Published in final edited form as:

Science. 2008 April 11; 320(5873): 243–246.

Deconstruction of Iterative Multidomain Polyketide Synthase Function

Jason M. Crawford^{1,*}, Paul M. Thomas², Jonathan R. Scheerer¹, Anna L. Vagstad¹, Neil L. Kelleher^{2,†}, and Craig A. Townsend^{1,†}

¹ Department of Chemistry, Johns Hopkins University, Baltimore, MD 21218, USA

² Department of Chemistry, University of Illinois, Urbana, IL 61801, USA

Abstract

PksA, which initiates biosynthesis of the environmental carcinogen aflatoxin B1, is one of the multidomain iterative polyketide synthases (IPKSs), a large, poorly understood family of biosynthetic enzymes. We found that dissection of PksA and its reconstitution from selected sets of domains allows the accumulation and characterization of advanced octaketide intermediates bound to the enzyme, permitting the reactions controlled by individual catalytic domains to be identified. A product template (PT) domain unites with the ketosynthase and thioesterase in this IPKS system to assemble precisely seven malonyl-derived building blocks to a hexanoyl starter unit and mediate a specific cyclization cascade. Because the PT domain is common among nonreducing IPKSs, these mechanistic features should prove to be general for IPKS-catalyzed production of aromatic polyketides.

Tens of thousands of natural products are known from microorganisms, plants, and animals that provide hormones, toxins, flavors and fragrances, pigments, drugs, and other materials of commercial value. A handful of biosynthetic pathways give rise to this rich diversity of useful structures. Among these, polyketides are synthesized from simple acyl-coenzyme A (acyl-CoA) substrates by polyketide synthases (PKSs) (1). We understand a great deal about the function of giant modular PKSs that synthesize complex products, for example, the antibiotic erythromycin and the immunosuppressant rapamycin (2). Each catalytic domain in these multidomain (type I) megaproteins is used once in an overall “assembly-line” process as a growing intermediate is advanced along the enzyme to yield a product. In contrast to these bacterial systems, in eukaryotes iterative PKSs (IPKSs) are generally the rule where a markedly smaller number of catalytic domains are similarly fused but individually reused in multiple catalytic cycles (iteration) that are “programmed” to yield specific products. How programming is achieved is a central unanswered question of iterative catalysis.

We undertook a “deconstruction” approach by domain dissection and reassembly of PksA, the nonreducing IPKS of aflatoxin biosynthesis, to reveal the global division of labor among the domains in these macromolecular machines to control polyketide chain length, cyclization of an extended poly- β -keto intermediate, and product release. A domain hypothesized to be a “product template” (PT) has been discovered to play a central role in product formation. These studies are combined with high-resolution mass spectrometry (MS) to monitor the highly

[†]To whom correspondence should be addressed. E-mail: ctownsend@jhu.edu (C.A.T.); kelleher@scs.uiuc.edu (N.L.K.).

*Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, USA.

31 December 2007; accepted 29 February 2008

10.1126/science.1154711

processive construction of a natural product scaffold from multiple acyl-CoA units directly on a polyketide synthase.

We previously used the UMA bioinformatics algorithm, which identifies linkers in a family of homologous proteins, to characterize the domain architecture of PksA (3). Two unrecognized domains were identified in addition to four from sequence homology. The first was an unusually long N-terminal domain that was found to be a starter unit–acyl carrier protein (ACP) transacylase (SAT), which selectively transfers a hexanoyl unit onto PksA to initiate polyketide synthesis (4,5). The second domain is internal, comprising about 350 amino acids, and lies just N-terminal to the acyl carrier protein (ACP, Fig. 1B). Like the SAT, this domain bore minimal primary sequence similarity to any known protein except similarly placed domains in other known and presumed nonreducing IPKSs. The proximity of this domain to the ACP and our hypothesized need for a stabilizing/cyclizing PT domain to control the reactivity of poly- β -keto intermediates such as **1** led us to investigate its possible role in the synthesis of norsolorinic acid (3), the first isolable intermediate in the biosynthesis of the environmental carcinogen aflatoxin B₁ (4) (Fig. 1A).

To begin dissection of nonreducing IPKS function, we used the cDNA for *pksA* to clone various mono-, di-, and tridomains for combinatorial reconstitution experiments. The protein fragments were expressed in *Escherichia coli* and purified by affinity chromatography as either N- or C-terminal His₆-tag fusions. UMA was used to guide the choice of termini between domains and to accommodate His₆-tag fusions in formerly flexible linker regions to minimize domain-domain interference. To deconstruct the assembly steps of product formation, we independently expressed and purified the ACP, PT, and TE/CLC monodomains; the PT-ACP and ACP-TE/CLC didomains; and the SAT-KS-MAT and PT-ACP-TE/CLC tridomains. The cloned fragments containing the ACP (ACP, PT-ACP, ACP-TE/CLC, or PT-ACP-TE/CLC) were post-translationally modified with a CoA-derived arm in vitro to their active *holo*-forms by means of the promiscuous 4'-phosphopantetheinyltransferase (PPT) Svp (6) (fig. S1).

The individual catalytic roles of the domains in PksA were probed by mixing the acyl-CoA substrates (hexanoyl- and malonyl-CoA) with the loading and condensation tridomain (SAT-KS-MAT) and all possible combinations, fused or independent, of the activated *holo*-carrier (ACP) and tailoring domains (PT and TE/CLC). The small-molecule products of the reactions were analyzed by high-performance liquid chromatography (HPLC). Controls lacking the *holo*-ACP (SAT-KS-MAT and SAT-KS-MAT + PT + TE/CLC), which carries the growing polyketide through the catalytic process, exhibited no detectable product output (Fig. 2, experiments 1 and 2). Incubation in the presence of the *holo*-ACP (SAT-KS-MAT + ACP), but lacking the tailoring domains, yielded a minute quantity of the octaketides, norsolorinic acid (3) and the shunt product, naphthopyrone **5**. The identity of **5**, oxidized to its corresponding naphthoquinone, was assumed on the basis of its mass identity to its isomer norsolorinic acid (**3**, electrospray ionization MS) and the precedent that inactivation of analogous fungal TE/CLC domains led to the accumulation of pyrone products (7). The structure of **5** was proved by comparison to an authentic standard (fig. S2).

The machinery encoding the correct chain length was present in the SAT-KS-MAT + ACP combination alone, although product output was very inefficient. Addition of the TE/CLC domain to this combination of domains reduced the quantity of the naphthopyrone (**5**) and enhanced the ratio of the correctly cyclized norsolorinic acid (**3**), identifying its catalytic role directly in product release through an intramolecular Claisen reaction. Again, product output was low (Fig. 2, experiments 3 to 5). In contrast, addition of the PT domain strongly drove cyclization and dehydration chemistry to form the first two rings, leaving the third and final ring cyclization step unaffected. In the absence of the TE/CLC domain, product release was spontaneous; that is, both the oxygen-carbon (**5**) and carbon-carbon (2 and 3) cyclization

products were formed (Fig. 2, experiments 6 and 7). Finally, the addition of both tailoring domains, PT and TE/CLC, resulted primarily in the correct cyclization chemistry and release of the norsolorinic acid anthrone (2) (Fig. 2, experiments 8 to 11). The anthrone is readily oxidized to the anthraquinone **3** during workup, although this oxidation might be driven enzymatically in vivo, as is known in other fungal pathways (8). Although the free-standing domains are able to interact properly and process faithfully wild-type polyketide formation, it is evident that domain fusion enhances catalytic efficiency (Fig. 2, experiments 4 to 11), which likely contributed to the evolutionary pressure to form iterative type I systems in fungi.

The fungal aromatic products resemble bacterial aromatic PKS (type II) products, such as tetracyclines and actinorhodins, which require biosynthetic tailoring steps. Combinations of the type II PKS components revealed that a chain-length factor (CLF) assists the KS in chain-length control (9–11), but fungal aromatic PKSs lack such domains. Whereas the PksA SAT and MAT (12) domains load the ACP with their respective starter and extender units, we demonstrate here that the fungal nonreducing KS domain carries out catalytic responsibilities similar to that of the KS/CLF complex. The initial aldol cyclization product(s), if set by the KS, are intrinsically reversible, but the PT acts to drive the dehydration/aromatization chemistry to irreversibly form rings A and B and to markedly increase the flux to product(s). Alternatively, the KS simply elongates the polyketide chain to a fixed length and the PT disciplines the possible chemical fates of this reactive species to template cyclization and aromatization chemistry to a specific product. Recent results from a related fungal IPKS support this picture and show that heterologous bacterial type II (dissociated) enzymes can intercept these fungal polyketide intermediates and carry out regioselective reactions (13).

To examine more closely the polyketide assembly process, we repeated all experiments as time-course studies monitoring the change in ultraviolet (UV)–visible absorption behavior (Fig. 3 and fig. S3). The fully reconstituted PksA (SAT-KS-MAT + PT-ACP-TE/CLC) gave little change in visible absorption over 4 hours. In marked contrast, the same two-part combination, but lacking the TE/CLC domain, rapidly became yellow and displayed a long λ_{max} at 388 nm. By varying the PT-ACP concentration with respect to the SAT-KS-MAT tridomain, the intensity of the chromophore absorption was found to correlate linearly with the PT-ACP concentration (fig. S4). After dialysis to remove small molecules, the absorbing species remained either covalently bound to or tightly associated with the polydomain complex. The accumulation of strongly absorbing species suggested that the TE/CLC operates efficiently to sweep a common intermediate generated by the PT domain through Claisen cyclization to norsolorinic acid (**3**), overriding the accumulation of the 388-nm absorbing species. The synthetic naphthopyrone **5** showed a virtually identical spectral signature under the assay conditions. However, addition of the TE/CLC domain to the SAT-KS-MAT + PT-ACP mixture, after the long-wavelength absorption had reached steady state, only partially reduced this intensity; this finding suggests that the bicyclic acyl intermediate **6** destined for the TE/CLC is contributing to the overall absorbance.

Combining the SAT-KS-MAT tridomain with only the *holo*-ACP again gave an intense absorption, but at a considerably shorter wavelength (Fig. 3C), suggesting the formation of less highly conjugated chromophore(s). The intensity of UV-visible absorption was again dependent on ACP concentration. In sum, these experiments support the view that accumulation of ACP-bound intermediate(s) of PksA is modulated by the presence or absence of the PT domain.

To determine the function of the PT domain in the context of a reconstituted system, we combined the SAT-KS-MAT tridomain with the ACP monodomain alone or with the PT-ACP didomain to measure intermediates attached to the ACP by MS. The TE/CLC domain shows broad hydrolytic substrate tolerance for *N*-acetyl cyste-amine thioesters, and (in the absence

of the correct Claisen substrate) likely serves an editing function. For this reason the TE/CLC domain was not included. Reaction mixtures flash-frozen at 30, 120, and 840 min were thawed and trypsin-digested before analysis by HPLC-MS to determine the nature of the covalent intermediates with part-per-million (ppm) mass accuracy. Figure 4 shows the relative intensities of all observed forms of the acylated tryptic peptides containing the pantetheinylated active-site serine of the ACP domain (all <1.5 ppm and eluting as discrete HPLC peaks) for the incubation of 20 μ M PT-ACP or ACP with 5 μ M SAT-KS-MAT in the presence of hexanoyl- and malonyl-CoA. For both reactions, the only iteratively extended products detected as phosphopantetheine-tethered intermediates were from the condensation of one hexanoyl starter unit with seven malonyl units (1). This product was also seen either singly or doubly dehydrated (6) while still tethered to the protein. The complex intermediates were confirmed by the observation of each specific phosphopantetheinyl ejection ion within 1.5 ppm (14). These phosphopantetheinyl ejection ions were further subjected to ion trap fragmentation to provide data corroborating their proposed structures (figs. S5 to S9).

Apart from the relative populations of the fully extended polyketide intermediates, both two-part combination reactions gave remarkably similar outcomes (Fig. 4). Elaboration of only the hexanoyl-primed ACP, and not the acetyl-primed (from KS-mediated loss of CO₂ from malonyl) or malonyl-primed ACP, supports the notion that the KS domain exerts control over the selection and extension of only specific starter units tethered to the ACP domain (in this case hexanoyl-S-ACP). These data support the assertion that the PT domain maintains the correct intramolecular aldol addition intermediate and promotes its dehydration, facilitating nonenzymatic release of either the naphthopyrone **5** or the anthrone **2**. Note that the ACP active-site tryptic peptide with the doubly dehydrated hexanoate + 7 acetate species (**6**, PT product) showed an absorbance under single-wavelength analysis (310 nm), in further support that this species is indeed the TE/CLC substrate. Combined with the drastic reduction in the flux to product(s) in the absence of the tailoring domains (Fig. 2), the mass spectrometric data support assignment of the PT domain as an aromatase/cyclase domain.

The deconstruction approach taken here has demonstrated the synthetic roles of all recognized domains in an IPKS catalytic cycle, notably the product template (PT) domain, which should prove to be general for the broad class of multidomain IPKSs. The insights afforded by the catalytic autonomy of these dissected, free-standing domains enable a rational strategy for engineering these enzymes to synthesize alternative products.

Supplemental Information

Refer to Web version on PubMed Central for supplementary material.

References and Notes

1. Staunton J, Weissman KJ. *Nat Prod Rep* 2001;18:380. [PubMed: 11548049]
2. Fischbach MA, Walsh CT. *Chem Rev* 2006;106:3468. [PubMed: 16895337]
3. Crawford JM, Vagstad AL, Ehrlich KC, Townsend CA. *Bioorg Chem* 2008;36:16. [PubMed: 18215412]
4. Udvary D, Merski M, Townsend CA. *J Mol Biol* 2002;323:585. [PubMed: 12381311]
5. Crawford JM, Dancy BCR, Hill EA, Udvary DW, Townsend CA. *Proc Natl Acad Sci USA* 2006;103:16728. [PubMed: 17071746]
6. Sanchez C, Du L, Edwards DJ, Toney MD, Shen B. *Chem Biol* 2001;8:725. [PubMed: 11451672]
7. Fujii I, Watanabe A, Sankawa U, Ebizuka Y. *Chem Biol* 2001;8:189. [PubMed: 11251292]
8. Chen ZG, Fujii I, Ebizuka Y, Sankawa U. *Phytochem* 1995;38:299.
9. McDaniel R, Ebert-Khosla S, Hopwood DA, Khosla C. *Science* 1993;262:1546. [PubMed: 8248802]
10. Tang Y, Tsai SC, Khosla C. *J Am Chem Soc* 2003;125:12708. [PubMed: 14558809]

11. Keatinge-Clay AT, Maltby DA, Medzihradszky KF, Khosla C, Stroud RM. Nat Struct Mol Biol 2004;11:888. [PubMed: 15286722]
12. Ma Y, et al. ChemBioChem 2006;7:1951. [PubMed: 17086560]
13. Ma SM, et al. J Am Chem Soc 2008;130:38. [PubMed: 18072779]
14. Dorrestein PC, et al. Biochemistry 2006;45:12756. [PubMed: 17042494]
15. Supported by NIH research grants ES001670 (C.A.T.) and GM067725 (N.L.K.), NIH Chemical Biology Interface training grant GM070421 (P.M.T.), and NIH postdoctoral fellowship GM079408 (J.R.S.).

Fig. 1. Aflatoxin B₁ biosynthesis. **(A)** Biosynthetic pathway. PksA accepts a C₆ fatty acid from the associated fatty acid synthase subunits (HexA/HexB) and homologates acetyl units while attached to either the acyl carrier protein (ACP) or ketosynthase (KS) to a fixed chain length (octaketide), as shown in the hypothetical intermediate **1** (shown as the keto form), R = ACP or KS. PksA controls specific cyclization chemistry to release the norsolorinic acid anthrone (**2**), which is auto-oxidized in vitro to norsolorinic acid (**3**). **3** is processed through a series of oxidative rearrangements to aflatoxin B₁ (**4**). **(B)** Fungal nonreducing PKS domain architecture (PksA). Domains include SAT, KS, malonyl-CoA:ACP transacylase (MAT), PT, ACP, and thioesterase/Claisen cyclase (TE/CLC). Many known and apparently homologous enzymes that share similar domain organization synthesize an assortment of products with alternative chain lengths and cyclization scaffolds (5).

Fig. 2.

Metabolite production from two-, three-, and four-part multi-domain combinations. **(A)** HPLC analysis. The enzymatic products were extracted into ethyl acetate, dried, and redissolved in dimethyl sulfoxide. The solution was diluted with mobile phase (80/20 acetonitrile, 0.1% aqueous trifluoroacetic acid), and products were separated over a Prodigy octadecyl sulfate 3 column (100 Å, 5 µm, 4.6 × 250 mm; Phenomenex, Torrance, CA). The naphthopyrone **5** and norsolorinic acid anthrone (**2**) were observed at 280 nm, and norsolorinic acid (**3**) was observed at 310 nm. The boxed area at lower right shows the absorption at 280 nm of anthrone **2** amplified (50×). **(B)** Spontaneous versus enzymatic formation of ring C. The putative enzyme-bound intermediate **6** undergoes spontaneous chemistry in the absence of the TE/CLC domain to form the O-C cyclization product **5** or the C-C cyclization products **2** and **3**. Addition of the TE/CLC, however, leads to the correct C-C cyclization products, confirming its role in catalyzing final cyclization and product release. The enzymatic anthrone product **2** auto-oxidizes to the anthraquinone **3**.

Fig. 3.

Chromophore formation. Enzymatic reactions were monitored over 15-min intervals for 240 min by UV-visible spectrophotometry. **(A)** SAT-KS-MAT control. All constructs by themselves similarly showed no change in absorbance. **(B)** Reconstitution of PksA. A 330-nm shoulder appears before reaching steady state in the presence of both tailoring domains (PT and TE/CLC). **(C)** PksA lacking the tailoring domains. Buildup of less highly conjugated chromophore(s) (~330 nm shoulder) occurs. **(D)** Chromophore formation in the presence of the PT domain. A 388-nm long-wavelength maximum from the formation of more highly conjugated chromophore(s) appears. Figure S3 shows spectra for all of the domain combinations, and the chromophore formation is highly similar for the same domain combinations whether they are tethered or not.

Fig. 4. Iterative intermediates. Relative intensities of acylated peptides from the ACP domain for reactions with the SAT-KS-MAT and either the ACP (A) or the PT-ACP (B). Mass spectrometric confirmation for the structure of each species is provided in figs. S5 to S9. Note that these 4-kD peptides bear different acyl intermediates, creating possible intensity biases stemming from differential ionization efficiencies during electrospray.