

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): Gregory L. Challis

Article Title: Mining microbial genomes for new natural products and biosynthetic pathways

Year of publication: 2008

Link to published version: <http://dx.doi.org/10.1099/mic.0.2008/018523-0>

Publisher statement: None

SGM Special Lecture

Fleming Prize Lecture
2007

Delivered at the 160th meeting
of the SGM, 27 March 2007

Correspondence
Gregory L. Challis
G.L.Challis@warwick.ac.uk

Mining microbial genomes for new natural products and biosynthetic pathways

Gregory L. Challis

Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK

Analyses of microbial genome sequences have revealed numerous examples of 'cryptic' or 'orphan' biosynthetic gene clusters, with the potential to direct the production of novel, structurally complex natural products. This article summarizes the various methods that have been developed for discovering the products of cryptic biosynthetic gene clusters in microbes and gives an account of my group's discovery of the products of two such gene clusters in the model actinomycete *Streptomyces coelicolor* M145. These discoveries hint at new mechanisms, roles and specificities for natural product biosynthetic enzymes. Our efforts to elucidate these are described. The identification of new secondary metabolites of *S. coelicolor* raises the question: what is their biological function? Progress towards answering this question is also summarized.

Introduction

Alexander Fleming's discovery of penicillin in 1928 and its subsequent development into a medicine by Florey and Chain in the 1940s provided the foundation for development of microbial natural products as a cornerstone of new drug discovery in the 20th century (Fleming, 1929; Chain *et al.*, 1940). By the end of the century many microbial natural products had found their way into the clinic as antibacterial, antifungal, antiparasitic, anticancer and immunosuppressive agents. Yet the turn of the century also witnessed a mass withdrawal of large pharmaceutical companies from new microbial natural product discovery and microbial natural products research (Koehn & Carter, 2005). Several factors spurred this retreat, including rediscovery of known natural products with high frequency, the technical challenges associated with purification and structure elucidation of natural products from microbial fermentations and the advent of combinatorial chemistry, which promised to provide a wealth of new compounds to screen for biological activity. At the very time natural product discovery programmes were ramping down in the late 1990s, large-scale microbial genome sequencing began to ramp up, fuelled by the pioneering application of whole-genome shotgun sequencing to *Haemophilus influenzae*, published in 1995, which demonstrated that microbial genome sequences could be obtained with hitherto unimagined rapidity (Fleischmann *et al.*, 1995). At present there are more than 580 complete microbial genome sequences. Dramatic and sustained increases in our understanding of the genetics and

enzymology of microbial natural product biosynthesis throughout the 1990s have also facilitated the identification and analysis of gene clusters likely to encode natural product biosynthetic pathways in sequenced microbial genomes (Fischbach & Walsh, 2006).

It was extremely fortunate that the Wellcome Trust funded me to carry out research as a postdoctoral fellow in the Department of Genetics at the John Innes Centre in 2000–2001, just as the *Streptomyces coelicolor* A3(2) genome sequencing project was nearing completion. *S. coelicolor* was one of the first sequenced microbes in which it was recognized that there are many more gene clusters encoding natural product-like biosynthetic pathways than there are known natural products of the organism (Bentley *et al.*, 2002). Similar observations have now been reported for several diverse sequenced micro-organisms (Bok *et al.*, 2006; Ikeda *et al.*, 2003; Keller *et al.*, 2005; Oliynyk *et al.*, 2007; Omura *et al.*, 2001; Paulsen *et al.*, 2005; Udvary *et al.*, 2007). This discovery provided a strong counter-argument to the idea that few novel compounds remain to be discovered from natural sources and suggested that the withdrawal of big pharmaceutical companies from natural product drug discovery was premature. At the inception of my independent academic career, we set out to investigate the so-called 'cryptic' or 'orphan' natural product biosynthetic gene clusters found within the genomes of *S. coelicolor* and other sequenced microbes that encode natural product biosynthesis-like proteins not associated with the production of known metabolites. Over the past 6 years numerous groups around the world have focused on similar objectives and today 'genome mining' for new natural products and biosynthetic pathways has become a dynamic and rapidly advancing field (Corre & Challis, 2007; Challis, 2008; Gross, 2007; Wilkinson & Micklefield, 2007).

Abbreviations: A, adenylation; ACP, acyl carrier protein; CDA, calcium-dependent antibiotic; FAS, fatty acid synthase; fhOrn, N^{β} -formyl- N^{β} -hydroxyornithine; hOrn, N^{β} -hydroxyornithine; NRPS, nonribosomal peptide synthetase; PCP, peptidyl carrier protein; PKS, polyketide synthase; TE, thioesterase.

In the first part of this article I will briefly review the different approaches various groups have taken for discovering the metabolic products of cryptic natural product biosynthetic gene clusters. An account of our discovery of new metabolites of *S. coelicolor* follows, along with a discussion of the investigations by us and others of the biosynthesis and biological functions of these metabolites.

Strategies for identifying the metabolic products of cryptic gene clusters

Many microbial natural products, in particular complex polyketides and nonribosomal peptides, are assembled by biosynthetic assembly lines involving modular mega-synthases and synthetases (Fischbach & Walsh, 2006). In many cases the number of modules in the assembly line corresponds exactly to the number of metabolic building blocks incorporated into the final product, although several exceptions to this paradigm have emerged in recent years (Haynes & Challis 2007). The presence or absence of domains with 'tailoring' activities in individual modules often allows prediction of the way in which an initially selected metabolic building block gets modified during the process of its incorporation into the natural product (Fischbach & Walsh, 2006). Models that predict the stereochemical outcome of some of these tailoring reactions, e.g. ketoreduction, have also emerged recently (Caffrey, 2003; Reid *et al.*, 2003). Models that predict the substrate specificity of the adenylation (A) and acyltransferase (AT) domains responsible for building block selection in each module of nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) assembly lines have also been reported and continue to be developed (Haydock *et al.*, 1995; Banskota *et al.*, 2006a, b; Stachelhaus *et al.*, 1999; Challis *et al.*, 2000; Rausch *et al.*, 2005). Insight into the structural features of the metabolic products of cryptic biosynthetic assembly lines can often be derived by application of the above bioinformatics analyses (Banskota *et al.*, 2006a, b; Bentley *et al.*, 2002; Challis & Ravel, 2000; Chen *et al.*, 2007; de Bruijn *et al.*, 2007; McAlpine *et al.*, 2005; Minowa *et al.*, 2007; Nguyen *et al.*, 2008; Paulsen *et al.*, 2005; Sudek *et al.*, 2006; Tohyama *et al.*, 2004; Udway *et al.*, 2007; Zirkle *et al.*, 2004). Such structural insights can lead to the prediction of putative physico-chemical properties of a metabolic product of a cryptic biosynthetic system. The search of fermentation broths for products of cryptic pathways can be narrowed to target only metabolites with the predicted physico-chemical properties, thus simplifying the analytical challenge (Fig.1). Several new metabolic products of cryptic biosynthetic gene clusters have been discovered using such methodologies (Lautru *et al.*, 2005; McAlpine *et al.*, 2005; Banskota *et al.*, 2006a, b).

Two other approaches that have been applied to cryptic biosynthetic systems where the substrates of enzymes in the pathways can be predicted are the 'genom isotopic approach' and *in vitro* reconstitution. In the genom isotopic

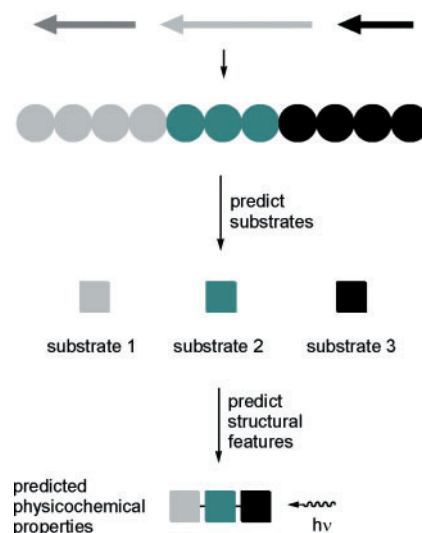


Fig. 1. Method for identifying the product(s) of cryptic biosynthetic gene clusters by predicting likely physico-chemical properties of the product from sequence analyses.

approach, stable-isotope-labelled putative precursors of the metabolic product are fed to the organism containing the cryptic biosynthetic gene cluster and 2D NMR experiments are used to screen extracts of the fermentation broth to identify metabolites containing the labelled precursors (Fig. 2; Gross *et al.*, 2007). NMR detection of the labelled metabolites can be used to guide fractionation of the extracts to facilitate their purification. This approach has been applied to isolation of the orfamides, novel macrocyclic lipopeptides predicted to be produced by *Pseudomonas fluorescens* Pf-5 from analysis of its genome

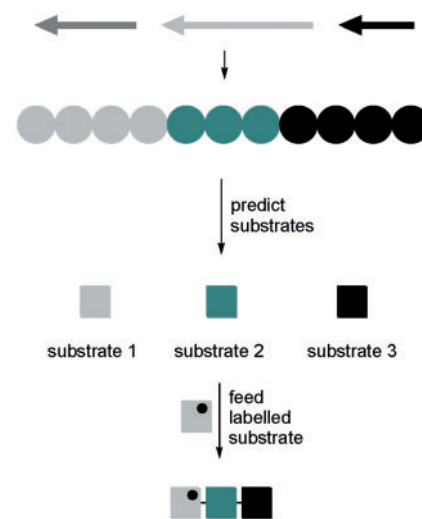


Fig. 2. Principles of the genom isotopic approach for identifying the product(s) of cryptic biosynthetic gene clusters.

sequence (Gross *et al.*, 2007). In the *in vitro* reconstitution approach, the predicted substrates of a biosynthetic enzyme, which has been produced in pure recombinant form, are incubated with it and the structures of the products are determined (Fig. 3). Epi-isozizaene is an example of a new compound that has been identified as the product of a cryptic sesquiterpene synthase discovered by the *S. coelicolor* genome sequencing project using the *in vitro* reconstitution approach (Lin *et al.*, 2006). This metabolite has recently been shown to be an intermediate in the assembly of the known *Streptomyces* sesquiterpene albaflavenone (Zhao *et al.*, 2008).

For some types of biosynthetic system, substrate specificity cannot be predicted with any degree of confidence from bioinformatics analyses. In these cases the directed approaches described above are not useful for finding the metabolic products of cryptic biosynthetic pathways and more generic approaches are required. Two related more generic approaches that have been used successfully to discover the products of cryptic biosynthetic gene clusters are gene knockout/comparative metabolic profiling and heterologous gene expression/comparative metabolic profiling (Corre & Challis, 2007). The first of these involves inactivation of a gene within the cryptic biosynthetic gene cluster hypothesized to be essential for metabolite biosynthesis, followed by comparison of the metabolites in the culture supernatants or extracts of the wild-type organism and the non-producing mutant using an appropriate analytical technique such as liquid chromatography-mass spectrometry (LC-MS). Metabolites present in the wild-type but lacking in the mutant are likely products of the cryptic gene cluster (Fig. 4), which can be isolated and structurally characterized. Germicidins are an example of

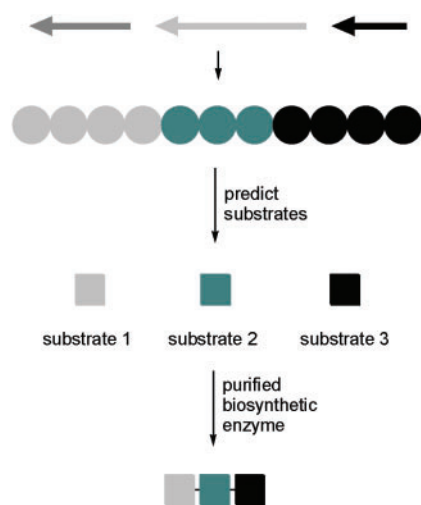


Fig. 3. The *in vitro* reconstitution approach for identifying the product(s) of cryptic biosynthetic gene clusters. The structure(s) of the product(s) resulting from incubation of the predicted substrate(s) with the purified enzyme are determined.

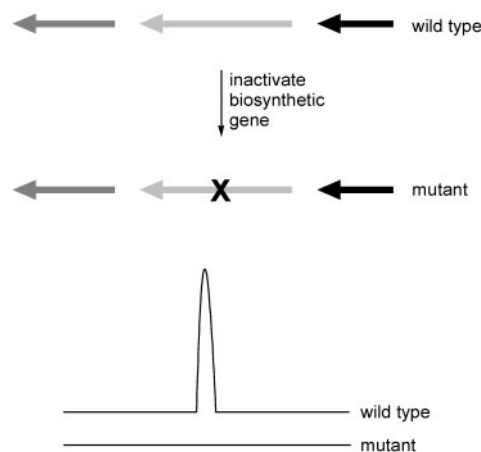


Fig. 4. The gene knockout/comparative metabolic profiling approach to identifying the product(s) of cryptic biosynthetic gene clusters.

metabolites discovered using this strategy (Song *et al.*, 2006). In the second approach the entire biosynthetic gene cluster is cloned, often in a single cosmid or BAC vector, and expressed in a heterologous host. The profile of metabolites in culture supernatants or extracts of the heterologous host containing and lacking the cloned cryptic biosynthetic gene cluster are compared using LC-MS or other appropriate analytical techniques. Metabolites present in the host containing the gene cluster, but absent in the host lacking the cluster, are likely products of the cryptic biosynthetic pathway (Fig. 5), which can be purified and structurally characterized as in the first approach. CBS40 is an example of a novel metabolite that has been identified by this approach (Hornung *et al.*, 2007). One potential obstacle which often has to be overcome in the

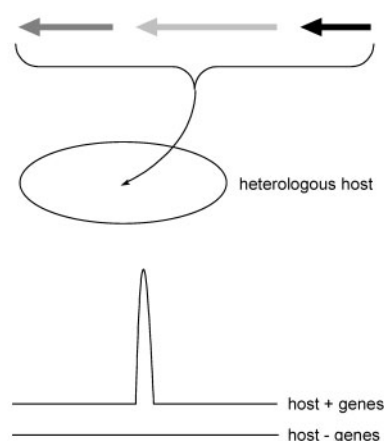


Fig. 5. The heterologous expression/comparative metabolic profiling approach to identifying the product(s) of cryptic biosynthetic gene clusters.

heterologous expression/comparative metabolic profiling approach is that natural product biosynthetic gene clusters are often large (>40 kb) and therefore it can be difficult to clone the entire cluster in a single vector. The use of multiple, mutually compatible expression vectors is one approach to overcoming this problem (Challis, 2006), although it has yet to be applied to the discovery of new metabolic products of cryptic biosynthetic gene clusters.

In the majority of the above approaches a common problem can be encountered: that the cryptic biosynthetic gene cluster is not expressed in the wild-type organism or heterologous host in laboratory culture. The exception is the *in vitro* reconstitution approach, which removes the biosynthetic genes from their natural regulatory context by expressing them under the control of a heterologous (and usually inducible) promoter. However, the *in vitro* reconstitution of an entire biosynthetic pathway usually involves separate overexpression of each gene and purification of the resulting overproduced protein, and many cryptic gene clusters contain multiple putative biosynthetic genes. Thus, the discovery of a fully elaborated metabolic product by this approach is likely to be very laborious. Two approaches to address this problem in *Aspergillus nidulans* have been reported (Bok *et al.*, 2006; Bergmann *et al.*, 2007). The first involved comparative profiling of gene expression in cryptic biosynthetic clusters in the wild-type, and mutants with the pleiotropic regulator of secondary metabolism *laeA* either deleted or overexpressed (Bok *et al.*, 2006). Gene clusters that are differentially expressed in the mutants compared with the wild-type are identified as putatively involved in secondary metabolic biosynthesis. This approach offers potential for the discovery of new metabolic products of cryptic biosynthetic pathways because overexpression of *laeA* causes increased expression of some *Aspergillus* cryptic gene clusters. However, this potential has yet to be demonstrated by the discovery of a novel *Aspergillus* natural product. The second approach involves expression of putative pathway-specific activator genes from within silent cryptic biosynthetic gene clusters under the control of an inducible promoter (Bergmann *et al.*, 2007). This approach has been shown to cause expression of a normally silent gene cluster in *A. nidulans* upon addition of the inducer. Aspyridones, the metabolic products of this gene cluster, were identified by comparative metabolic profiling of the wild-type and mutant strains and spectroscopic analyses showed them to have novel structures, thus proving the utility of this approach for discovering new natural products of cryptic biosynthetic gene clusters that are not expressed in laboratory cultures (Bergmann *et al.*, 2007; Fig. 6).

The various strategies summarized above for identifying the metabolic products of cryptic biosynthetic gene clusters have different strengths and weaknesses, depending on how much can be deduced about the structure of the products from bioinformatics analyses, the size of the gene cluster, and whether the gene cluster is well expressed in laboratory cultures. These factors need to be carefully considered

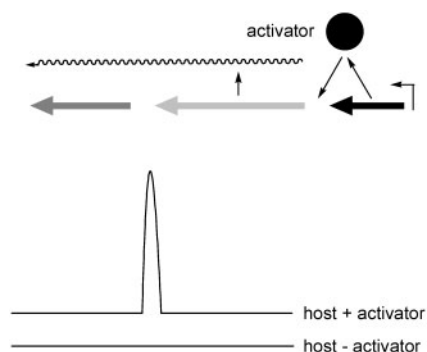


Fig. 6. The expression of pathway-specific activator/comparative metabolic profiling approach to identifying the product(s) of cryptic biosynthetic gene clusters. Placing a pathway-specific activator gene under the control of an inducible promoter results in activation of transcription of the normally silent biosynthetic gene cluster.

when choosing the best approach to take in attempting to identify the products of a cryptic biosynthetic gene cluster. Doubtless further approaches will be added to this already impressive array as research activity in this exciting new field continues to increase.

Discovery of coelichelin by *S. coelicolor* genome mining

In the course of a search of microbial genome sequences for gene clusters encoding novel NRPS systems, a cluster of 11 genes, one of which encodes a protein that is similar to several well-characterized NRPSs, was discovered in the partially completed genome sequence of *S. coelicolor* M145 (Fig. 7; Challis & Ravel, 2000). This gene cluster had not been reported to be involved in the biosynthesis of any known secondary metabolites of *S. coelicolor*. Sequence analysis of the NRPS-like protein encoded by the *cchH* gene in this cluster suggested that it contains 10 catalytic domains, organized into three functional modules (Fig. 7; Challis & Ravel, 2000). The first two modules both contain domains similar to known epimerization (E) domains in other NRPSs, suggesting that these modules incorporate D-amino acids into the product of CchH. Application of the predictive, structure-based models that Jacques Ravel and I developed as postdoctoral researchers in Craig Townsend's group (Challis *et al.*, 2000) suggested that the A domains in modules 1, 2 and 3 of CchH selected the amino acids L-N⁵-formyl-N⁵-hydroxyornithine (L-fhOrn), L-threonine and L-N⁵-hydroxyornithine (L-hOrn), respectively, and catalysed their ATP-dependent transfer onto the adjacent peptidyl carrier protein (PCP) domains in each module (Challis & Ravel, 2000; Fig. 7). Thus we proposed that CchH catalysed assembly of the novel tripeptide D-fhOrn-D-allo-Thr-L-hOrn, which could be formed as one of two essentially isomeric structures, depending on the regiospecificity of the condensation (C) domain in module 3 of the NRPS

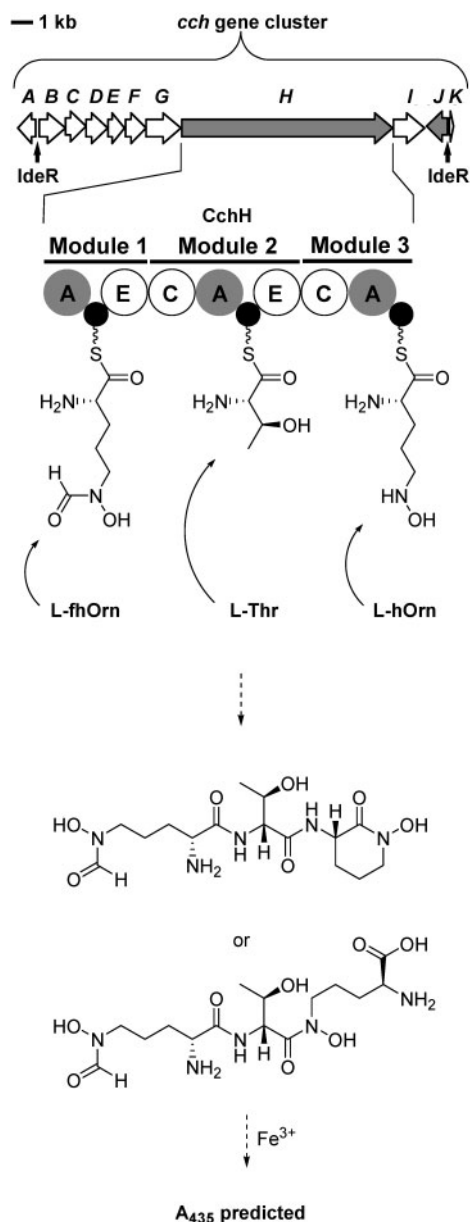


Fig. 7. Organization of the *cch* gene cluster in *S. coelicolor* and the NRPS encoded by the *cchH* gene. Vertical arrows indicate the location of putative ferrous iron-dependent repressor (IdeR) binding sites in two intergenic regions within the cluster. The substrates predicted to be recognized by the A domain in each module of the NRPS are shown attached to the adjacent PCP domains (black circles). Two predicted possible structures of the products of CchH, assuming collinearity between the order and number of modules and the sequence of the peptide product, are shown. These compounds were postulated to form complexes with ferric iron that absorb light at 435 nm, because they contain hydroxamic acid functional groups.

(Fig. 7; Challis & Ravel, 2000). At this time we also noted one unusual feature of CchH, which was that unlike virtually all other NRPS multienzymes, it lacked a

thioesterase (TE) domain for hydrolytic offloading of the fully assembled peptide product from the PCP domain in the last module of the synthetase (Challis & Ravel, 2000).

Sequence analysis of the proteins encoded by other genes in the 11-gene *cch* cluster supported the proposed structure of the CchH product. For example, *cchB* encodes a protein that could convert L-ornithine to its N^5 -hydroxy derivative (L-hOrn) and *cchA* encodes a protein that could catalyse N^5 -formylation of L-hOrn (Challis & Ravel, 2000). Thus, plausible pathways for provision of the proposed non-proteinogenic amino acid substrates of modules 1 and 3 of the CchH NRPS could be envisioned.

Sequence analysis of the intergenic regions between the *cchA* and *cchB* genes, and the *cchH* and *cchI* genes, showed that they contained inverted repeat sequences similar to known binding sites for ferrous iron-dependent repressor (IdeR) proteins in *Streptomyces pilosus* (Fig. 7; Barona-Gómez *et al.*, 2006; Günter *et al.*, 1993). The ferrous complexes of such repressor proteins bind to the inverted repeats and prevent expression of the adjacent genes. When ferrous iron becomes scarce inside the cell the apo-proteins of the repressor are formed, which allows expression of the adjacent genes. Thus this analysis indicated that the *cch* gene cluster would only be expressed when iron is deficient, suggesting appropriate culture conditions for production of the metabolite 'encoded' by the *cch* gene cluster.

The proposed structures of the products of CchH suggested a strategy for their selective detection in culture supernatants of *S. coelicolor*. Both proposed structures contain hydroxamic acid functional groups and it is well known that the deprotonated forms of such functional groups can complex ferric iron. Ferric tris-hydroxamate complexes exhibit characteristic absorbance maxima at 435 nm and it was envisaged that the proposed products of CchH could make such complexes (2-ligand : 1-iron or 3-ligand : 2-iron complexes). Thus, we envisaged that it should be possible to detect the products of CchH by addition of ferric iron to culture supernatants of *S. coelicolor*, followed by HPLC analysis monitoring absorbance at 435 nm (Fig. 7).

In the event that multiple metabolites absorbing at 435 nm were observed after addition of ferric iron to culture supernatants of *S. coelicolor* grown under iron-deficient conditions, we wanted to distinguish between those that were products of the *cch* gene cluster and those that were not. We therefore inactivated the *cchH* gene in *S. coelicolor* by single-crossover integration of a pKC1132 derivative containing an internal fragment of *cchH* (Lautru *et al.*, 2005). Growth of the resulting mutant and the wild-type strain in a medium rendered iron deficient by treatment with Chelex resin, followed by addition of ferric iron to the culture supernatants and comparative metabolic profiling using HPLC, identified a compound absorbing at 435 nm that was absent in the mutant, but present in the wild-type (Lautru *et al.*, 2005). This ferric complex was purified and the metal was removed (Lautru *et al.*, 2005). The iron-free

compound was structurally characterized by high-resolution and tandem ESI-MS (Lautru *et al.*, 2005). Extensive 1- and 2D NMR spectroscopy and molecular modelling of the corresponding gallium complex was also carried out (Lautru *et al.*, 2005). The resulting structure of the product of the cryptic *cch* gene cluster of *S. coelicolor* was deduced to be the tetrapeptide D-fhOrn-D-allo-Thr-L-hOrn-D-fhOrn (Fig. 8; Lautru *et al.*, 2005). This novel natural product was named coelichelin because it is a new *S. coelicolor* iron chelator.

Investigations of coelichelin biosynthesis

The finding that the product of the *cch* cluster is a tetrapeptide containing two residues of fhOrn, one residue of Thr and one residue of hOrn, rather than one residue of each of these amino acids, was quite a surprise. It suggested that coelichelin may be the first example of a tetrapeptide assembled by a trimodular NRPS. Several alternative models for the assembly of coelichelin by the trimodular NRPS CchH can be proposed. All involve the iterative use of module 1 to incorporate two molecules of fhOrn into coelichelin (Fig. 9). Furthermore, the direct linkage between the fhOrn and hOrn residues in the experimentally determined structure of coelichelin implies the direct or indirect interaction of modules 1 and 3 of CchH during the assembly process (i.e. formal 'skipping' of module 2). The *cchH* gene is the only gene within the *cch* cluster that encodes a protein with significant sequence similarity to an NRPS. Nevertheless, it is possible that a gene outside the *cch* cluster, located elsewhere on the chromosome of *S. coelicolor*, could encode a fourth NRPS module that participates in assembly of the coelichelin tetrapeptide structure. To exclude this possibility we integrated a single cosmid containing the entire *cch* gene cluster into the chromosome of the coelichelin non-producer *Streptomyces fungicidicus* (Lautru *et al.*, 2005). The resulting strain produced coelichelin under iron-deficient conditions, providing strong evidence that CchH is the only NRPS required for assembly of the tetrapeptide and that the *cch*

cluster contains all of the genes required for coelichelin biosynthesis in streptomycetes (Lautru *et al.*, 2005). Further evidence for the conclusion that the *cch* gene cluster is sufficient for coelichelin biosynthesis in streptomycetes comes from the finding that an essentially identical gene cluster in *Streptomyces ambofaciens* also directs production of this tetrapeptide (Barona-Gómez *et al.*, 2006).

As mentioned above, an unusual feature of the coelichelin NRPS is that it lacks a C-terminal TE domain for hydrolytic cleavage of the tetrapeptidyl thioester attached to the PCP domain in module 3 of CchH. The *cchJ* gene encodes an enzyme showing significant sequence similarity to Fes, IroD and IroE, enzymes from enteric bacteria that catalyse hydrolytic cleavage of the trilactone ring of ferric-enterobactin and related molecules (Lautru *et al.*, 2005; Lin *et al.*, 2005). A recent crystal structure of IroE shows that it belongs to the α,β -hydrolase superfamily of enzymes and contains an atypical Ser-His catalytic dyad at its active site (Larsen *et al.*, 2006). TE domains also belong to the α,β -hydrolase superfamily and contain Ser-His-Asp catalytic triads. Thus it occurred to us that CchJ could be acting as a thioesterase that catalyses hydrolytic cleavage of the fully assembled coelichelin tetrapeptidyl chain from the PCP domain in module 3 of CchH. To test this hypothesis, we replaced the *cchJ* gene on the chromosome of *S. coelicolor* M145 with an *oriT-aac(3)IV* cassette, which confers apramycin resistance, using PCR-targeting-based methodology (Gust *et al.*, 2003; Lautru *et al.*, 2005). The resulting mutant was unable to produce coelichelin under iron-deficient growth conditions (Lautru *et al.*, 2005). Complementation by integration of a plasmid containing the *cchJ* gene under the control of the constitutive *ermE** promoter into the chromosome of the mutant restored coelichelin production under iron-deficient growth conditions (S. Lautru & G. L. Challis, unpublished data). These results show that CchJ is required for coelichelin biosynthesis and are consistent with the hypothesis that it acts as the terminal thioesterase in this process (Fig. 9).

The *cchA* and *cchB* genes encode enzymes hypothesized to be capable of converting ornithine to the non-proteinogenic amino acids hOrn and fhOrn that are incorporated into the coelichelin structure (Fig. 10). To investigate this hypothesis, we replaced the *cchA* gene in *S. coelicolor* M145 with an *oriT-aac(3)IV* cassette using PCR-targeting methodology. The resulting mutant was unable to produce coelichelin under iron-deficient conditions. If the proposed function of CchA as a formyl-tetrahydrofolate-dependent formyltransferase capable of converting hOrn to fhOrn is correct, feeding of chemically synthesized fhOrn to the *cchA::oriT-aac(3)IV* mutant of *S. coelicolor* should restore coelichelin production. This indeed proved to be the case (D. Oves-Costales, C. Corre & G. L. Challis, unpublished data), providing strong evidence for the hypothesized function of CchA.

The majority of gene clusters containing an NRPS-encoding gene also contain a gene encoding a small

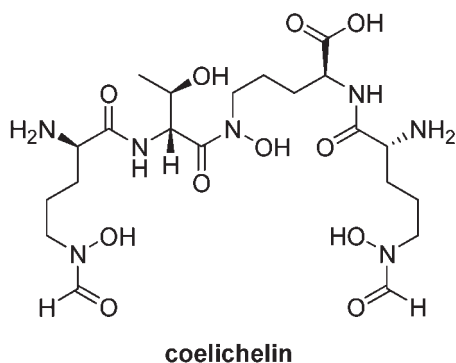


Fig. 8. The experimentally elucidated structure of coelichelin, the product of the *S. coelicolor* *cch* gene cluster.

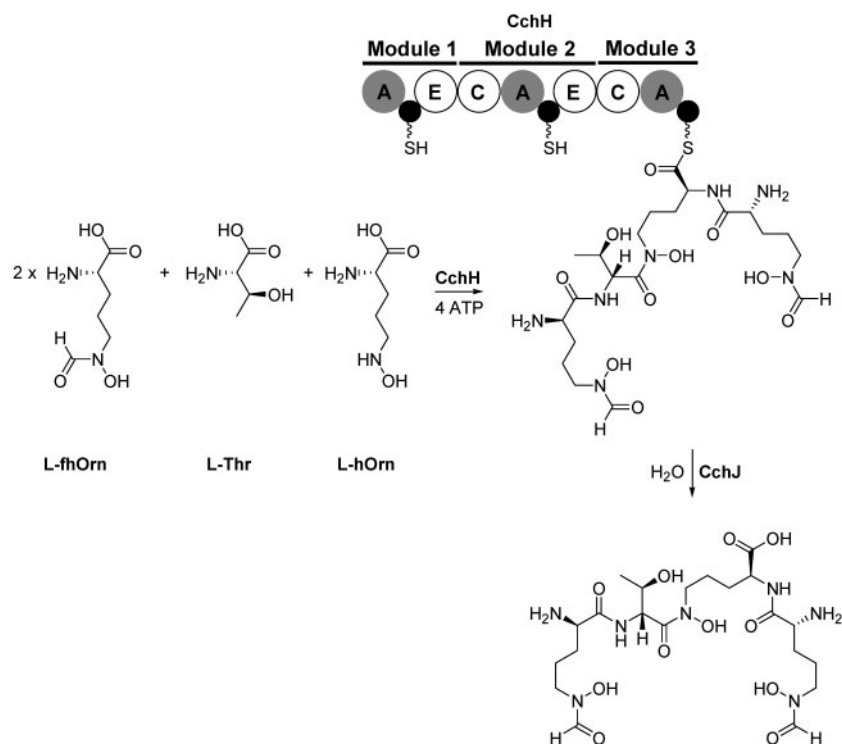


Fig. 9. Proposed assembly of a tetrapeptidyl thioester by CchH, involving iterative use of module 1 and formal skipping of module 2 in the second iteration. CchJ is proposed to catalyse hydrolytic cleavage of the assembled tetrapeptide from the PCP domain in module 3 of CchH.

(<100 amino acid) MbtH-like protein of unknown function. The *cch* gene cluster is no exception. The *cchK* gene encodes an MbtH-like protein. Since no conclusive studies examining the requirement for genes encoding MbtH-like proteins in nonribosomal peptide biosynthesis had been reported, we decided to investigate the requirement of *cchK* for coelichelin biosynthesis. Thus we deleted *cchK* on the chromosome of *S. coelicolor* (Lautru *et al.*, 2007). The resulting mutant still produced coelichelin, but at a lower level than the wild-type strain (Lautru *et al.*, 2007). It occurred to us that an orthologue of CchK encoded by the *cdaX* (SCO3218) gene within the *cda* gene cluster that directs calcium-dependent antibiotic (CDA) biosynthesis in *S. coelicolor* might be partially complementing the *cchK* deletion (Hojati *et al.*, 2002). We therefore replaced the *cdaX* gene in *S. coelicolor* with the *oriT-aac(3)IV* cassette (Lautru *et al.*, 2007). This resulted in a

strain that did not produce CDA on iron-replete medium, but did produce the antibiotic when iron deficiency was enforced by addition of 2,2'-bipyridyl to the medium (Lautru *et al.*, 2007). These results indicated that CchK could complement the *cdaX* deletion, suggesting that the CchK and CdaX MbtH-like proteins mediate cross-talk between the coelichelin and CDA biosynthetic pathways. To further investigate this hypothesis we constructed a *cchK/cdaX* double mutant of *S. coelicolor* (Lautru *et al.*, 2007). The production of both coelichelin and CDA was abolished in this mutant, showing that MbtH-like proteins are required for the biosynthesis of both these nonribosomal peptides in *S. coelicolor* (Lautru *et al.*, 2007). Expression of either *cdaX* or *cchK* in the double mutant restored production of both CDA and coelichelin under appropriate growth conditions, showing that it is indeed the MbtH-like proteins encoded by these genes that mediate cross-talk between the two biosynthetic pathways (Lautru *et al.*, 2007).

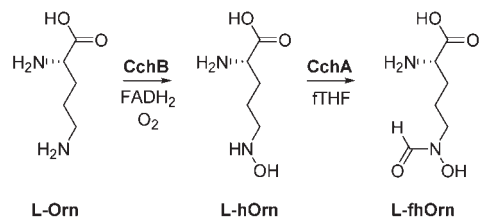


Fig. 10. Proposed roles of CchA and CchB in the biosynthesis of non-proteinogenic amino acid precursors of coelichelin from ornithine. fTHF, formyltetrahydrofolate.

While these studies firmly established the requirement for MbtH-like proteins in nonribosomal peptide biosynthesis in *S. coelicolor*, they did not shed light on the possible role of these proteins. One possibility is that CchK and CdaX are transcriptional regulators of the coelichelin and CDA biosynthetic genes. To investigate this hypothesis, RT-PCR analyses of transcript levels for biosynthetic genes in the *cda* and *cch* gene clusters were carried out in wild-type *S. coelicolor* and the *cchK/cdaX* mutant (Lautru *et al.*, 2007). No significant differences were observed, suggesting that MbtH-like proteins may not regulate transcription of

nonribosomal peptide biosynthetic genes (Lautru *et al.*, 2007). The X-ray crystal structure of an MbtH-like protein encoded by a gene within the pyoverdine biosynthetic gene cluster of *Pseudomonas aeruginosa* has recently been reported (Drake *et al.*, 2007; Fig. 11). However, this structure does not suggest possible functions for these proteins in nonribosomal peptide biosynthesis. One possibility is that MbtH-like proteins could interact directly with NRPSs in some way to increase the efficiency of peptide biosynthesis. Further experiments will be required to test this hypothesis and define a role for these essential proteins in nonribosomal peptide assembly.

The biological function of coelichelin

NMR studies of the Ga-coelichelin complex suggest that this novel *S. coelicolor* natural product may be capable of making a strong tris-hydroxamate complex with ferric iron. Indeed the formation of a 1:1 ferric-coelichelin tris-hydroxamate complex (ferricoelichelin) is supported by ESI-MS analyses and UV-visible spectroscopy, which shows a broad absorbance with a maximum at 435 nm (accounting for the red colour of the complex) that is characteristic of ferric tris-hydroxamate complexes (Barona-Gómez *et al.*, 2006; Fig. 12). Several natural products containing hydroxamic acid functional groups are known to play a role in microbial

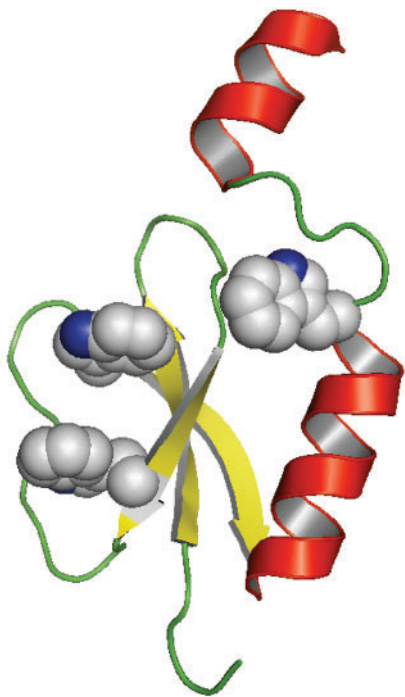


Fig. 11. Structure of the MbtH-like protein encoded by a gene within the pyoverdine biosynthetic gene cluster of *P. aeruginosa*. The side chains of three Trp residues that are universally conserved in MbtH-like proteins are highlighted in space-filling representation.

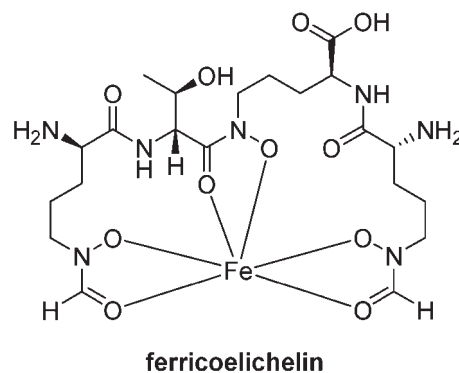


Fig. 12. Structure of the ferricoelichelin complex.

iron uptake. Although iron is the fourth most abundant element in the Earth's crust, in its ferric form it exists predominantly as highly insoluble oxide/hydroxide polymeric complexes from which saprophytic micro-organisms cannot directly remove the iron. Similarly in mammals free iron levels are very low because storage and transport proteins such as transferrin and lactoferrin bind ferric iron with high affinity, thus preventing pathogenic microbes from directly acquiring ferric iron from their hosts. Iron is an essential element for growth and proliferation of nearly all micro-organisms. As a consequence, many biosynthesize and excrete high-affinity ferric iron chelators called siderophores, which scavenge iron from deposits in the environment or their hosts (Miethke & Marahiel, 2007). Active transport systems translocate the iron-siderophore complexes into microbial cells and once they are internalized, the iron is removed by reductive and/or hydrolytic processes (Miethke & Marahiel, 2007).

Sequence analyses of the proteins encoded by the *cchCDEF* genes indicated that they are similar to lipoprotein receptor, permease and ATPase components of known ferric-siderophore transport systems in other bacteria (Miethke & Marahiel, 2007; Fig. 13). Together with the experimentally proven affinity of coelichelin for ferric iron and the sequence analyses of intergenic regions in the *cch* cluster suggesting that its expression is controlled by intracellular iron levels (Barona-Gómez *et al.*, 2006), this observation suggested that coelichelin plays a role in iron uptake by *S. coelicolor*. In other work we had identified the *des* gene cluster that directs the production of known tris-hydroxamic acid-containing desferrioxamines of *S. coelicolor* (Barona-Gómez *et al.*, 2004). Ferrioxamine complexes are known to be utilized as iron sources by *Streptomyces pilosus* (Muller & Raymond, 1984), thus it seemed likely that the desferrioxamine products of the *des* cluster as well as coelichelin could play a role in *S. coelicolor* iron acquisition.

To test these hypotheses we created independent mutants of *S. coelicolor* lacking the NRPS-encoding *cchH* gene and the *desD* gene (Barona-Gómez *et al.*, 2004, 2006), which encodes an enzyme that catalyses the key step in

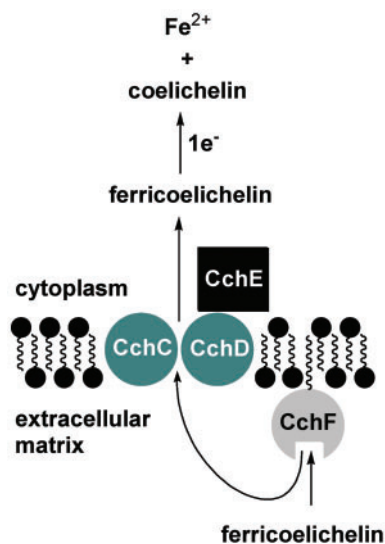


Fig. 13. Proposed function of CchF (lipoprotein receptor), CchC/CchD (membrane-spanning permeases) and CchE (ATPase) as components of an ABC transporter system that mediates the selective uptake of ferricoelichelin in *S. coelicolor*. Reductive release of iron from ferricoelichelin would yield ferrous iron for utilization in biochemical processes that are vital for the survival of the cell.

desferrioxamine biosynthesis (Kadi *et al.*, 2007). The *cchH* mutant was incapable of producing coelichelin, but still produced desferrioxamines (Barona-Gómez *et al.*, 2006). Conversely the *desD* mutant did not produce desferrioxamines (Barona-Gómez *et al.*, 2004), but still made coelichelin (Barona-Gómez *et al.*, 2006). We also constructed a mutant lacking both the *cchH* and *desD* genes, which did not produce coelichelin or desferrioxamines (Barona-Gómez *et al.*, 2006). The ability of these mutants to grow on a colloidal silica medium was examined. While both the single mutants could grow on this medium

provided ferric iron was added, the double mutant could not, even in the presence of high concentrations of ferric iron (Barona-Gómez *et al.*, 2006). Addition of purified ferrioxamines or ferricoelichelin to the double mutant restored growth (Barona-Gómez *et al.*, 2006). These data strongly suggest that both desferrioxamines and coelichelin function as siderophores that mediate iron uptake in *S. coelicolor*.

Discovery of new and known germicidins by *S. coelicolor* genome mining

Type III PKSs catalyse the iterative decarboxylation and concomitant condensation of malonyl-CoA building blocks to form a poly- β -ketomethylene intermediate of defined chain length that typically undergoes PKS-catalysed cyclization and dehydration reactions to yield an aromatic product (Austin & Noel, 2003). Important bacterial examples include DpgA from *Amycolatopsis orientalis*, which with the assistance of DpgB and DpgC catalyses the assembly of 3,5-dihydroxyphenylacetyl-CoA (a precursor of the non-proteinogenic amino acid 3,5-dihydroxyphenylglycine that is incorporated into the antibiotic vancomycin) from four molecules of malonyl-CoA (Chen *et al.*, 2001; Li *et al.*, 2001; Pfeifer *et al.*, 2001; Fig. 14), and RppA from *Streptomyces griseus*, which catalyses the assembly of 1,3,6,8-tetrahydroxynaphthalene from five molecules of malonyl-CoA (Funa *et al.*, 1999; Fig. 15). The molecular mechanisms by which type III PKSs control chain length and cyclization regiochemistry are still not well understood. Thus, it is difficult to predict the structures of the likely products of putative novel type III PKSs uncovered by genome sequencing.

Analysis of the *S. coelicolor* complete genome sequence identified three genes encoding proteins with significant sequence similarity to known type III PKSs (Bentley *et al.*, 2002). One of these proteins showed a high degree of sequence similarity to RppA and, not surprisingly, was subsequently shown to catalyse assembly of 1,3,6,8-

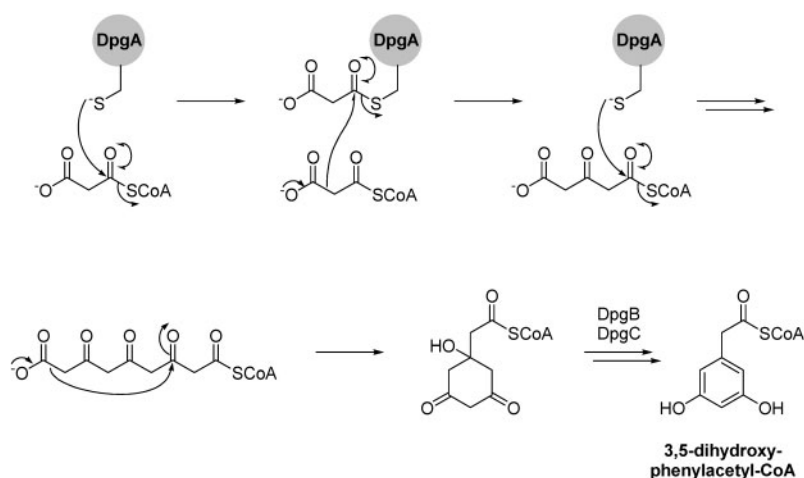


Fig. 14. Reactions catalysed by the type III PKS DpgA in the biosynthesis of 3,5-dihydroxyphenylacetyl-CoA. DpgB and DpgC assist in this process.

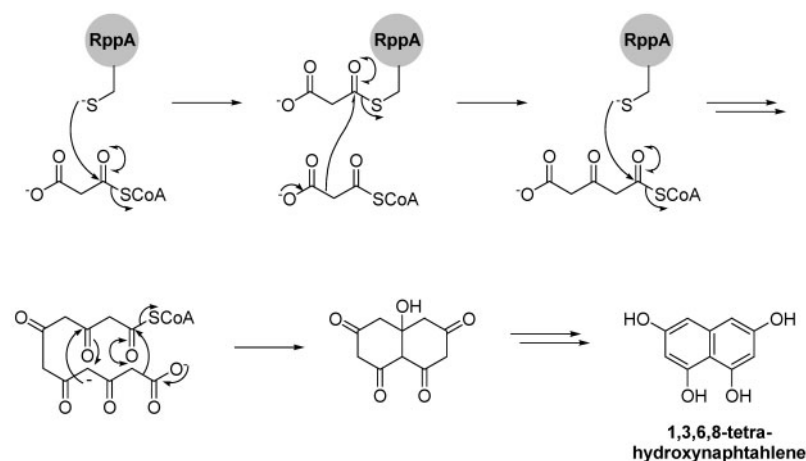


Fig. 15. Reactions catalysed by the type III PKS RppA in the assembly of 1,3,6,8-tetrahydroxynaphthalene in *S. griseus* and *S. coelicolor*.

tetrahydroxynaphthalene (Izumikawa *et al.*, 2003). To attempt to identify the product(s) of one of the two remaining cryptic putative type III PKS enzymes we replaced the *gcs* (SCO7221) gene on the chromosome of *S. coelicolor* with a cassette conferring apramycin resistance using PCR targeting (Song *et al.*, 2006). Comparison of the metabolites in the organic extracts from culture supernatants of wild-type *S. coelicolor* and the *gcs* mutant using LC-MS identified five compounds that were present in the wild-type, but lacking in the mutant (Song *et al.*, 2006). The major component of this complex was purified from the extract and analysed by ESI-TOF-MS, ESI-MS/MS, and 1- and 2D NMR spectroscopy (Song *et al.*, 2006). From these analyses the compound was deduced to be germicidin A, a known inhibitor of streptomyces spore germination, isolated along with its desmethyl analogue germicidin B from *Streptomyces viridochromogenes* (Petersen *et al.*, 1993; Fig. 16). Feeding of stable isotope-labelled precursors and LC-MS/MS analyses together with NMR and MS analyses of another purified component of the mixture established the structures of the other five members of the complex as the other known compound germicidin B and three new compounds, which were named isogermicidin A, isogermicidin B and germicidin C (Song *et al.*, 2006; Fig. 16).

New type III PKS enzymology in germicidin biosynthesis

The identification of germicidins as the products of a type III PKS was quite unexpected. The feeding studies with

stable isotope-labelled precursors indicated that germicidin A is assembled from one molecule of 2-methylbutyryl-CoA, one molecule of malonyl-CoA and one molecule of ethylmalonyl-CoA. (Fig. 17). There appears to be no precedent for the utilization of ethylmalonyl-CoA as an extender unit by type III PKSs, although it has been proposed to be used in this capacity by type I modular PKSs (Erb *et al.*, 2007). To establish whether the *gcs* gene is sufficient for germicidin production in streptomyces, we integrated a plasmid containing *gcs* under the control of the constitutive *ermE*^{*} promoter into the chromosome of *Streptomyces venezuelae*, which is known not to contain a *gcs* orthologue and does not produce germicidins (Song *et al.*, 2006). The resulting strain produced all five germicidins, showing that *gcs* is indeed the only gene required for germicidin biosynthesis in streptomyces (Song *et al.*, 2006).

These results suggested two alternative possible roles for the type III PKS encoded by *gcs* in germicidin biosynthesis. In the first role, the PKS catalyses condensation of an acyl-CoA starter unit (preferentially 2-methylbutyryl-CoA or isobutyryl-CoA) with a molecule of malonyl-CoA, followed by a molecule of ethylmalonyl- or methylmalonyl-CoA to form a triketide, which undergoes cyclization with loss of CoASH to yield the corresponding pyrone (Fig. 18). In the second role, the PKS catalyses elongation of the β -ketoacyl-ACP products of the FabH fatty acid biosynthetic enzyme with a molecule of ethylmalonyl- or methylmalonyl-CoA, followed by cyclization of the resulting triketide to form

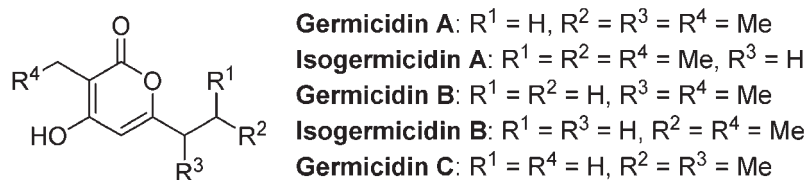


Fig. 16. Structures of germicidins isolated from *S. coelicolor* by genome mining.

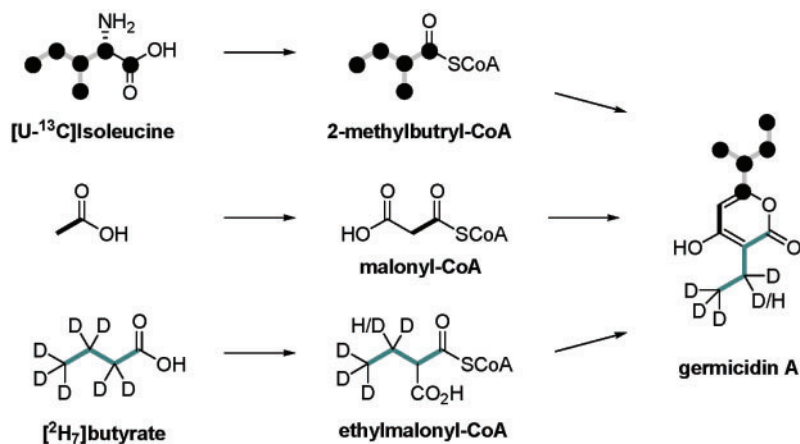


Fig. 17. Incorporation pattern of stable-isotope-labelled precursors into germicidin A and proposed direct acyl-CoA precursors of this antibiotic.

the pyrone (Fig. 19). Both of these hypothetical roles involve novel type III PKS enzymology. In the first potential role, the PKS would catalyse successive rounds of chain extension with different extender units, which appears to be without precedent for iterative PKSs. In the second potential role, the PKS would catalyse transacylation of a β -ketoacyl thioester (formed by the primary metabolic fatty acid biosynthesis enzyme FabH) from an ACP to the active-site cysteine residue of the PKS and elongation with ethylmalonyl-CoA (or methylmalonyl-CoA).

To discriminate between the two possible roles of the type III PKS, we analysed germicidin production in a mutant of *S. coelicolor* in which the gene encoding the endogenous FabH protein (which preferentially elongates branched-chain acyl-CoA thioesters, especially 2-methylbutyryl-CoA and 3-methylbutyryl-CoA, with malonyl-CoA) was replaced with the gene encoding the FabH enzyme from *Escherichia coli* (which preferentially elongates acetyl-CoA with malonyl-CoA) (Li *et al.*, 2005). In wild-type *S. coelicolor* fatty acids are derived predominantly from the

branched-chain 2-methylbutyryl-CoA, 3-methylbutyryl-CoA and isobutyryl-CoA starter units (74%, the remaining 26% are derived from straight-chain starter units; Fig. 20). On the other hand, in the *S. coelicolor* mutant in which the endogenous *fabH* gene has been replaced with its counterpart from *E. coli*, 88% of the fatty acids are derived from straight-chain starter units (acetyl-CoA, propionyl-CoA and possibly butyryl-CoA), while the remaining 12% derive from an isobutyryl-CoA starter unit (Li *et al.*, 2005; Fig. 20). In the same mutant, only germicidin B and isogermicidin B, which are derived from isobutyryl-CoA and butyryl-CoA starter units, respectively, are produced (Song *et al.*, 2006; Fig. 20). This result implicates FabH in germicidin assembly and strongly suggests that the germicidin type III PKS elongates ACP-bound β -ketoacyl thioester intermediates in fatty acid biosynthesis with ethylmalonyl-CoA and catalyses cyclization of the resulting triketides to form the corresponding pyrones. Sherman and coworkers have recently reported biochemical evidence that supports our hypothesis that the germicidin type III PKS utilizes an acyl-ACP starter unit (Grüschow *et al.*, 2007)

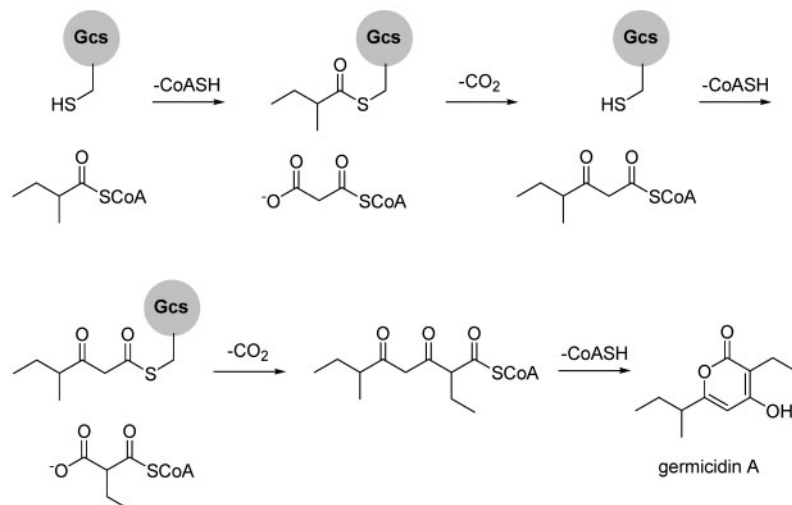


Fig. 18. Possible role for the type III PKS Gcs in germicidin A assembly. This role is inconsistent with the experimental data.

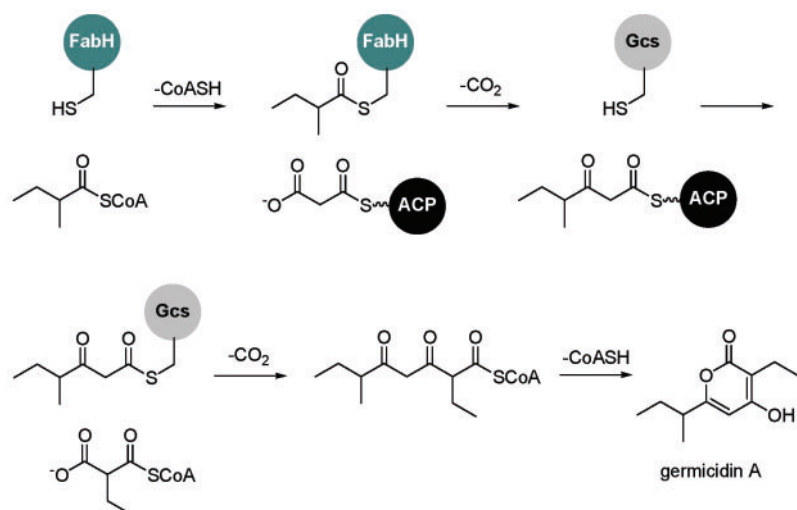
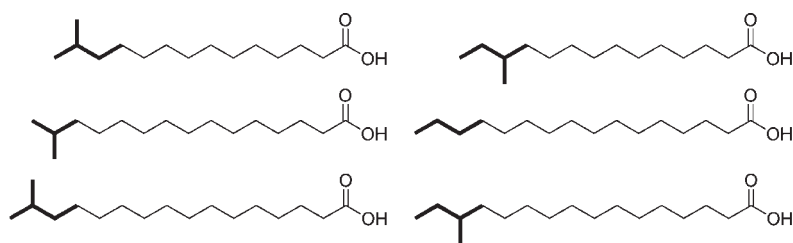


Fig. 19. Alternative possible role for the type III PKS Gcs in germicidin A assembly and possible roles of the FabH enzyme and ACP involved in fatty acid biosynthesis in *S. coelicolor*. These roles are consistent with the available experimental data.

Contemporaneous with our studies on germicidin biosynthesis, Joseph Noel and co-workers discovered a unique type I fatty acid synthase (FAS)-type III PKS hybrid that is involved in differentiation-inducing factor biosynthesis in *Dictyostelium discoideum* (Austin *et al.*, 2006). The arrangement of catalytic domains in this hybrid multi-enzyme suggests that the type III PKS-like domain extends an acyl-ACP thioester, assembled by the type I FAS domains, with three malonyl-CoA units and catalyses subsequent cyclization and aromatization reactions, providing a second potential example of a type III PKS that

utilizes an acyl-ACP starter unit assembled by a FAS. Very recently, Horinouchi and coworkers have demonstrated biochemically that an acyl-ACP thioester assembled on a type I FAS is elongated with three molecules of malonyl-CoA, and subsequently cyclized and aromatized, by a type III PKS in the biosynthesis of phenolic lipids in *Azotobacter vinelandii* (Miyanaga *et al.*, 2008). Thus the emerging view is that utilization of FAS-assembled acyl-ACP starter units by type III PKSs may be a relatively common phenomenon. On the other hand, utilization of ethylmalonyl-CoA as an extender unit by a type III PKS

Wild-type *S. coelicolor*



S. coelicolor mutant in which *fabH* gene is replaced with *E. coli* homologue

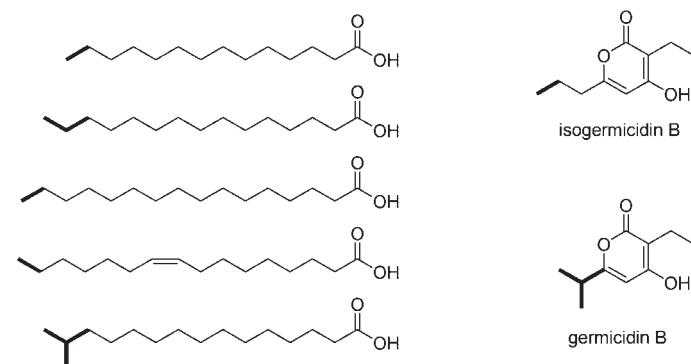


Fig. 20. Main fatty acids produced by wild-type *S. coelicolor* and an *S. coelicolor* mutant in which the *fabH* gene has been replaced by its functional homologue from *E. coli*. The structures of the germicidins produced by the mutant are also shown. Putative starter units for the fatty acids and germicidins produced are highlighted with bold bonds.

thus far seems to remain a unique property of germicidin synthase.

Conclusions

Mining of the *S. coelicolor* genome has yielded one of the first examples of a novel natural product to be discovered by a genomics-guided approach and a rich vein of new biosynthetic chemistry. Our efforts to mine the genome of *S. coelicolor* and other streptomycetes are ongoing and have resulted in the discovery of further new natural products with fascinating biological activities and novel mechanisms of biosynthetic assembly. The details of these discoveries will be disclosed in the near future.

Over the past four years, the field of genome mining for new natural products has exploded. There have been numerous reports of the discovery of novel natural products from sequenced microbes by genomics-guided approaches, hinting that we may be on the verge of a second golden age of new bioactive natural product discovery. Experimental techniques for discovery of the products of cryptic biosynthetic gene clusters continue to be developed and refined. One important challenge that lies ahead is the development of general methods for activating the expression of silent cryptic biosynthetic gene clusters.

Much new biosynthetic chemistry has been discovered over the past 20 years by sequencing-based approaches. Progress in this area has accelerated recently thanks to the avalanche of genomic information since the turn of the century. Doubtless, much new, exciting and intriguing biosynthetic chemistry remains to be discovered in the future through the continued exploitation of information from genome sequencing projects.

Acknowledgements

I would like to express my sincere gratitude to Professor Sir David Hopwood FRS and Professor Mervyn Bibb, who nominated me for the 2007 Fleming Prize Lectureship of the SGM. I am indebted to several talented co-workers whose names appear on the articles from my group cited in this paper for their dedicated and enthusiastic input into our genome mining projects over the past seven years. I thank Daniel-Oves Costales for helpful comments on the manuscript. The Wellcome Trust (Grant ref. 053086), BBSRC (Grant refs EGH16081 and B16610) and the European Union (Integrated Project Actinogen Contract no. 005224 and Marie Curie Intra-European Fellowship Contract no. MEIF-CT-2003-501686) are gratefully acknowledged for funding our work in the genome mining area.

References

- Austin, M. B. & Noel, J. P. (2003). The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep* **20**, 79–110.
- Austin, M. B., Saito, T., Bowman, M. E., Haydock, S., Kato, A., Moore, B. S., Kay, R. R. & Noel, J. P. (2006). Biosynthesis of *Dictyostelium*

discoideum differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase. *Nat Chem Biol* **2**, 494–502.

Banskota, A. H., McAlpine, J. B., Sorensen, D., Aouidate, M., Pirae, M., Alarco, A.-M., Omura, S., Shiomi, K., Farnet, C. M. & Zazopoulos, E. (2006a). Isolation and identification of three new 5-alkenyl-3,3(2H)-furanones from two *Streptomyces* species using a genomic screening approach. *J Antibiot* **59**, 168–176.

Banskota, A. H., McAlpine, J. B., Sorensen, D., Ibrahim, A., Aouidate, M., Pirae, M., Alarco, A.-M., Farnet, C. M. & Zazopoulos, E. (2006b). Genomic analyses lead to novel secondary metabolites. Part 3 ECO-0501, a novel antibacterial of a new class. *J Antibiot* **59**, 533–542.

Barona-Gómez, F., Wong, U., Giannakopoulos, A. E., Derrick, P. J. & Challis, G. L. (2004). Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. *J Am Chem Soc* **126**, 16282–16283.

Barona-Gómez, F., Lautru, S., Francou, F.-X., Leblond, P., Pernodet, J.-L. & Challis, G. L. (2006). Multiple biosynthetic and uptake systems mediate siderophore-dependent iron acquisition in *Streptomyces coelicolor* A3(2) and *Streptomyces ambofaciens* ATCC 23877. *Microbiology* **152**, 3355–3366.

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. & other authors (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147.

Bergmann, S., Schuemann, J., Scherlach, K., Lange, C., Brakhage, A. A. & Hertweck, C. (2007). Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. *Nat Chem Biol* **3**, 213–217.

Bok, J. W., Hoffmeister, D., Maggio-Hall, L. A., Renato, M., Glasner, J. D. & Keller, N. P. (2006). Genomic mining for *Aspergillus* natural products. *Chem Biol* **13**, 31–37.

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

Chain, E., Florey, H. W., Gardner, A. D., Heatley, N. G., Jennings, M. A., Orr-Ewing, J. & Sanders, A. G. (1940). Penicillin as a chemotherapeutic agent. *Lancet* **2**, 226–228.

Challis, G. L. (2006). Engineering *E. coli* to produce nonribosomal peptide antibiotics. *Nat Chem Biol* **2**, 398–400.

Challis, G. L. (2008). Genome mining for new natural product discovery. *J Med Chem* **51** (in press).

Challis, G. L. & Ravel, J. (2000). 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* **187**, 111–114.

Challis, G. L., Ravel, J. & Townsend, C. A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* **7**, 211–224.

Chen, H., Tseng, C. C., Hubbard, B. K. & Walsh, C. T. (2001). Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. *Proc Natl Acad Sci U S A* **98**, 14901–14906.

Chen, X. H., Koumoutsi, A., Scholz, R., Eisenreich, A., Schneider, K., Heinemeyer, I., Morgenstern, B., Voss, B., Hess, W. R. & other authors (2007). Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. *Nat Biotechnol* **25**, 1007–1014.

Corre, C. & Challis, G. L. (2007). Heavy tools for genome mining. *Chem Biol* **14**, 7–9.

de Bruijn, I., de Kock, M. J. D., Yang, M., de Waard, P., van Beek, T. A. & Raaijmakers, J. M. (2007). Genome-based discovery, structure

- prediction and functional analysis of cyclic lipopeptide antibiotics in *Pseudomonas* species. *Mol Microbiol* **63**, 417–428.
- Drake, E. J., Cao, J., Qu, J., Shah, M. B., Straubinger, R. M. & Gulick, A. M. (2007). The 1.8 Å crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of *Pseudomonas aeruginosa*. *J Biol Chem* **282**, 20425–20434.
- Erb, T. J., Berg, I. A., Brecht, V., Mueller, M., Fuchs, G. & 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 U S A* **104**, 10631–10636.
- Fischbach, M. A. & Walsh, C. T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* **106**, 3468–3496.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A. & other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.
- Fleming, A. (1929). On the antibacterial action of cultures of *Penicillium*, with special reference to their use in the isolation of *B. influenzae*. *Br J Exp Pathol* **10**, 226–236.
- Funa, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y. & Horinouchi, S. (1999). A new pathway for polyketide synthesis in microorganisms. *Nature* **400**, 897–899.
- Gross, H. (2007). Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Appl Microbiol Biotechnol* **75**, 267–277.
- Gross, H., Stockwell, V. O., Henckels, M. D., Nowak-Thompson, B., Loper, J. E. & Gerwick, W. H. (2007). The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters. *Chem Biol* **14**, 53–63.
- Grüschow, S., Buchholz, T. J., Seufert, W., Dordick, J. S. & Sherman, D. H. (2007). Substrate profile analysis and ACP-mediated acyl transfer in *Streptomyces coelicolor* type III polyketide synthases. *ChemBioChem* **8**, 863–868.
- Günter, K., Toupet, C. & Schupp, T. (1993). Characterization of an iron-regulated promoter involved in desferrioxamine B synthesis in *Streptomyces pilosus*: repressor-binding site and homology to the diphtheria toxin gene promoter. *J Bacteriol* **175**, 3295–3302.
- Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. (2003). PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A* **100**, 1541–1546.
- 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. (1995). Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA: acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett* **374**, 246–248.
- Haynes, S. W. & Challis, G. L. (2007). Non-linear enzymatic logic in natural product modular MEGA-synthases and synthetases. *Curr Opin Drug Discov Devel* **10**, 203–218.
- Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F., Wilkinson, B., Sidebottom, P. J., Rudd, B. A. M. & other authors (2002). Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem Biol* **9**, 1175–1187.
- Hornung, A., Bertazzo, M., Dziarnowski, A., Schneider, K., Welzel, K., Wohler, S.-E., Holzenkaempfer, M., Nicholson, G. J., Bechthold, A. & other authors (2007). A genomic screening approach to the structure-guided identification of drug candidates from natural sources. *ChemBioChem* **8**, 757–766.
- Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. & Omura, S. (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**, 526–531.
- Izumikawa, M., Shipley, P. R., Hopke, J. N., O'Hare, T., Xiang, L., Noel, J. P. & Moore, B. S. (2003). Expression and characterization of the type III polyketide synthase 1,3,6,8-tetrahydroxynaphthalene synthase from *Streptomyces coelicolor* A3(2). *J Ind Microbiol Biotechnol* **30**, 510–515.
- Kadi, N., Oves-Costales, D., Barona-Gomez, F. & Challis, G. L. (2007). A new family of oligomerisation macrocyclisation biocatalysts. *Nat Chem Biol* **3**, 652–685.
- Keller, N. P., Turner, G. & Bennett, J. W. (2005). Fungal secondary metabolism – from biochemistry to genomics. *Nat Rev Microbiol* **3**, 937–947.
- Koehn, F. E. & Carter, G. T. (2005). The evolving role of natural products in drug discovery. *Nat Rev Drug Discov* **4**, 206–220.
- Larsen, N. A., Lin, H., Wei, R., Fischbach, M. A. & Walsh, C. T. (2006). Structural characterization of enterobactin hydrolase IroE. *Biochemistry* **45**, 10184–10190.
- Lautru, S., Deeth, R. J., Bailey, L. & Challis, G. L. (2005). Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat Chem Biol* **1**, 265–269.
- Lautru, S., Oves-Costales, D., Pernodet, J.-L. & Challis, G. L. (2007). MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. *Microbiology* **153**, 1405–1412.
- Li, T.-L., Choroba, O. W., Hong, H., Williams, D. H. & Spencer, J. B. (2001). Biosynthesis of the vancomycin group of antibiotics: characterisation of a type III polyketide synthase in the pathway to (S)-3,5-dihydroxyphenylglycine. *Chem Commun* 2156–2157.
- Li, Y., Florova, G. & Reynolds, K. A. (2005). Alteration of the fatty acid profile of *Streptomyces coelicolor* by replacement of the initiation enzyme 3-ketoacyl acyl carrier protein synthase III (FabH). *J Bacteriol* **187**, 3795–3799.
- Lin, H., Fischbach, M. A., Liu, D. R. & Walsh, C. T. (2005). In vitro characterization of salmochelin and enterobactin trilactone hydrolases IroD, IroE, and Fes. *J Am Chem Soc* **127**, 11075–11084.
- Lin, X., Hopson, R. & Cane, D. E. (2006). Genome mining in *Streptomyces coelicolor*: molecular cloning and characterization of a new sesquiterpene synthase. *J Am Chem Soc* **128**, 6022–6023.
- McAlpine, J. B., Bachmann, B. O., Pirae, M., Tremblay, S., Alarco, A.-M., Zazopoulos, E. & Farnet, C. M. (2005). Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. *J Nat Prod* **68**, 493–496.
- Miethke, M. & Marahiel, M. A. (2007). Siderophore-based iron acquisition and pathogen control. *Microbiol Mol Biol Rev* **71**, 413–451.
- Minowa, Y., Araki, M. & Kanehisa, M. (2007). Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. *J Mol Biol* **368**, 1500–1517.
- Miyanaaga, A., Funa, N., Awakawa, T. & Horinouchi, S. (2008). Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. *Proc Natl Acad Sci U S A* **105**, 871–876.
- Muller, G. & Raymond, K. N. (1984). Specificity and mechanism of ferrioxamine-mediated iron transport in *Streptomyces pilosus*. *J Bacteriol* **160**, 304–312.
- Nguyen, T., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C. & Piel, J. (2008). Exploiting the mosaic structure of *trans*-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat Biotechnol* **26**, 225–233.
- Oliynyk, M., Samborsky, M., Lester, J. B., Mironenko, T., Scott, N., Dickens, S., Haydock, S. F. & Leadlay, P. F. (2007). Complete

- genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL2338. *Nat Biotechnol* **25**, 447–453.
- Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H. & other authors (2001).** Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc Natl Acad Sci U S A* **98**, 12215–12220.
- Paulsen, I. T., Press, C. M., Ravel, J., Kobayashi, D. Y., Myers, G. S. A., Mavrodi, D. V., DeBoy, R. T., Seshadri, R., Ren, Q. & other authors (2005).** Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat Biotechnol* **23**, 873–878.
- Petersen, F., Zähler, H., Metzger, J. W., Freund, S. & Hummel, R. P. (1993).** Germicidin, an autoregulative germination inhibitor of *Streptomyces viridochromogenes* NRRL B-1551. *J Antibiot* **46**, 1126–1138.
- Pfeifer, V., Nicholson, G. J., Ries, J., Recktenwald, J., Schefer, A. B., Shawky, R. M., Schroder, J., Wohlleben, W. & Pelzer, S. (2001).** A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (S)-3,5-dihydroxyphenylglycine. *J Biol Chem* **276**, 38370–38377.
- Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D. H. (2005).** Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res* **33**, 5799–5808.
- Reid, R., Piagentini, M., Rodriguez, E., Ashley, G., Viswanathan, N., Carney, J., Santi, D. V., Hutchinson, C. R. & McDaniel, R. (2003).** A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* **42**, 72–79.
- Song, L., Barona-Gomez, F., Corre, C., Xiang, L., Udvary, D. W., Austin, M. B., Noel, J. P., Moore, B. S. & Challis, G. L. (2006).** 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* **128**, 14754–14755.
- Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. (1999).** The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6**, 493–505.
- Sudek, S., Haygood, M. G., Youssef, D. T. A. & Schmidt, E. W. (2006).** Structure of trichamide, a cyclic peptide from the bloom-forming cyanobacterium *Trichodesmium erythraeum*, predicted from the genome sequence. *Appl Environ Microbiol* **72**, 4382–4387.
- Tohyama, S., Eguchi, T., Dhakal, R. P., Akashi, T., Otsuka, M. & Kakinuma, K. (2004).** Genome-inspired search for new antibiotics. Isolation and structure determination of new 28-membered polyketide macrolactones, halstoctacosanolides A and B, from *Streptomyces halstedii* HC34. *Tetrahedron* **60**, 3999–4005.
- Udvary, D. W., Zeigler, L., Asolkar, R. N., Singan, V., Lapidus, A., Fenical, W., Jensen, P. R. & Moore, B. S. (2007).** Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Natl Acad Sci U S A* **104**, 10376–10381.
- Wilkinson, B. & Micklefield, J. (2007).** Mining and engineering natural-product biosynthetic pathways. *Nat Chem Biol* **3**, 379–386.
- Zhao, B., Lin, X., Lei, L., Lamb, D. C., Kelly, S. L., Waterman, M. R. & Cane, D. E. (2008).** Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). *J Biol Chem* **283**, 8183–8189.
- Zirkle, R., Black, T. A., Grolach, J., Ligon, J. M. & Molnar, I. (2004).** Analysis of a 108-kb region of the *Saccharopolyspora spinosa* genome covering the obscurin polyketide synthase locus. *DNA Seq* **15**, 123–134.