity in two types of natural product biosynthetic systems, known as the modular polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs),⁶ has facilitated the prediction of structural features of natural products assembled by new examples of these systems uncovered by genomics. Modular PKSs and NRPSs are both multienzymes containing numerous enzymatic domains organized into functional units termed modules, usually distributed over one or a few giant polypeptides.⁶ In all of the early-studied modular PKS and NRPS systems, there is a logical correspondence between (i) the number of modules in the biosynthetic system and the number of building blocks incorporated into the natural product and (ii) the occurrence of different kinds of optional domains within modules (Figure 2) and structural features of the natural product. Such “colinearity” is exemplified by the 6-deoxyerythronolide B modular PKS,¹⁶ which uses seven modules to incorporate seven propionate-derived building blocks into the erythromycin macrolactone, and the tyrocidine NRPS,¹⁷ which uses 10 modules to incorporate 10 amino acids into a cyclic decapeptide antibiotic. In both these systems, each domain appears to be used once in the overall assembly process. Every module in these systems contains a domain that specifically recognizes the substrate of the module and covalently tethers it to an adjacent carrier protein domain. In modular PKSs, the acyltransferase (AT) domain selects an acyl-CoA unit and catalyzes its transfer to an acyl carrier protein (ACP) domain, whereas in NRPSs the adenylation (A) domain recognizes a specific amino, aryl, or other acid from among the cellular pool and catalyzes its covalent attachment via a thioester linkage to a peptidyl carrier protein (PCP) domain (Figure 2).⁶ Sequence comparisons of large numbers of AT and A domains of known substrate specificity allowed conserved motifs associated with domains that recognize particular substrates to be identified.^{18–20} Such conserved sequence motifs have proved to be of significant value for predicting the substrates incorporated by “cryptic” modular PKSs and NRPSs, uncovered by genomics, into their structurally uncharacterized products.

Substrate specificity predictions combined with the frequently observed colinear enzymatic logic in modular PKS and NRPS systems provide powerful tools, which in some cases can predict many of the structural features of the products of novel modular PKS and NRPS systems from their primary sequences. However,

[†] Contact information. Phone: +44 (0) 2476 574024. Fax: +44 (0) 2476 524112. E-mail: g.l.challis@warwick.ac.uk.

^a Abbreviations: NCE, new chemical entity; PKS, polyketide synthase; NRPS, nonribosomal peptide synthetase; AT, acyl transferase; ACP, acyl carrier protein; A, adenylation; PCP, peptidyl carrier protein; KS, ketosynthase; DH, dehydratase; KR, ketoreductase; ER, enoylreductase; C, condensation; MT, methyl transferase; E, epimerization; HPLC, high-pressure liquid chromatography; DAD, diode array detector; MS, mass spectrometry; IR, infrared; NMR, nuclear magnetic resonance; PCR, polymerase chain reaction; RT-PCR, reverse transcriptase polymerase chain reaction; DAD, diode array detector; UV, ultraviolet.

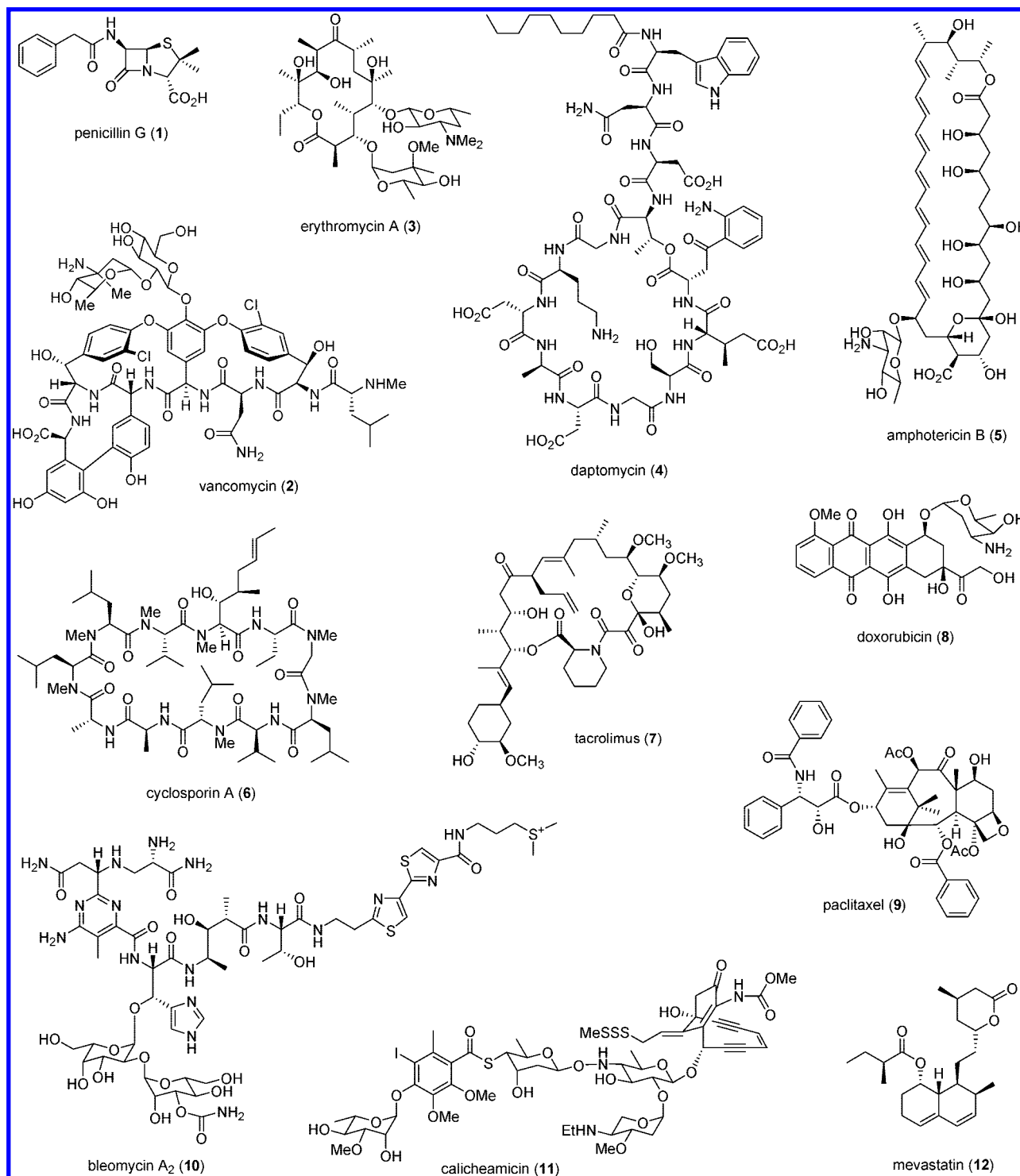


Figure 1. Structures of some important natural products used in the clinic.

these tools have their limits. It is not always possible to accurately predict the substrate specificity of an AT or A domain from its primary sequence, especially if it recognizes a rare or new substrate, and several examples of nonlinear enzymatic logic, such as iterative domain or module use and domain or module skipping, have recently been discovered in modular PKS and NRPS systems.²¹ A further complicating factor is presented by tailoring enzymes that modify the initial PKS and NRPS products by catalyzing reactions such as hydroxylation or O-methylation.⁶ The site of these modifications can often be hard to predict.

For other types of biosynthetic system, the enzymatic logic is less clear and, as a consequence, little or nothing can often be predicted about the structure of putative products “encoded” by cryptic systems. A good example is the terpene cyclases.²² These enzymes catalyze the conversion of linear, methyl-branched polyene substrates, such as geranyl pyrophosphate, farnesyl pyrophosphate, geranylgeranyl pyrophosphate, squalene, and 2, 3-oxidosqualene, to diverse cyclic products (Figure 3). While the substrate of a terpene cyclase of unknown function can often be predicted by comparing its sequence with those of known function, the mechanistic logic employed by different

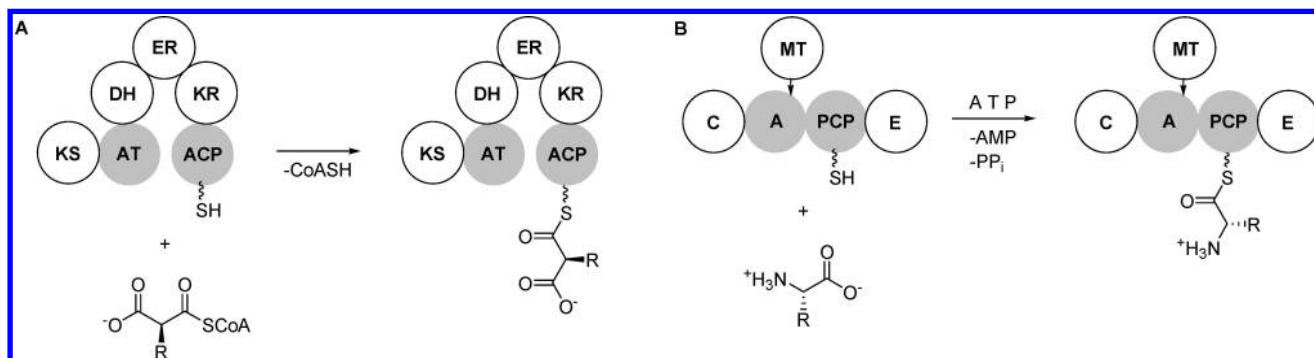


Figure 2. Analogous enzymatic logic used by the AT domain in modular PKSs and the A domain in NRPSs to select substrates from the cellular pool and tether them as thioesters to adjacent carrier protein (ACP/PCP) domains. (A) Domains in a typical PKS module. The AT and ACP domains are present in all modules. The ketosynthase (KS) domain is present in all chain extension modules. The dehydratase (DH), enoyl reductase (ER), and ketoreductase (KR) domains are optional domains. Modules can contain none of these domains, just the KR domain, the KR and DH domains, or the KR, DH, and ER domains. These optional domains modify the structure of the growing polyketide chain during assembly on the PKS in a usually predictable fashion. (B) Domains in a typical NRPS module. The A and PCP domains are present in all modules. The condensation (C) domain is present in all chain extension modules. The epimerization (E) domain is an optional domain that inverts the α -carbon stereochemistry of the substrate. The methyltransferase (MT) domain is also optional. When present, it is inserted into the C-terminal end of the A domain and catalyzes methylation of the α -amino group of the substrate. In some NRPS modules the C domain is replaced with a heterocyclization domain and additional optional domains with usually predictable function are incorporated.

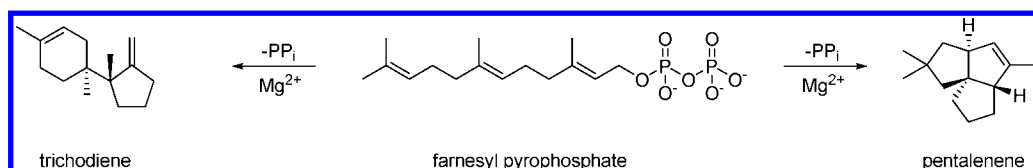


Figure 3. Trichodiene and pentalenene are examples of two structurally diverse natural products assembled by different sesquiterpene synthases from the same substrate farnesyl pyrophosphate.

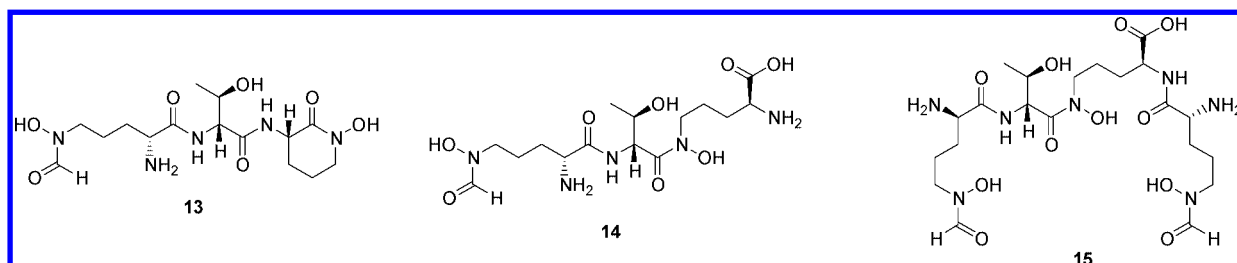


Figure 4. Proposed (13/14) and experimentally determined (15) structures of coelichelin, a novel nonribosomal peptide natural product isolated from *S. coelicolor* by genome mining.

cyclases to direct specific formation of different products from the same substrate is not yet well enough understood to predict the likely structure of the product of a cryptic cyclase.

Despite the current inability to predict structural features of the products of some types of biosynthetic system from sequence data, significant recent progress has been made in the exploitation of genomic information for the discovery of natural products of both modular and other types of cryptic biosynthetic pathway. The major advances in different facets of this exciting new field are reviewed in the following four sections.

Genome Mining for New Peptide Natural Products

The application of sequence-based substrate specificity predictions to the analysis of a cryptic NRPS system was first reported in 2000.²³ A cluster of genes (the *cch* cluster) encoding an NRPS system not associated with the production of a known natural product was discovered in the partially completed genome sequence of *S. coelicolor*. Comparative sequence analyses revealed the domain and module architecture of this NRPS system, and the substrate specificity of the A domain within each module of the NRPS was predicted.²³ It was

hypothesized that the trimodular NRPS system assembled one of two novel D-D-L-tripeptides from the amino acids L- δ -N-formyl- δ -N-hydroxyornithine (fhOrn), L-Thr, and L- δ -N-hydroxyornithine (hOrn), respectively (Figure 4).²³ The predicted hydroxamic acid functional groups in these hypothetical structures suggested that they might bind ferric iron and play a role in transport of this essential inorganic nutrient into the *S. coelicolor* cell. Ferric hydroxamate complexes exhibit characteristic absorbances in their UV-vis spectra that result from ligand to metal charge transfer, suggesting a strategy for selective detection of the natural product in culture supernatants of *S. coelicolor*. The potential role of the natural product in ferric iron acquisition also suggested appropriate growth conditions for its production.

The gene encoding the NRPS was inactivated in *S. coelicolor*, and comparative profiling of ferric iron-binding hydroxamate metabolites accumulated in culture supernatants of wild type *S. coelicolor* and the knockout mutant grown under appropriate (iron-deficient) conditions identified a compound present in the wild type and absent in the mutant.²⁴ This novel natural product, named coelichelin, was purified by semipreparative HPLC, and

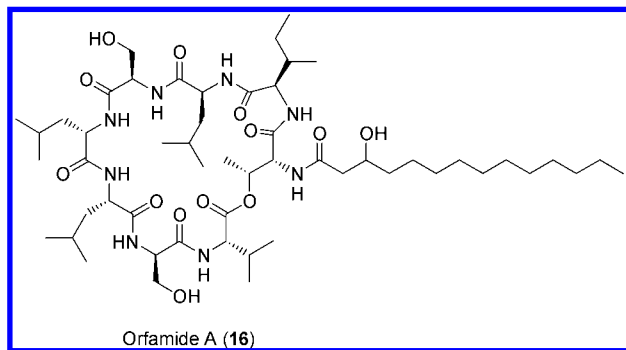


Figure 5. Structure of orfamide A, a novel nonribosomal peptide antibiotic isolated from *P. fluorescens* by genome mining.

its structure was elucidated by mass spectrometry and 1- and 2-D NMR spectroscopy (Figure 4).²⁴ The structure fully validates the prior prediction that the natural product contains D-fhOrn, D-allo-Thr, and L-hOrn, but interestingly is a tetrapeptide incorporating two molecules of D-fhOrn rather than a tripeptide containing one molecule of D-fhOrn as had been predicted on the basis of the colinearity paradigm for NRPS enzymatic logic.^{23,24} Heterologous expression of the *cch* gene cluster in another streptomycete supported the hypothesis that coelichelin NRPS employs nonlinear enzymatic logic in assembly of the tetrapeptide.²⁴ Although no medically relevant biological activity has been reported for coelichelin, it is structurally quite similar to the well-known angiotensin converting enzyme inhibitor foroxymithine²⁵ and has been shown to play a role in ferric iron acquisition by *S. coelicolor*.²⁶

Recently, A domain substrate specificity predictions for a cryptic NRPS system encoded within the *Pseudomonas fluorescens* Pf-5 genome sequence have been used to isolate another novel peptide natural product.²⁷ Analysis of a decamodular NRPS system encoded by the *ofaA*, *ofaB*, and *ofaC* genes suggested that it was likely to catalyze assembly of a leucine-rich lipodecapeptide. This insight allowed selection of a bioassay to detect the natural product and guide its purification from *P. fluorescens*.²⁷ In an interesting and potentially quite general new approach for isolation of new peptide natural products by genome mining, ¹H–¹⁵N HMBC NMR was also used to guide purification of the peptide from cultures of *P. fluorescens* that had been fed L-[¹⁵N]Leu.²⁷ Structure elucidation of orfamide A, the major peptide identified by these approaches, confirmed the accuracy of the A domain substrate specificity predictions and showed it to be a novel cyclic lipodecapeptide (Figure 5).²⁷ Inactivation of the *ofaA* gene in *P. fluorescens* abolished production of orfamide A, unambiguously demonstrating that the NRPS encoded by this gene is required for biosynthesis of the natural product.²⁷ Evaluation of the biological activity of orfamide A showed that it is a biosurfactant that lyses zoospores of *Phytophthora ramorum*, the causative agent of “sudden oak death”, and has moderate antifungal activity against an amphotericin B-resistant strain of *Candida albicans*.²⁷

A third example of the isolation of novel peptide natural products by genome mining has been reported by Müller and co-workers.^{28,29} In 2001, the identification of novel NRPS- and PKS-encoding gene fragments, not associated with the production of known natural products, in the genome of the myxobacterium *Stigmatella aurantiaca* was reported.²⁸ Comparative metabolic profiling of wild type *S. aurantiaca* and mutants with these gene fragments inactivated identified potential products of some of the putative novel gene clusters, although the structures of these compounds were not reported at that time.²⁸

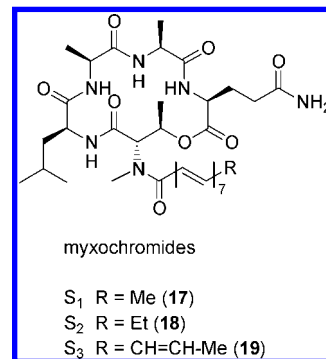


Figure 6. Structures of novel myxochromide nonribosomal peptides isolated from *S. cellulorum* using a genome mining approach.

Recently, the structures of one group of compounds, myxochromides S_{1–3}, identified by this approach were elucidated, and cloning, sequencing, and analysis of the gene cluster directing their biosynthesis were reported.²⁹ Myxochromides S_{1–3} 17–19 (Figure 6) are novel cyclic pentapeptides that are structurally related to myxochromide A,³⁰ a cyclic hexapeptide isolated from the myxobacterium *Myxococcus virescens* Mxv48. Myxochromide S₁ showed no antibacterial or antifungal activity but was found to be weakly cytotoxic.²⁹ The NRPS that assembles the pentapeptide core contains six modules and appears to employ module skipping during the peptide assembly process in another example of nonlinear enzymatic logic in modular synthetases.³¹

Aeruginosides 126A and 126B, novel members of the well-known aeruginosin family of 2-carboxy-6-hydroxyoctahydroindole-containing cyanobacterial serine protease inhibitors, have recently been identified as the products of a cryptic NRPS system discovered in the genome of *Planktothrix agardhii* CYA126/8.³² Predictions of NRPS A domain substrate specificity have also been used to identify previously unknown biosynthetic systems for known natural products encoded within the genomes of sequenced microorganisms.^{33–35}

Recently, novel ribosomally biosynthesized peptides have been isolated from various genera of bacteria by genome mining, for example, trichamide 20, a macrocyclic thiazole-containing peptide produced by the marine cyanobacterium *Trichodesmium erythraeum*, and haloduracin 21, a two-component lantibiotic produced by *Bacillus halodurans* C-125, both of which were predicted to be produced by genome sequence analysis (Figure 7).^{36,37} In these cases, prediction of the amino acids incorporated into the natural products was relatively straightforward because it relied on the genetic code and prediction of the points of excision of the natural products from larger pre-peptides by sequence comparisons with highly similar systems. While no medically relevant biological activity has been reported for trichamide, haloduracin showed activity against *Lactococcus lactis*.³⁷

The prediction that an *S. coelicolor* gene encoding a protein belonging to a third major family of carboxylic acid-activating and amide bond-forming enzymes,^{38–40} unrelated in sequence to either ribosomal or nonribosomal peptide synthetases, is involved in the biosynthesis of known desferrioxamines, including the clinically used iron chelator desferral, has also been confirmed by a gene knockout/comparative metabolic profiling approach.⁴¹

Genome Mining for New Polyketide Natural Products

The first potentially novel polyketide natural products to be identified by a genome mining approach belong to the enediyne group of potent antitumor antibiotics. The enediyne moiety in

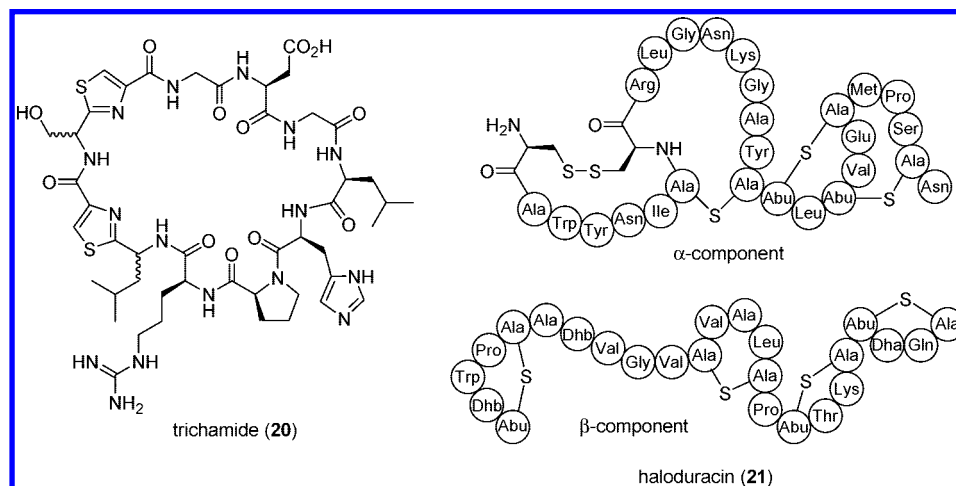


Figure 7. Structures of trichamide and haloduracin, novel ribosomally biosynthesized peptide natural products of *T. erythraeum* and *B. halodurans*, respectively, discovered by genome mining.

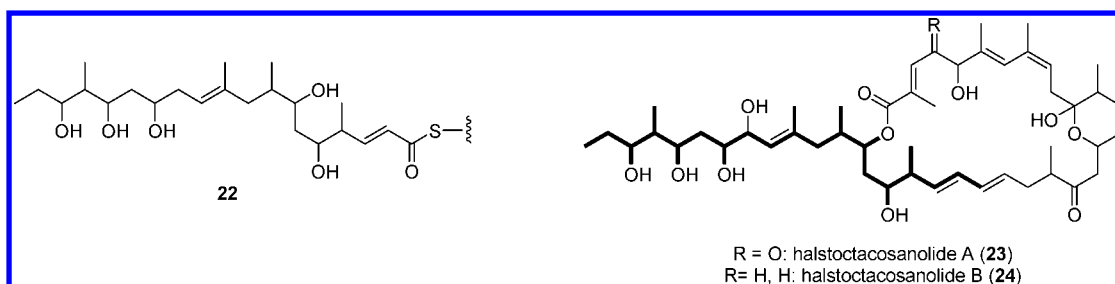


Figure 8. Structure of predicted chain elongation intermediate for a cryptic modular PKS discovered in *S. halstedii* (left) and the experimentally determined structures of the metabolites assembled by this PKS (right). The portion of the experimentally determined structures highlighted in bold corresponds to the predicted intermediate.

such natural products is proposed to be assembled by a characteristic iterative PKS and four universally conserved accessory proteins of unknown function. Relatively little is currently known about the mechanism of this process. Using a genome scanning approach, researchers at Ecopia BioSciences and their collaborators identified several gene clusters encoding the five putative enediynes assembly proteins in several actinomycetes not previously known to produce enediyne natural products.⁴² Bioassays showed that in appropriate growth media the strains harboring these gene clusters produced agents that damage DNA, consistent with the hypothesis that they produce enediynes.⁴² However, none of these compounds appear to have been isolated and structurally characterized, so it is still not clear whether the DNA-damaging agents identified are enediynes or other natural products and whether they are new or known compounds.

The application of AT domain specificity and optional processing domain activity predictions in modular PKS systems to the isolation of novel polyketide natural products was first reported in 2004.⁴³ During a search of the *Streptomyces halstedii* genome for the gene cluster that directs production of the polyketide vicenistatin, Kakinuma and co-workers identified several genes encoding a modular PKS system not involved in vicenistatin biosynthesis.⁴³ Analysis of the domains present in each of the eight modules identified as components of this PKS and prediction of the substrate specificity of the AT domain in each module led to a hypothetical partial structure **22** for the polyketide likely to be assembled by the PKS.⁴³ Comparison of **22** with the structures of other known polyketide natural products suggested that it was novel. Thus, it was decided to examine whether a polyketide containing this partial structure

could be isolated from *S. halstedii*. Two novel polyketides with moderate antibacterial activity, halstoctacosanolide A and B **23** and **24**, were isolated and shown by extensive 1- and 2-D NMR analyses to contain the predicted partial structure bearing an extra hydroxyl group (Figure 8).⁴³ Recently the complete gene cluster believed to direct halstoctacosanolide biosynthesis in *S. halstedii* has been cloned and sequenced and its involvement on the biosynthesis of these novel metabolites has been confirmed by a gene inactivation experiment.⁴⁴

More recently, scientists at Ecopia BioSciences have reported three examples of the isolation of novel polyketides **25–29** from different actinobacteria using a genome mining approach (Figure 9).^{45–47} In each case modular genes encoding PKS systems were identified by genome scanning and the associated gene clusters were cloned and fully sequenced. Analysis of the domain and module architecture of each PKS system and prediction of AT domain substrate specificities, together with functional analyses of ancillary proteins encoded within the gene clusters, led to the prediction of putative novel structures for each of the polyketides. Each of the predicted natural products was subsequently isolated and structurally characterized using spectroscopic techniques.^{45–47} All of the isolated compounds are novel and show significant biological activity: **25** has strong antifungal activity,⁴⁵ **26–28** inhibit electron transport,⁴⁶ and **29** possesses strong antibacterial activity against Gram-positive pathogens including several strains of methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococci* (VRE).⁴⁷ In each case, physicochemical properties deduced from the predicted structures were used to guide the isolation procedures. A furanone similar to **26–28** has also been isolated from the

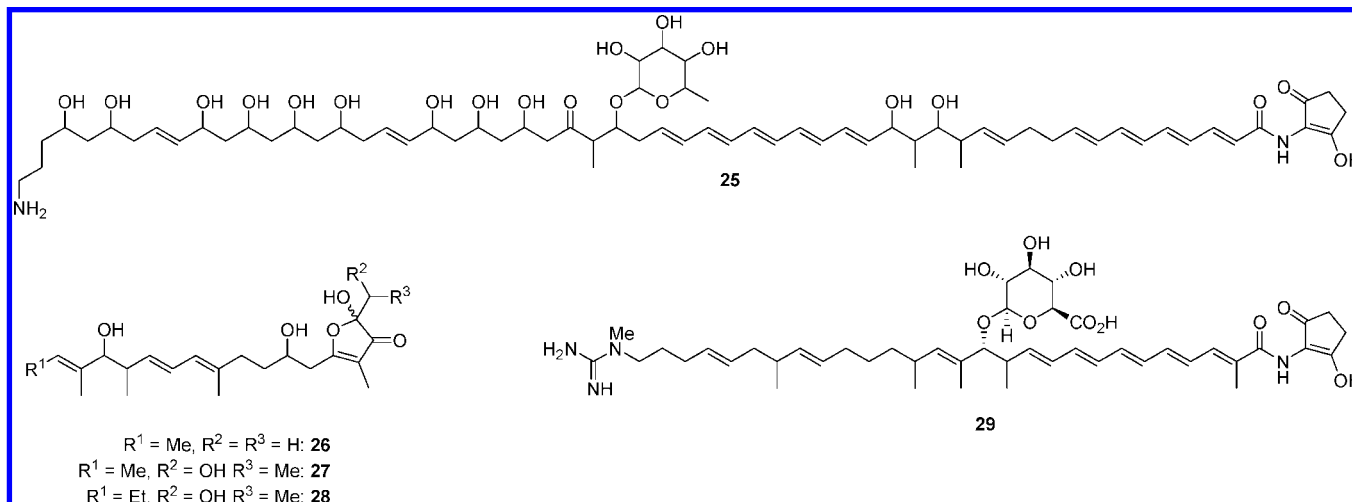


Figure 9. Structures of novel modular PKS products isolated by Ecopia from various actinobacteria by genome mining.

myxobacterium *Stigmatella aurantiaca* by a gene knockout/comparative metabolic profiling genome mining approach.^{28,79}

Two modular PKSs not associated with production of known natural products have also been identified in the genomes of *Saccharopolyspora* species.^{48,49} Sequencing of the genes encoding these PKSs and analyses of the module and domain architecture of the PKSs, together with predictions of AT domain substrate specificity, suggest that these PKSs are likely to assemble novel natural products.^{41,42} In both cases, the PKS genes were inactivated and the profile of metabolites produced by the wild type and mutant strains was compared.^{48,49} No significant differences could be identified in either case, suggesting that the gene clusters are not expressed under the growth conditions examined. In the case of the PKS gene cluster identified in *Sacc. erythraea*, module and domain swapping experiments were used to show that components of the PKS were active, strongly suggesting that the PKS is functional and would produce a metabolite if growth conditions allowing expression of the PKS could be identified.⁴⁹

The modular PKS examples discussed above illustrate the predictive power of the genome mining approach for identifying potential polyketide NCEs and show how deduced physicochemical or biological properties for the predicted structures can be used to identify the corresponding metabolites. However, for some classes of PKS enzyme, such as those that use their active sites iteratively or that employ nonlinear enzymatic logic in their assembly processes, it is not possible to predict the structure(s) likely to be assembled by enzymes of unknown function. Nevertheless, the genome mining approach can still be applied in attempts to identify the products of novel synthases uncovered by genome sequencing projects.

A recent example is the discovery of germicidins **30–34** (Figure 10), which inhibit germination of streptomycete spores,⁵¹ as the unexpected products of a putative type III PKS of unknown function encoded by the *sco7221* coding sequence in the *S. coelicolor* genome.⁵⁰ Bacterial type III PKSs typically employ a single active site to catalyze the iterative condensation of malonyl-CoA building blocks to form poly- β -ketomethylene intermediates that undergo aldol or Dieckmann cyclizations, followed by dehydration and enolization reactions to yield aromatic products such as 1,3,6,8-tetrahydroxynaphthalene, phloroglucinol, or 3,5-dihydroxyphenylacetyl-CoA.^{52–54} Biochemical studies with the purified, recombinant protein encoded by *sco7221* showed that it was active with a variety of acyl-CoA starter units and malonyl-CoA as an extender unit (D. W.

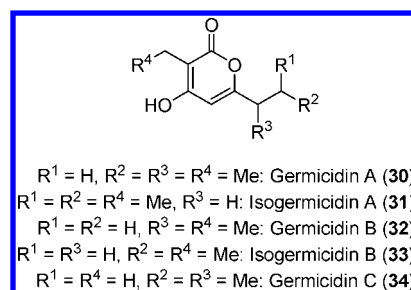


Figure 10. Structures of known (**30** and **32**) and novel (**31**, **33** and **34**) germicidins identified as the products of a cryptic type III PKS encoded within the *S. coelicolor* genome.

Udwary and B. S. Moore, personal communication). They did not, however, provide strong indications for the likely metabolic product of the PKS. Comparative LC–MS profiling of metabolites in organic extracts of wild type *S. coelicolor* and a mutant lacking the *sco7221* coding sequence identified five compounds lacking in the mutant but present in the wild type.⁵⁰ NMR and MS analyses of the purified compounds showed them to be both new and known members of the germicidin family of natural products (Figure 10).⁵⁰ Feeding experiments with labeled precursors and analysis of *S. coelicolor* mutants with altered fatty acid biosynthetic machinery led to the hypothesis that germicidin A **30**, the major metabolite produced by this type III PKS, is assembled by elongation of a specific β -ketoacyl-ACP intermediate in fatty acid biosynthesis with ethylmalonyl-CoA.⁵⁰ The resulting triketide undergoes enolization and cyclization to give the pyrone structure of germicidin A. The use of β -ketoacyl-ACP starter units and the use of ethylmalonyl-CoA as an extender unit are both without precedent for type III PKSs. The utilization of acyl-ACP starter units by germicidin synthase has recently been demonstrated in vitro.⁵⁵ Moreover, type III PKSs normally catalyze several iterations of chain extension, whereas germicidin synthase catalyzes only one chain extension reaction and can therefore be thought of as modular rather than iterative. As a consequence of these unusual features, the involvement of a type III PKS in germicidin biosynthesis could not have been predicted.

A second example is an antibiotic produced by *Bacillus subtilis* that inhibits prokaryotic, but not eukaryotic, protein synthesis and was first discovered in the mid-1990s.⁵⁶ Its structure was not reported at that time because of its chemical instability. Recently, two independent lines of enquiry showed

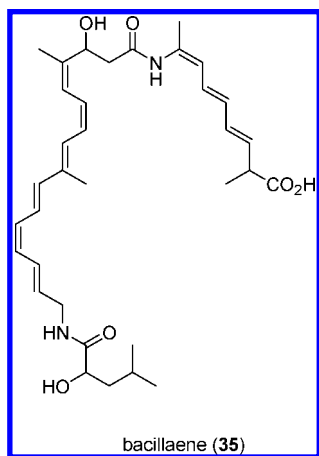


Figure 11. Structure of bacillaene, a novel antibiotic produced by a cryptic hybrid PKS/NRPS system encoded by the *pksX* gene cluster in *Bacillus* spp. A dihydro derivative and several double bond isomers of bacillaene and its dihydro derivative were also isolated.

that this antibiotic was the product of the *pksX* gene cluster, which encodes a cryptic hybrid PKS/NRPS system in *B. subtilis* and *Bacillus amyloliquefaciens* FZB42.^{57,58} Because of several unusual features of the PksX synthases/synthetases, including unconventional module organization, novel domains, and employment of nonlinear enzymatic logic in the assembly process, the prediction of structural elements and thus physicochemical properties of the product could not be used to guide its isolation. This hindrance was overcome by using comparative metabolic profiling of *B. subtilis* strains containing the *pksX* gene cluster with those lacking the *pksX* gene cluster, in conjunction with a minimal chromatography-based purification strategy. Thus, bacillaene, the novel product of this cryptic cluster, was isolated and structurally characterized (Figure 11).⁵⁹ The known polyketide antibiotics diffidin and macrolactin have recently been identified as the products of two further cryptic type I modular PKS systems encoded within the genome of *B. amyloliquefaciens* FZB42 using a gene knockout/comparative metabolic profiling approach.^{57,60}

These examples demonstrate that even when few or no elements of the structure of the products of cryptic biosynthetic systems can be predicted, the genome mining approach can still be applied to the identification of new metabolic products of a microorganism and the discovery of novel and unanticipated biocatalytic properties for well-known enzyme classes. Such discoveries will expand the tool-kit of enzymes that can be employed in ongoing efforts to produce analogues of natural products by genetic engineering.

Genome Mining for New Terpenoid Natural Products

Terpenes are a large group of structurally diverse polycyclic hydrocarbon metabolites produced by many plants and microorganisms. They are biosynthesized from linear polyprenyl pyrophosphate precursors of differing lengths.⁵² Terpene cyclases catalyze the key transformations of these linear precursors into the polycyclic terpenoid carbon skeletons via a series of cation-mediated cyclization and rearrangement reactions.^{22,61} Although the substrate utilized by terpene cyclases of unknown function can usually be predicted using sequence comparisons, it is usually not possible to predict many or any structural features of the products. Thus, genome mining for terpenoid natural products cannot currently be easily targeted toward new structures.

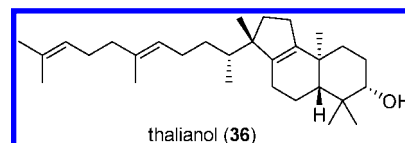


Figure 12. Structure of thalianol, the novel triterpenoid product of a cryptic oxidosqualene cyclase encoded within the *A. thaliana* genome.

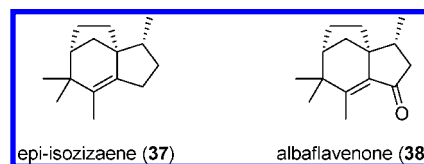


Figure 13. Structures of epi-isozizaene, a novel sesquiterpene synthase product discovered by *Streptomyces coelicolor* genome mining, and albaflavenone, a known *Streptomyces* metabolic product likely to be derived from epi-isozizaene by cytochrome p-450-mediated oxidation.

Analysis of the genome sequence of the model plant *Arabidopsis thaliana* has identified 13 genes encoding likely oxidosqualene cyclases.¹⁵ While some of these are involved in the biosynthesis of known *Arabidopsis* natural products, e.g., cycloartenol, the function of others is cryptic. In 2004, Matsuda and co-workers reported the cloning and expression in yeast of a cDNA encoding one of the cryptic *A. thaliana* oxidosqualene cyclases.¹⁵ Incubation of oxidosqualene with a crude homogenate of the expression host yielded a single triterpenoid product that was subsequently produced by expression of this cDNA in an engineered yeast strain lacking lanosterol synthase (to prevent low product yields resulting from a competition between the heterologous oxidosqualene cyclase and lanosterol synthase for the cellular pool of oxidosqualene).¹⁵ Structure elucidation of this product by MS and NMR showed it to be the novel triterpenoid thalianol **36** (Figure 12).¹⁵ An epoxidized derivative of thalianol was also isolated from the engineered expression host.¹⁵ Since the *A. thaliana* genome appears to encode many more oxidosqualene synthases than are required to account for the number of triterpenoid metabolites it is known to produce, it was hypothesized that many of the genes encoding cryptic oxidosqualene cyclases are only expressed under highly specific conditions.¹⁵ Consequently, heterologous expression of the cDNA encoding such cryptic cyclases represents a powerful approach for defining their function and potentially discovering new plant triterpenoids. Indeed, one of the other cryptic *A. thaliana* oxidosqualene cyclases has recently been shown to catalyze lanosterol biosynthesis,⁶² which was previously unknown in plants, and another of the cryptic *A. thaliana* oxidosqualene cyclases has been shown to catalyze assembly of the iridal skeleton via an unprecedented Grob fragmentation, using the heterologous expression approach.⁶³

An analogous heterologous expression approach has recently been utilized to elucidate the function of a cryptic sesquiterpene synthase encoded within the *S. coelicolor* genome.⁶⁴ Thus, incubation of the purified recombinant protein derived from overexpression of the *S. coelicolor* *sco5222* coding sequence in *E. coli* with farnesyl pyrophosphate led to the formation of epi-isozizaene **37** (Figure 13), which has never been isolated as a natural product.⁶⁴ However, a gene encoding a putative cytochrome P-450 is adjacent to *sco5222* and the two genes are probably co-transcribed,⁷ suggesting that epi-isozizaene could be an intermediate in the biosynthesis of its oxidized derivative albaflavenone **38** (Figure 13), a known natural product of *Streptomyces albidoflavus*.⁶⁵ Interestingly, neither epi-isozizaene nor albaflavenone has been detected in volatile extracts of

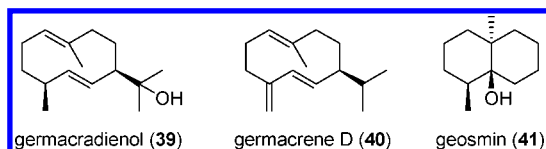


Figure 14. Structures of germacradienol, germacrene D, and geosmin, discovered as products of a cryptic sesquiterpene synthase encoded by the *S.coelicolor* genome.

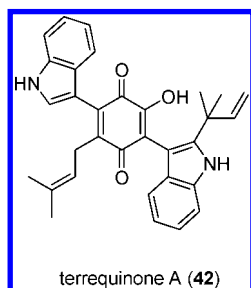


Figure 15. Structure of terrequinone, a previously unknown natural product of *Aspergillus nidulans* discovered by genome mining.

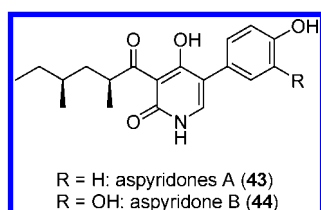


Figure 16. Structures of aspyridones, novel hybrid PKS–NRPS products discovered by activation of a silent cryptic gene cluster identified in the *Aspergillus nidulans* genome sequence.

S. coelicolor, suggesting that the gene encoding the sesquiterpene synthase is not expressed under the growth conditions employed.⁶⁴

A similar heterologous expression approach was used earlier to investigate the function of another cryptic sesquiterpene synthase encoded by the *sco6073* coding sequence of the *S. coelicolor* genome.⁶⁶ These experiments showed that this synthase catalyzes conversion of farnesyl pyrophosphate to the novel compound germacradienol **39** and the known germacrene D **40** (Figure 14).⁶⁶ However, concurrent investigations of the function of this synthase using a gene knockout/comparative metabolic profiling approach showed that it was involved in the biosynthesis of geosmin **41** (Figure 14),⁶⁷ a well-known sesquiterpene-derived hydrocarbon alcohol produced by many *Streptomyces* species. Subsequently, the purified synthase has been shown to catalyze conversion of farnesyl pyrophosphate into geosmin in addition to germacradienol and germacrene D.⁶⁸

Mining Fungal Genomes for New Natural Products

All of the natural products in the three sections above were discovered by mining bacterial or plant genomes. Fungi are well-

known as prolific producers of bioactive natural products, but discoveries of new natural products by mining fungal genomes have only been reported in the past 2 years. Analyses of the genome sequences of *Aspergillus* species have identified numerous cryptic gene clusters encoding enzymes likely to be involved in natural product biosynthesis but not associated with the production of known *Aspergillus* metabolites.^{12,13} Recognizing that the ability to successfully mine *Aspergillus* genomes for new natural products requires the cryptic gene clusters to be expressed, Keller and co-workers applied DNA microarray analyses to examine expression of secondary metabolite gene clusters in *Aspergillus nidulans*.¹³ Several known and putative secondary metabolic gene clusters were identified as being under the control of *LaeA*, a recently identified pleiotropic regulator of secondary metabolite production in *Aspergillus* spp.¹³ By deletion or overexpression of *laeA*, they showed that *LaeA* activates expression of some of the gene clusters and represses expression of the others.¹³ A gene encoding a putative dimethylallyl-L-tryptophan synthase in one of the clusters that is positively regulated by *LaeA* was inactivated, and comparison of metabolites produced by the mutant and wild type identified a yellow compound lacking in the former that was present in the latter.¹³ MS and NMR studies identified this compound as the antitumor agent terrequinone A **42** (Figure 15), which was not previously known as a metabolite of *A. nidulans* but has been isolated from another *Aspergillus* species.¹³ Terrequinone A is derived from a hybrid biosynthetic pathway involving a novel NRPS assembly line and prenyl transferases.^{69,70} The real power of this approach is in the potential it offers for activating normally silent secondary metabolic gene clusters in diverse *Aspergillus* species by manipulation of *laeA* expression levels. However, this potential remains to be demonstrated because the terrequinone biosynthetic gene cluster is clearly expressed in wild type *A. nidulans*.

In a related approach, Hertweck and co-workers discovered a silent gene cluster encoding a cryptic hybrid PKS–NRPS system in the genome of *A. nidulans*. They identified a gene encoding a putative pathway-specific activator protein within this gene cluster. Ectopic integration of a plasmid containing this gene under the control of an inducible promoter into the genome *A. nidulans* resulted in activation of expression of the cryptic gene cluster under inducing conditions. HPLC–DAD–MS analysis revealed two new metabolites in the induced strains that were isolated and determined by NMR, IR, and MS analyses to be the novel compounds aspyridone A **43** and aspyridone B **44** (Figure 16).⁷¹ This work illustrates the considerable potential of regulatory gene manipulation strategies for identification of the products of silent cryptic secondary metabolic gene clusters in microbial genomes.

Another recent report from the same group resulted in the isolation of a second novel family of natural products from *A. nidulans* by a genome mining approach.⁷² It was observed that multiple genes in the genome of *A. nidulans* appear to encode

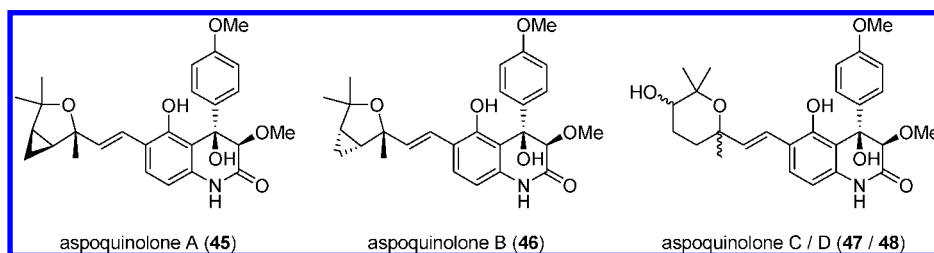


Figure 17. Structures of aspoquinolones, novel *Aspergillus nidulans* natural products discovered by genome mining.

anthranilate synthases.⁷² Anthranilate is known to be a precursor of quinazoline, quinoline, and acridine alkaloids,⁷² but no such compounds have been isolated from *A. nidulans*. The metabolome of *A. nidulans* HKI0410 was screened for nitrogen containing metabolites using HPLC-DAD and HPLC-UV-MS. This resulted in the identification of four new prenylated quinoline-2-one alkaloids, aspoquinolones A-D **45–48** (Figure 17), the structures of which were elucidated by 1- and 2-D NMR spectroscopy.⁷² These compounds exhibited potent antitumor activities and are the first prenylated quinoline-2-ones isolated from *Aspergillus* species. Indeed, very few examples of such compounds are known from filamentous fungi. In *Penicillium* species, prenylated quinoline-2-ones are known to be derived from anthranilate.⁷² Thus, it seems likely that one of the anthranilate-synthase orthologues identified in the genome of *A. nidulans* is involved in the biosynthesis of aspoquinolones A–D.

Given the large number of putative natural product biosynthetic genes identified in *Aspergillus* genomes,¹³ it is likely that many new compounds will be identified in these and other filamentous fungi by ongoing genome mining efforts as more and more genome sequences become available.

Conclusion

From the genome sequence data currently available in public databases, it is already clear that many microbes and plants have the potential to biosynthesize novel natural products. With the ever-increasing pace at which genome sequence data are being acquired, driven by continuing advances in sequencing technology, an abundance of cryptic natural product biosynthetic systems will be discovered. Continued development of genome mining tools will make these cryptic systems a potentially invaluable resource for the discovery of NCEs. Genome-screening approaches, such as PCR-based identification of coding sequences for putative halogenase enzymes,⁷³ which appear to be associated exclusively with natural product biosynthetic pathways in bacteria, continue to be developed and applied to the discovery of novel metabolites. Such approaches offer an attractive alternative to whole genome sequencing for the discovery of cryptic biosynthetic gene clusters.

The predictive sequence analysis tools for modular biosynthetic assembly lines allow rapid assessment, in some cases, of the likely structural novelty of the products of cryptic systems, and progress in the development of these tools continues to be made.⁷⁴ In many cases, they provide sufficient structural insights to predict physicochemical properties that can be exploited in isolation strategies. The discovery of salinilactam A, a novel modular PKS product from mining of the *Salinispora tropica* CNB-440 genome sequence, provides the most recent demonstration of the continuing importance of this approach.⁷⁵ Incorporation of isotopically labeled predicted precursors promises to offer a very useful extension to this approach for guiding metabolite isolation.^{27,76} The recent emergence of nonlinear enzymatic logic in modular systems may, however, significantly impair predictive power of these tools,²¹ depending on the frequency at which these nonlinear phenomena are encountered.

Gene knockouts coupled with comparative metabolic profiling of wild type and mutant strains have proved to be a powerful approach for identifying products of cryptic biosynthetic systems in several microorganisms.^{24,29,50,59} For many plants and microbes, however, gene inactivation is far from straightforward. In these cases, heterologous expression of single biosynthetic genes or whole gene clusters offers an alternative with demonstrated potential.^{15,62,64}

Establishing whether gene clusters encoding cryptic biosynthetic systems are expressed is an important prerequisite for initiating a genome mining campaign. This can be rapidly assessed by RT-PCR, Northern blotting, or DNA microarray analysis. Manipulation of the cellular levels of pleiotropic regulators of secondary metabolism by gene knockout or overexpression has proven potential as a strategy for the activation of silent cryptic pathways,¹³ but the application of this approach for new metabolite discovery has yet to be demonstrated. An alternative approach for the activation of silent clusters is the manipulation of pathway specific regulatory genes, which has demonstrated utility for the identification of novel natural products by genome mining.⁷¹ Heterologous expression of genes in synthetic operons under the control of a strong constitutive promoter offers yet another alternative for activation of silent clusters as demonstrated by the recent identification of gaudimycins (novel angucyclines) as the products of two silent type II polyketide synthase biosynthetic systems identified in *Streptomyces* spp.⁷⁷ Much recent progress has been made in the heterologous expression of entire biosynthetic gene clusters under the control of inducible promoters in the genetically tractable and fast-growing host *Escherichia coli*.⁷⁸ Application of this technology to identification of the products of cryptic biosynthetic gene clusters has yet to be reported but may be an attractive option for clusters that are silent under laboratory conditions in their natural host.

In conclusion, the future of genome mining for new natural product discovery seems bright and it is likely that the coming years will see an explosion of interest in this exciting field. Genome mining approaches are likely to be incorporated into the drug discovery process of biotechnology companies and may also provide the stimuli needed to catalyze the re-entry of big pharmaceutical companies into natural product drug discovery.

Acknowledgment. I thank Dr. C. Corre for helpful comments on the manuscript. Support for genome mining research in the author's lab by BBSRC (Grants B16610 and EGH16081) and the European Union (Integrated Project Actinogen, Contract No. 005224) is gratefully acknowledged.

Biography

Gregory L. Challis received a B.Sc. in Chemistry (First Class Honors) from Imperial College of Science, Technology and Medicine, London, U.K., in 1994, and a D.Phil. in Organic Chemistry from the University of Oxford, U.K., in 1998, where he worked under the supervision of Prof. Sir Jack E. Baldwin FRS. He was a Wellcome Trust International Prize Traveling Research Fellow in the Department of Chemistry, Johns Hopkins University, Baltimore, MD, from 1998 to 2000 and in the Department of Genetics, John Innes Centre, Norwich, U.K., from 2000 to 2001. He was appointed Lecturer in Chemical Biology in the Department of Chemistry, University of Warwick, U.K., in 2001. In 2003 he was promoted to Senior Lecturer in Chemical Biology, and in 2006 he was promoted to his current position of Professor of Chemical Biology at the same institution. His research interests encompass diverse aspects of natural product chemistry and biology.

Note Added in Proof. Cane and co-workers have recently shown that the cytochrome P-450 encoded by the gene adjacent to the *sco5222* coding sequence is indeed responsible for the conversion of *epi*-isozizaene to albaflavenone, as proposed, and that albaflavenone is produced by *S. coelicolor*. See: Zhao, B.; Lin, X.; Lei, L.; Lamb, D. C.; Kelly, S. L.; Waterman, M. R.; Cane, D. E.; Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). *J. Biol. Chem.* **2008**, DOI: 10.1074/jbc.M710421200.

References

- (1) Newman, D. J.; Cragg, G. M.; Snader, K. M. Natural products as sources of new drugs over the period 1981–2002. *J. Nat. Prod.* **2003**, *66*, 1022–1037.
- (2) Butler, M. S. Natural products to drugs: natural product derived compounds in clinical trials. *Nat. Prod. Rep.* **2005**, *22*, 162–195.
- (3) Koehn, F. E.; Carter, G. T. The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discovery* **2005**, *4*, 206–220.
- (4) Wilkinson, B.; Micklefield, J. Mining and engineering natural-product biosynthetic pathways. *Nat. Chem. Biol.* **2007**, *3*, 379–386.
- (5) Gross, H. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Appl. Microbiol. Biotechnol.* **2007**, *75*, 267–277.
- (6) Fischbach, M. A.; Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **2006**, *106*, 3468–3496.
- (7) Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C.-H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O’Neil, S.; Rabinowitch, E.; Rajandream, M.-A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **2002**, *417*, 141–147.
- (8) Omura, S.; Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Takahashi, C.; Shinose, M.; Takahashi, Y.; Horikawa, H.; Nakazawa, H.; Osonoe, T.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12215–12220.
- (9) Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Omura, S. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **2003**, *21*, 526–531.
- (10) Oliynyk, M.; Samborsky, M.; Lester, J. B.; Mironenko, T.; Scott, N.; Dickens, S.; Haydock, S. F.; Leadlay, P. F. Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL2338. *Nat. Biotechnol.* **2007**, *25*, 447–453.
- (11) Paulsen, I. T.; Press, C. M.; Ravel, J.; Kobayashi, D. Y.; Myers, G. S. A.; Mavrodii, D. V.; DeBoy, R. T.; Seshadri, R.; Ren, Q.; Madupu, R.; Dodson, R. J.; Durkin, A. S.; Brinkac, L. M.; Daugherty, S. C.; Sullivan, S. A.; Rosovitz, M. J.; Gwinn, M. L.; Zhou, L.; Schneider, D. J.; Cartinhour, S. W.; Nelson, W. C.; Weidman, J.; Watkins, K.; Tran, K.; Khouri, H.; Pierson, E. A.; Pierson, L. S.; Thomashow, L. S.; Loper, J. E. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat. Biotechnol.* **2005**, *23*, 873–878.
- (12) Keller, N. P.; Turner, G.; Bennett, J. W. Fungal secondary metabolism—from biochemistry to genomics. *Nat. Rev. Microbiol.* **2005**, *3*, 937–947.
- (13) Bok, J. W.; Hoffmeister, D.; Maggio-Hall, L. A.; Renato, M.; Glasner, J. D.; Keller, N. P. Genomic mining for *Aspergillus* natural products. *Chem. Biol.* **2006**, *13*, 31–37.
- (14) Aubourg, S.; Lecharny, A.; Bohlmann, J. Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Mol. Genet. Genomics* **2002**, *267*, 730–745.
- (15) Fazio, G. C.; Xu, R.; Matsuda, S. P. T. Genome mining to identify new plant triterpenoids. *J. Am. Chem. Soc.* **2004**, *126*, 5678–5679.
- (16) Staunton, J.; Wilkinson, B. Biosynthesis of erythromycin and rapamycin. *Chem. Rev.* **1997**, *97*, 2611–2629.
- (17) Mootz, H. D.; Marahiel, M. A. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **1997**, *179*, 6843–6850.
- (18) Haydock, S. F.; Aparicio, J. F.; Molnar, I.; Schwecke, T.; Khaw, L. E.; König, A.; Marsden, A. F.; Galloway, I. S.; Staunton, J.; Leadlay, P. F. Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* **1995**, *374*, 246–248.
- (19) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **1999**, *6*, 493–505.
- (20) Challis, G. L.; Ravel, J.; Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **2000**, *7*, 211–224.
- (21) Haynes, S. W.; Challis, G. L. Non-linear enzymatic logic in natural product modular mega-synthases and synthetases. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 203–218.
- (22) Christianson, D. W. Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **2006**, *106*, 3412–3442.
- (23) Challis, G. L.; Ravel, J. Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. *FEMS Microbiol. Lett.* **2000**, *187*, 111–114.
- (24) Lautru, S.; Deeth, R. J.; Bailey, L. M.; Challis, G. L. Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat. Chem. Biol.* **2005**, *1*, 265–269.
- (25) Umezawa, H.; Aoyagi, T.; Ogawa, K.; Obata, T.; Iinuma, H.; Naganawa, H.; Hamada, M.; Takeuchi, T. Foroxymithine, a new inhibitor of angiotensin-converting enzyme, produced by actinomycetes. *J. Antibiot.* **1985**, *38*, 1813–1815.
- (26) Barona-Gomez, F.; Lautru, S.; Francou, F.-X.; Leblond, P.; Pernodet, J.-L.; Challis, G. L. Multiple biosynthetic and uptake systems mediate siderophore-dependent iron acquisition in *Streptomyces coelicolor* A3(2) and *Streptomyces ambifaciens* ATCC 23877. *Microbiology* **2006**, *152*, 3355–3366.
- (27) Gross, H.; Stockwell, V. O.; Henckels, M. D.; Nowak-Thompson, B.; Loper, J. E.; Gerwick, W. H. The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters. *Chem. Biol.* **2007**, *14*, 53–63.
- (28) Silakowski, B.; Kunze, B.; Müller, R. Multiple hybrid polyketide synthase/nonribosomal peptide synthetase gene clusters in the myxobacterium *Stigmatella aurantiaca*. *Gene* **2001**, *275*, 233–240.
- (29) Wenzel, S. C.; Kunze, B.; Höfle, G.; Silakowski, B.; Scharfe, M.; Blöcker, H.; Müller, R. Antibiotics from gliding bacteria. Part 101. Structure and biosynthesis of myxochromides S1-3 in *Stigmatella aurantiaca*: evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis. *ChemBioChem* **2005**, *6*, 375–385.
- (30) Trowitzsch-Kienast, W.; Gerth, K.; Wray, V.; Reichenbach, H.; Höfle, G. Antibiotics from gliding bacteria. LV. Myxochromide A: a highly unsaturated lipopeptide from *Myxococcus virescens*. *Liebigs Ann. Chem.* **1993**, 1233–1237.
- (31) Wenzel, S. C.; Meiser, P.; Binz, T. M.; Mahmud, T.; Müller, R. Nonribosomal peptide biosynthesis: point mutations and module skipping lead to chemical diversity. *Angew. Chem., Int. Ed.* **2006**, *45*, 2296–2301.
- (32) Ishida, K.; Christiansen, G.; Yoshida, W. Y.; Kurmayer, R.; Welker, M.; Valls, N.; Bonjoch, J.; Hertweck, C.; Boerner, T.; Hemscheidt, T.; Dittmann, E. Biosynthesis and structure of aeruginoside 126A and 126B, cyanobacterial peptide glycosides bearing a 2-carboxy-6-hydroxyoctahydroindole moiety. *Chem. Biol.* **2007**, *14*, 565–576.
- (33) May, J. J.; Wendrich, T. M.; Marahiel, M. A. The *dhb* operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. *J. Biol. Chem.* **2001**, *276*, 7209–7217.
- (34) Silakowski, B.; Kunze, B.; Nordsiek, G.; Blocker, H.; Höfle, G.; Müller, R. The myxochelin iron transport regulon of the myxobacterium *Stigmatella aurantiaca* Sg a15. *Eur. J. Biochem.* **2000**, *267*, 6476–6485.
- (35) de Bruijn, I.; de Kock, M. J. D.; Yang, M.; de Waard, P.; van Beek, T. A.; Raaijmakers, J. M. Genome-based discovery, structure prediction and functional analysis of cyclic lipopeptide antibiotics in *Pseudomonas* species. *Mol. Microbiol.* **2007**, *63*, 417–428.
- (36) Sudek, S.; Haygood, M. G.; Youssef, D. T. A.; Schmidt, E. W. Structure of trichamide, a cyclic peptide from the bloom-forming cyanobacterium *Trichodesmium erythraeum*, predicted from the genome sequence. *App. Environ. Microbiol.* **2006**, *72*, 4382–4387.
- (37) McCleren, A. L.; Cooper, L. E.; Quan, C.; Thomas, P. M.; Kelleher, N. L.; van der Donk, W. A. Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17243–17248.
- (38) Challis, G. L. A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *ChemBioChem* **2005**, *6*, 601–611.
- (39) Oves-Costales, D.; Kadi, N.; Fogg, M. J.; Song, L.; Wilson, K. S.; Challis, G. L. Enzymatic logic of anthrax stealth siderophore biosynthesis: AsbA catalyzes ATP-dependent condensation of citric acid and spermidine. *J. Am. Chem. Soc.* **2007**, *129*, 8416–8417.
- (40) Kadi, N.; Oves-Costales, D.; Barona-Gomez, F.; Challis, G. L. A new family of ATP-dependent oligomerization-macrocyclization biocatalysts. *Nat. Chem. Biol.* **2007**, *3*, 652–656.
- (41) Barona-Gomez, F.; Wong, U.; Giannakopoulos, A. E.; Derrick, P. J.; Challis, G. L. Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. *J. Am. Chem. Soc.* **2004**, *126*, 16282–16283.
- (42) Zazopoulos, E.; Huang, K.; Staffa, A.; Liu, W.; Bachmann, B. O.; Nonaka, K.; Ahlert, J.; Thorson, J. S.; Shen, B.; Farnet, C. M. A genomics-guided approach for discovering and expressing cryptic metabolic pathways. *Nat. Biotechnol.* **2003**, *21*, 187–190.

- (43) Tohyama, S.; Eguchi, T.; Dhakal, R. P.; Akashi, T.; Otsuka, M.; Kakinuma, K. Genome-inspired search for new antibiotics. Isolation and structure determination of new 28-membered polyketide macrolactones, halstocacosanolides A and B, from *Streptomyces halstedii* HC34. *Tetrahedron* **2004**, *60*, 3999–4005.
- (44) Tohyama, S.; Kakinuma, K.; Eguchi, T. The complete biosynthetic gene cluster of the 28-membered polyketide macrolactones, halstocacosanolides, from *Streptomyces halstedii* HC34. *J. Antibiot.* **2006**, *59*, 44–52.
- (45) McAlpine, J. B.; Bachmann, B. O.; Pirae, M.; Tremblay, S.; Alarco, A.-M.; Zazopoulos, E.; Farnet, C. M. Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. *J. Nat. Prod.* **2005**, *68*, 493–496.
- (46) Banskota, A. H.; McAlpine, J. B.; Sorensen, D.; Aoudite, M.; Pirae, M.; Alarco, A.-M.; Omura, S.; Shiomi, K.; Farnet, C. M.; Zazopoulos, E. Isolation and identification of three new 5-alkenyl-3,3(2H)-furanones from two *Streptomyces* species using a genomic screening approach. *J. Antibiot.* **2006**, *59*, 168–176.
- (47) Banskota, A. H.; McAlpine, J. B.; Sorensen, D.; Ibrahim, A.; Aoudite, M.; Pirae, M.; Alarco, A.-M.; Farnet, C. M.; Zazopoulos, E. Genomic analyses lead to novel secondary metabolites. Part 3 ECO-0501, a novel antibacterial of a new class. *J. Antibiot.* **2006**, *59*, 533–542.
- (48) Zirkle, R.; Black, T. A.; Gorlach, J.; Ligon, J. M.; Molnar, I. Analysis of a 108-kb region of the *Saccharopolyspora spinosa* genome covering the obscurin polyketide synthase locus. *DNA Sequence* **2004**, *15*, 123–134.
- (49) Boakes, S.; Oliynyk, M.; Cortes, J.; Böhm, I.; Rudd, B. A. M.; Revill, W. P.; Staunton, J.; Leadlay, P. F. A new modular polyketide synthase in the erythromycin producer *Saccharopolyspora erythraea*. *J. Mol. Microbiol. Biotechnol.* **2005**, *8*, 73–80.
- (50) Song, L.; Barona-Gomez, F.; Corre, C.; Xiang, L.; Udway, D. W.; Austin, M. B.; Noel, J. P.; Moore, B. S.; Challis, G. L. Type III polyketide synthase β -ketoacyl-ACP starter unit and ethylmalonyl-CoA extender unit selectivity discovered by *Streptomyces coelicolor* genome mining. *J. Am. Chem. Soc.* **2006**, *128*, 14754–14755.
- (51) Petersen, F.; Zähler, H.; Metzger, J. W.; Freund, S.; Hummel, R. P. Germicidin, an autoregulative germination inhibitor of *Streptomyces viridochromogenes* NRRL B-1551. *J. Antibiot.* **1993**, *46*, 1126–1138.
- (52) Tseng, C. C.; McLoughlin, S. M.; Kelleher, N. L.; Walsh, C. T. Role of the active site cysteine of DpgA, a bacterial type III polyketide synthase. *Biochemistry* **2004**, *43*, 970–980.
- (53) Austin, M. B.; Izumikawa, M.; Bowman, M. E.; Udway, D. W.; Ferrer, J.-L.; Moore, B. S.; Noel, J. P. Crystal structure of a bacterial type III polyketide synthase and enzymatic control of reactive polyketide intermediates. *J. Biol. Chem.* **2004**, *279*, 45162–45174.
- (54) Zha, W.; Rubin-Pitel, S. B.; Zhao, H. Characterization of the substrate specificity of PhlD, a type III polyketide synthase from *Pseudomonas fluorescens*. *J. Biol. Chem.* **2006**, *281*, 32036–32047.
- (55) Grushow, S.; Buchholz, T. J.; Seufert, W.; Dordick, J. S.; Sherman, D. H. Substrate profile analysis and ACP-mediated acyl transfer in *Streptomyces coelicolor* type III polyketide synthases. *ChemBioChem* **2007**, *8*, 863–868.
- (56) Patel, P. S.; Huang, S.; Fisher, S.; Pirnik, D.; Alkonis, C.; Dean, L.; Meyers, E.; Fernandes, P.; Mayer, F. Bacillaene, a novel inhibitor of prokaryotic protein synthesis produced by *Bacillus subtilis*: production, taxonomy, isolation, physico-chemical characterization and biological activity. *J. Antibiot.* **1995**, *48*, 997–1003.
- (57) Chen, X.-H.; Vater, J.; Piel, J.; Franke, P.; Scholz, R.; Schneider, K.; Koumoutsis, A.; Hitzeroth, G.; Grammel, N.; Strittmatter, A. W.; Gottschalk, G.; Suessmuth, R. D.; Borriss, R. Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J. Bacteriol.* **2006**, *188*, 4024–4036.
- (58) Straight, P. D.; Fischbach, M. A.; Walsh, C. T.; Rudner, D. Z.; Kolter, R. A singular enzymatic megacomplex from *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 305–310.
- (59) Butcher, R. A.; Schroeder, F. C.; Fischbach, M. A.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Clardy, J. The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1506–1509.
- (60) Schneider, K.; Chen, X.-H.; Vater, J.; Franke, P.; Nicholson, G.; Borriss, R.; Suessmuth, R. D. Macrolactin is the polyketide biosynthesis product of the pks2 cluster of *Bacillus amyloliquefaciens* FZB42. *J. Nat. Prod.* **2007**, *70*, 1417–1423.
- (61) Dewick, P. M. The biosynthesis of C5–C25 terpenoid compounds. *Nat. Prod. Rep.* **2002**, *19*, 181–222.
- (62) Kolesnikova, M. D.; Xiong, Q.; Lodeiro, S.; Hua, L.; Matsuda, S. P. T. Lanosterol biosynthesis in plants. *Arch. Biochem. Biophys.* **2006**, *447*, 87–95.
- (63) Xiong, Q.; Wilson, W. K.; Matsuda, S. P. T. An Arabidopsis oxidosqualene cyclase catalyzes iridoid skeleton formation by Grob fragmentation. *Angew. Chem., Int. Ed.* **2006**, *45*, 1285–1288.
- (64) Lin, X.; Hopson, R.; Cane, D. E. Genome mining in *Streptomyces coelicolor*: molecular cloning and characterization of a new sesquiterpene synthase. *J. Am. Chem. Soc.* **2006**, *128*, 6022–6023.
- (65) Gurtler, H.; Pedersen, R.; Anthoni, U.; Christophersen, C.; Nielsen, P. H.; Wellington, E. M.; Pedersen, C.; Bock, K. Alabaflavone, a sesquiterpene ketone with a zizaene skeleton produced by a streptomycete with a new rope morphology. *J. Antibiot.* **1994**, *47*, 434–439.
- (66) Cane, D. E.; Watt, R. M. Expression and mechanistic analysis of a germacradienol synthase from *Streptomyces coelicolor* implicated in geosmin biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1547–1551.
- (67) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1541–1546.
- (68) Jiang, J.; He, X.; Cane, D. E. Geosmin biosynthesis. *Streptomyces coelicolor* germacradienol/germacrene D synthase converts farnesyl diphosphate to geosmin. *J. Am. Chem. Soc.* **2006**, *128*, 8128–8129.
- (69) Balibar, C. J.; Howard-Jones, A. R.; Walsh, C. T. Terrequinone A biosynthesis through L-tryptophan oxidation, dimerization and bisprenylation. *Nat. Chem. Biol.* **2007**, *3*, 584–592.
- (70) Schneider, P.; Weber, M.; Rosenberger, K.; Hoffmeister, D. A one-pot chemoenzymatic synthesis for the universal precursor of antidiabetics and antiviral bis-indolylquinones. *Chem. Biol.* **2007**, *14*, 635–644.
- (71) Bergmann, S.; Schuermann, J.; Scherlach, K.; Lange, C.; Brakhage, A. A.; Hertweck, C. Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. *Nat. Chem. Biol.* **2007**, *3*, 213–217.
- (72) Scherlach, K.; Hertweck, C. Discovery of aspoquinolones A–D, prenylated quinoline-2-one alkaloids from *Aspergillus nidulans*, motivated by genome mining. *Org. Biomol. Chem.* **2006**, *4*, 3517–3520.
- (73) Hornung, A.; Bertazzo, M.; Dziarnowski, A.; Schneider, K.; Welzel, K.; Wohler, S.-E.; Holzenkaempfer, M.; Nicholson, G. J.; Bechthold, A.; Suessmuth, R. D.; Vente, A.; Pelzer, S. A genomic screening approach to the structure-guided identification of drug candidates from natural sources. *ChemBioChem* **2007**, *8*, 757–766.
- (74) Minowa, Y.; Araki, M.; Kanehisa, M. Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. *J. Mol. Biol.* **2007**, *368*, 1500–1517.
- (75) Udway, D. W.; Zeigler, L.; Asolkar, R. N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P. R.; Moore, B. S. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10376–10381.
- (76) Corre, C.; Challis, G. L. Heavy tools for genome mining. *Chem. Biol.* **2007**, *14*, 7–9.
- (77) Palmu, K.; Ishida, K.; Mantsala, P.; Hertweck, C.; Metsä-Ketela, M. Artificial reconstruction of two cryptic angucycline antibiotic biosynthetic pathways. *ChemBioChem* **2007**, *8*, 1577–1584.
- (78) Challis, G. L. Engineering *Escherichia coli* to produce nonribosomal peptide antibiotics. *Nat. Chem. Biol.* **2006**, *2*, 398–400.
- (79) Kunze, B.; Reichenbach, H.; Müller, R.; Höfle, G. Aurafuron A and B, new bioactive polyketides from *Stigmatella aurantiaca* and *Archangium gephyra* (myxobacteria). *J. Antibiot.* **2005**, *58*, 244–251.

JM700948Z