Published in final edited form as:

Curr Opin Chem Biol. 2012 April ; 16(1-2): 117–123. doi:10.1016/j.cbpa.2012.01.018.

Combinatorial Biosynthesis of Polyketides – A Perspective

Fong T. Wong^a and Chaitan Khosla^{a,b,*}

^aDepartment of Chemical Engineering, Stanford University, Stanford, California 94305

^bDepartment of Chemistry, Stanford University, Stanford, California 94305

Abstract

Since their discovery, polyketide synthases have been attractive targets of biosynthetic engineering to make “unnatural” natural products. Although combinatorial biosynthesis has made encouraging advances over the past two decades, the field remains in its infancy. In this enzyme-centric perspective, we discuss the scientific and technological challenges that could accelerate the adoption of combinatorial biosynthesis as a method of choice for the preparation of encoded libraries of bioactive small molecules. Borrowing a page from the protein structure prediction community, we propose a periodic challenge program to vet the most promising methods in the field, and to foster the collective development of useful tools and algorithms.

Introduction

Polyketides are a structurally diverse but biosynthetically related family of natural products that includes a number of medically important substances such as lovastatin (a cholesterol-lowering agent), erythromycin (an antibiotic), and FK506 (an immunosuppressant) [1]. Their structural and stereochemical complexity makes systematic chemical manipulation a formidable undertaking. Consequently, there has been considerable interest in the potential of harnessing combinatorial biosynthesis to introduce novel functionality into these bioactive compounds and to produce altogether new chemotypes.

In this review, combinatorial biosynthesis is defined as the genetic manipulation of two or more enzymes involved in polyketide biosynthesis. According to this enzyme-centric definition, combinatorial biosynthesis could even yield the natural product itself, as long as the corresponding polyketide synthase (PKS) harbors two or more genetically modified enzymes. These enzymatic modifications can be accomplished by either genetic manipulation of the original enzyme or by replacing it with a homolog (although the latter approach is more common at the present time). By contrast, a product-centric definition of combinatorial biosynthesis would encompass natural product analogs with two or more functional group transformations, regardless of how these modifications are achieved. For example, products of combinatorial biosynthesis could be derived via precursor directed biosynthesis or through other metabolic engineering strategies [2]. We have chosen an enzyme-centric definition because, in our opinion, it highlights the fundamental and technological challenges to exploiting the functional modularity of PKSs [3]. Specifically,

© 2012 Elsevier Ltd. All rights reserved.

*Corresponding Author, khosla@Stanford.edu, Telephone: (650)-723-6538.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

combinatorial biosynthesis can be achieved by manipulating enzymes responsible for primer unit incorporation, chain elongation, and chain termination.

The State of the Art

The feasibility of combinatorial biosynthesis has been demonstrated in the context of different types of multifunctional PKSs [4-8]. Whereas the architectures of these PKS sub-families are variable, all multifunctional PKSs harbor one or more ketosynthases (KS), acyltransferases (AT) and acyl carrier proteins (ACP). In addition, most PKSs also include auxiliary enzymes such as reductases, dehydratases, transferases, cyclases, and thioesterases. To highlight the scope of combinatorial biosynthesis, here we primarily focus on multimodular PKSs with assembly line architectures. Like an automobile assembly line, these PKSs have multiple way stations (called “modules”), each of which harbors distinct protein domains. Except in a few rare cases, each module is deployed only once in the PKS catalytic cycle; this one-to-one correspondence facilitates convenient mapping of each enzyme domain in the PKS to a unique reaction in the polyketide biosynthetic pathway. A prototypical example of this PKS sub-family is the 6-deoxyerythronolide B synthase (DEBS) (Figure 1) [9]. A particularly impressive showcase for the enzymatic complexity of assembly line PKSs is the FR901464 biosynthetic synthase. (Figure 2) [10]. FR901464 is synthesized by an assembly line encompassing a PKS with several atypical architectural and enzymatic features, including a nonribosomal peptide synthetase and an HMG-CoA reductase.

Phylogenetic and structural analysis of assembly line PKSs suggests that nature has harnessed gene duplication, mutation, and recombination to pursue combinatorial biosynthesis over evolutionary time [11]. In a presumably analogous laboratory investigation, ca. 50 analogs of 6-deoxyerythronolide B were produced by engineering two or more domains of DEBS [5]. The latter study was enabled by the establishment of tools for the reconstitution of complete PKS pathways into genetically amenable hosts such as *Streptomyces coelicolor* [12, 13].

Notwithstanding encouraging progress over the past two decades [7, 14-17], the promise of rationally guided combinatorial biosynthesis remains unrealized. In the sections that follow, we discuss key ecological, enzymological, and technological challenges that must be addressed in order to efficiently synthesize libraries of “unnatural” natural products.

Ecological challenges

Until recently, a major obstacle to combinatorial biosynthesis was the availability of DNA sequences of an adequately large number of cloned PKS genes. Less than 20 multifunctional PKS gene clusters had been fully sequenced by the turn of the millennium. As high-throughput sequencing techniques gained momentum, this number increased exponentially. Whereas the growth in PKSs corresponding to structurally characterized natural products has remained modest, the emergence of whole genome sequencing methods has resulted in the discovery of cryptic gene clusters at an explosive pace (Figure 3). Not only has there been an immense growth in the repertoire of enzyme domains and modules, but new assembly line architectures have also been discovered (e.g., “AT-less” PKSs [18, 19]). Today, an aspiring biosynthetic engineer has access to a virtually infinite palette of genetic raw material, although much of it remains to be functionally decoded.

Notwithstanding breathtaking advances in mining nature's PKS gene clusters [20], the ability to identify complete PKSs from unculturable microorganisms remains seriously constrained. The development of resource-efficient strategies for cloning and sequencing large (20-100 kb) contigs from metagenomic sources will enable at least two related types of

opportunities in combinatorial biosynthesis. First, the DNA encoding unprecedented chemotypes could become accessible. For example, close structural analogs of marine natural products such as discodermolide [21] or spongistatin [22] have not yet been isolated from cultured microorganisms. Combinatorial biosynthesis of discodermolide or spongistatin analogs is therefore predicated upon cloning their complete gene clusters. Second, thus far, the vast majority of cloned PKS genes have been isolated from terrestrial bacteria, primarily the actinomycetes, bacilli and myxobacteria. As the genetic content of the earth's oceans is mined for PKSs, new biocatalytic strategies will surely emerge, which in turn could be exploited through combinatorial biosynthesis. For example, an enzyme that catalyzes a Favorskii rearrangement was found in the enterocin biosynthetic pathway from a marine actinomycete [23]. This has led to the engineering of new types of polyketide analogs.

Enzymological challenges

At its core, combinatorial biosynthesis of assembly line PKSs is an exercise in enzyme engineering that rests upon two crucial assumptions. First, individual enzymes along the assembly line must have relaxed substrate specificity. Second, the mechanisms that promote channeling of biosynthetic intermediates from one enzyme to the next must be sufficiently conserved in order to permit the engineering of chimeric assembly lines. Available evidence suggests that both hypotheses are plausible, but lack thorough validation. In the remainder of this section, we review the experimental evidence supporting these hypotheses.

At least three different lines of evidence can be cited in support of the hypothesis that PKS enzymes have broad tolerance for the growing polyketide chain supplied to them. First the substrate specificity of a few PKS modules has been quantified, and is known to be relatively modest (i.e., many substrate analogs have k_{cat}/K_M values within 10-100 fold that of the natural substrate) [24, 25]. Therefore, unless the relevant reaction is a major rate-limiting step in the biosynthetic pathway, a structurally altered biosynthetic intermediate should be well tolerated. Second, precursor directed biosynthesis has been used to convert unnatural primer units or diketides of a number of natural PKSs into the corresponding polyketide analogs [26-28]. Third and perhaps most intriguingly, successive modules of certain PKSs, such as the mycolactone synthase [29], show exceptionally high conservation (>90% identity) in KS and ACP domain sequences, suggesting that module duplication may have been sufficient for the evolution of long, variably functionalized polyketide backbones.

Available structural models for DEBS suggest that the growing polyketide chain is channeled across extraordinary lengths (50-100 nm) before the product is released. In contrast to model systems such as tryptophan synthase and carbamoyl phosphate synthetase, where intermediates are channeled through relatively short (1-10 nm) mostly buried tunnels [30, 31], PKSs rely on selective and dynamic protein-protein interactions. For example, at least three types of protein-protein interactions have been identified in DEBS that strongly influence unidirectional movement of the growing polyketide chain. As shown in Figure 1, the N/C-terminal linkers flanking the three DEBS proteins dock together to constrain the growing chain to a path where module 3 follows module 2 and module 5 follows module 4 [32, 33]. At the same time, the ACP domain of each module preferentially docks onto the downstream module of the assembly line, thus ensuring that the growing chain does not iteratively pass through any module more than once every catalytic cycle [34]. Last but not least, chain elongation by each module also relies on selective intramodular domain-domain interactions [35]. The design of catalytically active PKS chimeras requires each type of non-covalent interaction to be preserved.

Although some progress has been made towards understanding the structure and mechanism of assembly line PKSs, the list of major unanswered questions is considerably longer. Foremost amongst them is the elucidation of PKS structural dynamics. The occurrence of large conformational changes in a multimodular PKS, as it progresses through its catalytic cycle, is all but certain. However, atomic-level insights into this remarkable phenomenon are limited. Other fundamental challenges include the development of broadly applicable experimental strategies to identify rate-limiting steps in natural or engineered PKS assembly lines, and the contribution of intramolecularity to the reaction rates on the assembly line. Last but not least, decoding the function of orphan gene clusters is becoming feasible, as relationships between PKS structure and function become clearer [36]. Deeper insights into PKS enzymology could eventually enable complete automation of this capability.

Technological challenges

In addition to the above challenges in basic biology and chemistry, the toolbox for combinatorial manipulation of PKSs must also be improved. The rapid growth in PKS sequences (Figure 3) will likely accelerate further, as methods for automated assembly of genome-sized contigs from GC-rich organisms are improved. Together with the rapidly decreasing cost of oligonucleotides, this could enable assembly of expression constructs for full-length (i.e. 30-100 kb) natural or modified PKS genes [37]. Of course, heterologous hosts capable of functionally expressing these PKS pathways must be available. Whereas no single host is capable of expressing all types of PKS pathways, *Escherichia coli* and hosts such as *Streptomyces coelicolor* and its close relative *Streptomyces lividans* appear to have a broad scope for this purpose. The key remaining challenge is to improve the polyketide productivity of these hosts. Our own experience suggests that productivity in heterologous hosts is most often not limited by the PKS itself, because specific polyketide productivity is higher in native hosts, even though PKS protein levels are higher in the heterologous host. This is an important challenge for the metabolic engineer [38].

Along with improved methods for PKS cloning and expression, superior methods for detecting and characterizing polyketide products are also needed. Advances in microscale NMR spectroscopy already allow structure elucidation of new natural products with as little as a few nanomoles, thus reducing the sample size by 2-3 orders of magnitude [39]. Similarly, new techniques for ionization and detection of natural products via mass spectrometry open the door to the characterization of very small samples [39]. In both approaches, the problem of contaminants can be particularly vexing. Therefore, robust workflows need to be developed for analyzing trace quantities of a new compound made by an engineered bacterium. Here too, the use of heterologous hosts is an advantage, as isotope-tagging methods could be implemented to differentiate polyketide products from background contaminants.

For combinatorial biosynthesis to gain widespread acceptance as a method for producing small molecule libraries, automation at all levels of experimental design is essential. In recent years, programs such as antiSMASH [40] and Clustscan [41] have been developed to accurately annotate PKSs by identifying domains and even predicting the structures of their natural products. Large databases of sequenced PKSs could also spawn new evolutionary strategies for designing assembly lines with unnatural specificity [11, 42]. Perhaps most intriguingly, newer computational tools (such as SBSPKS [43], [44]) are attempting to perform structure-based analysis of PKSs [45]. With further development and installation of appropriate user interfaces, these programs could eventually serve as portals for combinatorial library design.

Combinatorial biosynthesis: A call to action

After nearly two decades, combinatorial biosynthesis remains in its infancy. In the future, large-scale combinatorial biosynthesis will require a catalog of validated domains, didomains, modules, and linkers capable of performing the spectrum of catalytic chemistry that is observed in nature's assembly lines. Their salient characteristics will be well documented in order to allow the engineer to rationally choose the right components for rationally designing a chimeric assembly line. The catalog will be accompanied by an instruction manual, and reference data from a set of control experiments.

As is perhaps obvious to any natural products biosynthetic chemist, this is an ambitious undertaking. How might one get from here to there? We suggest borrowing a page from the protein structure prediction community and its longstanding CASP challenge program intended to advance automated methods for protein structure prediction. Started in 1994, CASP is approaching its tenth biennial competition, and has been a major driving force for the emergence of the most widely used methods for structure prediction and homology modeling [46]. Importantly, efforts such as CASP have not only fostered useful tools for protein engineers, but have also facilitated scientific progress in the field. An analogous format for combinatorial biosynthesis could be contemplated. In each competition cycle, two (or more) well-defined problems would be presented to the community, and all resources (DNA, vectors, hosts, sequences, reference compounds, etc.) would be made available to labs wishing to participate in the competition. At least two types of problems are envisioned in each competition round;

- The first set would require prediction of the product structure of an orphan PKS, where the actual structure is known but not yet published.
- The second set would require engineering a PKS that produces the highest titer of a target synthon in a defined heterologous host.

The best solutions will not only be widely publicized, but are also likely to gain rapid acceptance as benchmarks for next-generation challenges. We welcome suggestions regarding how such an approach might be further tailored to catalyze rapid advances in combinatorial biosynthesis.

Acknowledgments

Research has been supported by NIH grant GM087934 and by a National Science Scholarship from the Agency of Science, Technology and Research (A*STAR), Singapore, to F.T.W

References

- special interest
 - • outstanding interest
1. Staunton J, Weissman KJ. Polyketide biosynthesis: a millennium review. *Nat Prod Rep*. 2001; 18:380–416. [PubMed: 11548049]
 2. Pickens LB, Tang Y, Chooi Y. *Metabolic Engineering for the Production of Natural Products*. *Annu Rev Chem Biomol Eng*. 2011; 2:211–236. [PubMed: 22432617]
 3. Khosla C, Kapur S, Cane DE. Revisiting the modularity of modular polyketide synthases. *Curr Opin Chem Biol*. 2009; 13(2):135–143. [PubMed: 19217343]
 4. McDaniel R, Ebert-Khosla S, Hopwood DA, Khosla C. Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. *Nature*. 1995; 375:549–554. [PubMed: 7791871]

5. McDaniel R, Thamchaipenet A, Gustafsson C, Fu H, Betlach M, Betlach M, Ashley G. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel “unnatural” natural products. *Proc Natl Acad Sci U S A*. 1999; 96:1846–1851. [PubMed: 10051557]
6. Tang Y, Lee TS, Khosla C. Engineered Biosynthesis of Regioselectively Modified Aromatic Polyketides Using Bimodular Polyketide Synthases. *PLoS Biol*. 2004; 2:e31. [PubMed: 14966533]
7. Gulder, T.; Freeman, M.; Piel, J. The Catalytic Diversity of Multimodular Polyketide Synthases: Natural Product Biosynthesis Beyond Textbook Assembly Rules; [Without Title]. Springer; Berlin/Heidelberg: 2011. p. 1-53.
8. Crawford JM, Townsend CA. New insights into the formation of fungal aromatic polyketides. *Nat Rev Micro*. 2010; 8:879–889.
9. Khosla C, Tang Y, Chen AY, Schnarr NA, Cane DE. Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annu Rev Biochem*. 2007; 76:195–221. [PubMed: 17328673]
10. • Zhang F, He H, Tang M, Tang Y, Zhou Q, Tang G. Cloning and Elucidation of the FR901464 Gene Cluster Revealing a Complex Acyltransferase-less Polyketide Synthase Using Glycerate as Starter Units. *J Am Chem Soc*. 2011; 133:2452–2462. Zhang et al describe the 93 kbp multimodular FR901464 PKS gene cluster encompassing an impressive collection of unique architecture and enzymatic features, such as AT-less NRPS/PKS and HMG-CoA reductase. [PubMed: 21291275]
11. Ridley CP, Lee HY, Khosla C. Evolution of polyketide synthases in bacteria. *Proc Natl Acad Sci U S A*. 2008; 105:4595–4600. [PubMed: 18250311]
12. Kao C, Katz L, Khosla C. Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science*. 1994; 265(5171):509–512. [PubMed: 8036492]
13. Zhang H, Boghigian BA, Armando J, Pfeifer BA. Methods and options for the heterologous production of complex natural products. *Nat Prod Rep*. 2011; 28:125–151. [PubMed: 21060956]
14. Jackie Tsoi C, Khosla C. Combinatorial biosynthesis of “unnatural” natural products: the polyketide example. *Chem Biol*. 1995; 2:355–362. [PubMed: 9383437]
15. Katz, L. Chapter 6 The DEBS Paradigm for Type I Modular Polyketide Synthases and Beyond. In: Hopwood, David A., editor. *Methods in Enzymology; Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates*. Academic Press; 2009. p. 113-142.
16. Cane DE. Programming of Erythromycin Biosynthesis by a Modular Polyketide Synthase. *J Biol Chem*. 2010; 285:27517–27523. [PubMed: 20522551]
17. • Fisch KM, Bakeer W, Yakasai AA, Song Z, Pedrick J, Wasil Z, Bailey AM, Lazarus CM, Simpson TJ, Cox RJ. Rational Domain Swaps Decipher Programming in Fungal Highly Reducing Polyketide Synthases and Resurrect an Extinct Metabolite. *J Am Chem Soc*. 2011; 133:16635–16641. Fisch et al, 2011, using combinatorial biosynthesis between two closely-related highly reduced fungal polyketide synthases, has shown that control of methylation pattern and chain length for these iterative synthases can be interestingly contributed to tailoring domains. Exploiting these properties, an extinct metabolite was also restored. [PubMed: 21899331]
18. Piel J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep*. 2010; 27:996–1047. [PubMed: 20464003]
19. Cheng, Y.; Coughlin, JM.; Lim, S.; Shen, B. Chapter 8 Type I Polyketide Synthases That Require Discrete Acyltransferases. In: Hopwood, David A., editor. *Methods in Enzymology; Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates*. Academic Press; 2009. p. 165-186.
20. Dorrestein PC, Carroll KS. ‘Omics’ of natural products and redox biology. *Curr Opin Chem Biol*. 2011; 15:3–4. [PubMed: 21282071]
21. Florence GJ, Gardner NM, Paterson I. Development of practical syntheses of the marine anticancer agents discodermolide and dictyostatin. *Nat Prod Rep*. 2008; 25:342–375. [PubMed: 18389141]
22. Bai R, Cichacz ZA, Herald CL, Pettit GR, Hamel E. Spongistatin 1, a highly cytotoxic, sponge-derived, marine natural product that inhibits mitosis, microtubule assembly, and the binding of vinblastine to tubulin. *Mol Pharmacol*. 1993; 44:757–766. [PubMed: 8232226]

23. Xiang L, Kalaitzis JA, Moore BS, Benkovic SJ. EncM, a Versatile Enterocin Biosynthetic Enzyme Involved in Favorskii Oxidative Rearrangement, Aldol Condensation, and Heterocycle-Forming Reactions. *Proc Natl Acad Sci U S A*. 2004; 101:15609–15614. [PubMed: 15505225]
24. Chuck J, McPherson M, Huang H, Jacobsen JR, Khosla C, Cane DE. Molecular recognition of diketide substrates by a β -ketoacyl-acyl carrier protein synthase domain within a bimodular polyketide synthase. *Chem Biol*. 1997; 4:757–766. [PubMed: 9375254]
25. Wu N, Kudo F, Cane DE, Khosla C. Analysis of the Molecular Recognition Features of Individual Modules Derived from the Erythromycin Polyketide Synthase. *J Am Chem Soc*. 2000; 122(20): 4847–4852.
26. Jacobsen JR, Hutchinson CR, Cane DE, Khosla C. Precursor-Directed Biosynthesis of Erythromycin Analogs by an Engineered Polyketide Synthase. *Science*. 1997; 277(5324):367–369. [PubMed: 9219693]
27. Marsden AFA, Wilkinson B, Cortés J, Dunster NJ, Staunton J, Leadlay PF. Engineering Broader Specificity into an Antibiotic-Producing Polyketide Synthase. *Science*. 1998; 279(5348):199–202. [PubMed: 9422686]
28. Gregory MA, Petkovic H, Lill RE, Moss SJ, Wilkinson B, Gaisser S, Leadlay PF, Sheridan RM. Mutasynthesis of rapamycin analogues through manipulation of a gene governing starter unit biosynthesis. *Angew Chem Intl Ed Engl*. 2005; 44:4757–4760.
29. Stinear TP, Mve-Obiang A, Small PLC, Frigui W, Pryor MJ, Brosch R, Jenkin GA, Johnson PDR, Davies JK, Lee RE, Adusumilli S, Garnier T, Haydock SF, Leadlay PF, Cole ST. Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc Natl Acad Sci U S A*. 2004; 101(5):1345–1349. [PubMed: 14736915]
30. Miles EW. Tryptophan synthase: A multienzyme complex with an intramolecular tunnel. *Chem Rec*. 2001; 1:140–151. [PubMed: 11893063]
31. Thoden JB, Huang X, Rauschel FM, Holden HM. Carbamoyl-phosphate Synthetase. *J Biol Chem*. 2002; 277:39722–39727. [PubMed: 12130656]
32. Gokhale RS, Tsuji SY, Cane DE, Khosla C. Dissecting and Exploiting Intermodular Communication in Polyketide Synthases. *Science*. 1999; 284:482–485. [PubMed: 10205055]
33. Tsuji SY, Wu N, Khosla C. Intermodular Communication in Polyketide Synthases: Comparing the Role of Protein-Protein Interactions to Those in Other Multidomain Proteins. *Biochemistry*. 2001; 40:2317–2325. [PubMed: 11327851]
34. Wu N, Cane DE, Khosla C. Quantitative Analysis of the Relative Contributions of Donor Acyl Carrier Proteins, Acceptor Ketosynthases, and Linker Regions to Intermodular Transfer of Intermediates in Hybrid Polyketide Synthases. *Biochemistry (NY)*. 2002; 41:5056–5066.
35. Kapur S, Chen AY, Cane DE, Khosla C. Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc Natl Acad Sci U S A*. 2010; 107:22066–22071. Intramodular chain elongation and intermodular chain translocation were shown to occur via distinct KS-ACP interfaces. The modularity of the ACP domain could be exploited for combinatorial biosynthesis. [PubMed: 21127271]
36. Laureti L, Song L, Huang S, Corre C, Leblond P, Challis GL, Aigle B. Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc Natl Acad Sci U S A*. 2011; 108:6258–6263. Constitutive expression of pathway-specific LAL regulator was used to activate the cryptic Stambomycin PKS pathway. Resolution of the resulting metabolite products allowed for elucidation of unique chemical features that were not predicted from sequence analyses. This study demonstrates an efficient strategy for discovering new metabolites and even PKS features through the activation of cryptic PKSs. [PubMed: 21444795]
37. Ellis T, Adie T, Baldwin GS. DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr Biol*. 2011; 3:109–118.
38. Keasling JD. Manufacturing Molecules Through Metabolic Engineering. *Science*. 2010; 330:1355–1358. [PubMed: 21127247]
39. Molinski TF. NMR of natural products at the ‘nanomole-scale’. *Nat Prod Rep*. 2010; 27:321–329. [PubMed: 20179874]

- 40•. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 2011; 39:W339–W346. A user-friendly interface for a comprehensive analysis of biosynthetic gene clusters is introduced. The genomic analysis combines current gene analysis tools for identification of secondary metabolite pathways with the additional analysis of flanking regions to include possible accessory proteins. [PubMed: 21672958]
41. Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D. ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Res.* 2008; 36:6882–6892. [PubMed: 18978015]
- 42•. Zucko J, Cullum J, Hranueli D, Long PF. Evolutionary dynamics of modular polyketide synthases, with implications for protein design and engineering. *J Antibiot.* 2011; 64:89–92. Instead of dissecting how nature accomplishes chemical innovation, this study goes back to basics and examines natural selection pressure on amino acid sequences between similar PKS modules. The synonymous-to-non-synonymous nucleotide substitution ratios across PKS modules were calculated using a sliding window approach. Interesting KSAT linker motifs, possibly involved in functionality of the dimeric complex, were also observed in this study. [PubMed: 21102600]
- 43•. Anand S, Prasad MVR, Yadav G, Kumar N, Shehara J, Ansari MZ, Mohanty D. SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Res.* 2010; 38:W487–W496. Two servers are described for the 3D visualization of complete PKS modules (Model_3D_PKS) and for sequence-based analysis of interpolyketide linkers' interactions (Dock_Dom_Anal). [PubMed: 20444870]
44. Starcevic A, Diminic J, Zucko J, Elbekali M, Schlosser T, Lisfi M, Vukelic A, Long P, Hranueli D, Cullum J. A novel docking domain interface model predicting recombination between homoeologous modular biosynthetic gene clusters. *J Ind Microbiol Biotechnol.* 2011; 38:1295–1304. [PubMed: 21107638]
45. Yadav G, Gokhale RS, Mohanty D. Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. *J Mol Biol.* 2003; 328:335–363. [PubMed: 12691745]
46. Moult J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr Opin Struct Biol.* 2005; 15:285–289. [PubMed: 15939584]

Figure 1.

The 6-deoxyerythronolide B synthase (3 genes, 32 kbp) is a canonical multimodular PKS [9]. The growing chain is shown as it moves down the assembly line. N and C terminal linkers are also shown. KS: ketosynthase, AT: acyltransferase, DH: dehydratase, ER: enoyl reductase, KR: ketoreductase, ACP: acyl carrier protein. The ketoreductase domain in module 3 is inactive (shown in lower caps).

Figure 2.

Modules in biosynthetic pathway of FR901464. FR901464 synthase is encoded by 20 ORFs spanning 93 kb. 3 ORFs encoding accessory tailoring proteins are not shown here [10]. The growing polyketide chain is shown attached to individual ACP domains. Acronyms in lower case refer to non-functional domains. The *trans*-acting AT acts on eight modules, modules 1 and 3-9. KS: ketosynthase, AT: acyltransferase, DH: dehydratase, ER: enoyl reductase, KR: ketoreductase, ACP: acyl carrier protein. GAT: glyceryl transferase/phosphatase, ECH: enoyl-CoA reductase OX: FAD-dependent monooxygenase, MT: methyltransferase, PCP: peptidyl carrier protein, C: condensation, A: adenylation.

Figure 3.

Growth in the number of cloned and sequenced multimodular PKS gene clusters. The cumulative increases over the past decade in the number of orphan PKS gene clusters and structurally characterized polyketides are shown. Data was obtained by calculating the number of “polyketide synthase” entries published each year in the nucleotides database (pubmed, URL: <http://www.ncbi.nlm.nih.gov/nuccore>). The entries were manually screened and sorted into orphan versus characterized PKSs.