Corrections

COMMENTARY. For the article “The drifting human genome,” by Jianzhi Zhang, which appeared in issue 51, December 18, 2007, of Proc Natl Acad Sci USA (104:20147–20148; first published December 10, 2007; 10.1073/pnas.0710524105), the authors note that, due to a printer’s error, the DOI appeared incorrectly. The DOI 10.1073/pnas.0710524105 should have appeared as 10.1073/pnas.0710524104. The online version has been corrected.

BIOCHEMISTRY. For the article “Heterogeneity in EGF-binding affinities arises from negative cooperativity in an aggregating system,” by Jennifer L. Macdonald and Linda J. Pike, which appeared in issue 1, January 8, 2008, of Proc Natl Acad Sci USA (105:112–117; first published December 28, 2007; 10.1073/pnas.0707080105), the authors note that in Fig. 3, in the parameter list inset at lower right, the labels $K_{21}$ and $L_{20}$ are reversed. The correct values for all four parameters are $K_{21} = 4.6 \pm 0.6 \times 10^{9}$, $L_{20} = 5.3 \pm 2.2 \times 10^{11}$, $K_{21} = 5.3 \pm 0.4 \times 10^{9}$, and $K_{22} = 3.4 \pm 1.1 \times 10^{8}$. This error does not affect the conclusions of the article. The corrected figure and its legend appear below.

**Fig. 3.** Binding of EGF to cells expressing increasing levels of wild-type EGF receptors. CHO-K1 tet-on EGFR cells were induced to express EGF receptors by using increasing doses of doxycycline. $^{125}$I-EGF-binding isotherms were generated from each set of cells, and all six isotherms were globally fit to Eq. 1, with only the value of $R_0$ varying among curves. Data points represent the mean ± SD of triplicate determinations. Solid lines represent the fitted curve through the data points of the same color.

The correct values for all four parameters are $K_{21} = 4.6 \pm 0.6 \times 10^{9}$, $L_{20} = 5.3 \pm 2.2 \times 10^{11}$, $K_{21} = 5.3 \pm 0.4 \times 10^{9}$, and $K_{22} = 3.4 \pm 1.1 \times 10^{8}$. This error does not affect the conclusions of the article. The corrected figure and its legend appear below.

**Fig. 3.** Binding of EGF to cells expressing increasing levels of wild-type EGF receptors. CHO-K1 tet-on EGFR cells were induced to express EGF receptors by using increasing doses of doxycycline. $^{125}$I-EGF-binding isotherms were generated from each set of cells, and all six isotherms were globally fit to Eq. 1, with only the value of $R_0$ varying among curves. Data points represent the mean ± SD of triplicate determinations. Solid lines represent the fitted curve through the data points of the same color.

BIOCHEMISTRY. For the article “Crystal structure and functional analysis of tetracenomycin ARO/CYC: Implications for cyclization specificity of aromatic polyketides,” by Brian Douglas Ames, Tyler Paz Korman, Wenjun Zhang, Peter Smith, Thanh Vu, Yi Tang, and Shiou-Chuan Tsai, which appeared in issue 14, April 8, 2008, of Proc Natl Acad Sci USA (105:5349–5354; first published April 3, 2008; 10.1073/pnas.0709223105), the authors note that on page 5353, left column, in Biological Significance and Proposed Mechanism, line 5, “intercalculating” should have appeared as “intercalating.” Additionally, on page 5354, left column, in Gene Expression and Protein Purification, in the last line, “10 mM Hepes (pH 7.0) and NaOH” should instead read: “10 mM Hepes (pH 7.0)/NaOH.” Finally, in the main article text, the numbering and links to some supporting information (SI) figures and tables were incorrect and have been corrected. These errors were printer’s errors and do not affect the conclusions of the article.

BIOCHEMISTRY. For the article “EB1 promotes Aurora-B kinase activity through blocking its inactivation by protein phosphatase 2A,” by Lei Sun, Jinmin Gao, Xin Dong, Min Liu, Dengwen Li, Xingjuan Shi, Jin-Tang Dong, Xianyu Lu, Chunyong Liu, and Jun Zhou, which appeared in issue 20, May 20, 2008, of Proc Natl Acad Sci USA (105:7153–7158; first published May 13, 2008; 10.1073/pnas.0710018105), the authors note that Jinmin Gao (J.G.), and not Jun Zhou (J.Z.), should have been included among those credited with contributing equally to this work. The corrected author contributions footnote appears below. In addition, in the Acknowledgments, a grant number was inadvertently omitted. “This work was supported by grants from the National Key Scientific Program of China (2007CB914503 and 2007CB914802),” should instead read: “This work was supported by grants from the National Key Scientific Program of China (2006CB910103, 2007CB914503, and 2007CB914802).”


www.pnas.org/cgi/doi/10.1073/pnas.0804020105
Crystal structure and functional analysis of tetracenomycin ARO/CYC: Implications for cyclization specificity of aromatic polyketides

Brian Douglas Ames*, Tyler Paz Korman*, Wenjun Zhang†, Peter Smith*, Thanh Vu*, Yi Tang†, and Shiou-Chuan Tsai*‡

*Department of Molecular Biology and Biochemistry, Department of Chemistry, University of California, Irvine, CA 92697; and ‡Department of Chemical Engineering, University of California, Los Angeles, CA 90095

Edited by Christopher T. Walsh, Harvard Medical School, Boston, MA, and approved February 12, 2008 (received for review September 27, 2007)

Polyketides are a class of natural products with highly diverse chemical structures and pharmaceutical activities. Polyketide cyclization, promoted by the aromatase/cyclase (ARO/CYC), helps diversify aromatic polyketides. How the ARO/CYC promotes highly specific cyclization is not well understood because of the lack of a first-ring ARO/CYC structure. The 1.9 Å crystal structure of Tcm ARO/CYC reveals that the enzyme belongs to the Bet v1-like superfamily (or STAR domain family) with a helix–grip fold, and contains a highly conserved interior pocket. Docking, mutagenesis, and an in vivo assay show that the size, shape, and composition of the pocket are important to orient and specifically fold the polyketide chain for C9-C14 first-ring and C7-C16 second-ring cyclizations. Two pocket residues, R69 and Y35, were found to be essential for promoting first- and second-ring cyclization specificity. Different pocket residue mutations affected the polyketide product distribution. A mechanism is proposed based on the structure-mutation-docking results. These results strongly suggest that the regiospecific cyclizations of the first two rings and subsequent aromatizations take place in the interior pocket. The chemical insights gleaned from this work pave the foundation toward defining the molecular rules for the ARO/CYC cyclization specificity, whose rational control will be important for future endeavors in the engineered biosynthesis of novel anticancer and antibiotic aromatic polyketides.

amoratase | cyclase | natural product

Ring formation is a common strategy to diversify natural products, and the molecular basis of cyclization specificity is a vigorously pursued topic in natural product biosynthesis. The aromatic polyketides are a class of natural products with diverse therapeutic activity and chemical structures (1), such as tetracycline (antibiotic) and doxorubicin (anticancer). These aromatic natural products, with diverse cyclization patterns, are produced by type II polyketide synthase (PKS). The polyketide chain is assembled by the “minimal” PKS (MinPKS), which is comprised of the ketosynthease/chain length factor (KS/CLF) heterodimer, and an acyl carrier protein (ACP). Then, the polyketide chain must be folded and regiospecifically cyclized en route to the final aromatic product (Fig. 1A). Genetic studies indicate that the aromatase/cyclase (ARO/CYC) may be important for cyclization specificity (1), whose absence can lead to an array of differently cyclized polyketides (2). How the ARO/CYC controls the cyclization pattern is not well defined, and past attempts to exploit ARO/CYC for controlling cyclization pattern has been hampered by the lack of ARO/CYC crystal structures. C9-C14 and C7-C12 first-ring cyclizations are the two most commonly observed cyclization patterns in type II PKS. Accordingly, there are two types of ARO/CYCs: the mono-domain ARO/CYCs are often associated with C9-C14 first-ring cyclization (Fig. 1, whereas the di-domain ARO/CYCs, which contain two repeats of the ARO/CYC domain, are often associated with PKS that contain ketoreductase (KR) and biosynthesize C7-C12 first-ring cyclized polyketides (3). The best studied mono-domain ARO/CYC is the tetracenomycin ARO/CYC (Tcm ARO/CYC, the N-terminal domain of the bifunctional protein TcmN), whose inclusion in vivo and in vitro promotes C9-C14 first-ring (and C7-C16 second-ring) cyclization (3, 4) (Fig. 1A and B). The Tcm MinPKS produces a linear decaketide (C20) intermediate that is transferred to Tcm ARO/CYC, which is proposed to fold, cyclize, and aromatize the ACP-tethered polyketide by aldol condensation and dehydration reactions (5) (Fig. 1B). After formation of the aromatic third ring and polyketide release from ACP to form Tcm F2 (3), cyclization of the fourth ring is accomplished by TcmI (a cyclase that is structurally and functionally distinct from Tcm ARO/CYC) to produce Tcm F1 (2) (6). Further enzymatic modification of TcmF1 results in the production of TcmC (1). Heterologous expression of the Tcm MinPKS alone results in the production of two differently cyclized products SEK15 (6) and SEK15b (7) (3); however, the inclusion of Tcm ARO/CYC (in the absence of downstream enzymes) produces RM80 (4)/RM80b (5), which have the “natural” cyclization and aromatization pattern for the first two rings (4) (Fig. 1A). How Tcm ARO/CYC promotes so many tandem reactions and achieves the C9-C14 cyclization specificity is not well understood, and it is unclear whether ARO/CYC ever binds the polyketide substrate or whether it serves only as a helper protein that stabilizes a protein–protein interface (such as between KS/CLF and ACP) where cyclization takes place.

Here, we present the 1.9 Å crystal structure and mutational analysis of Tcm ARO/CYC. We show that this ARO/CYC has an interior pocket capable of binding a 20-carbon polyketide substrate. This provides the first structural basis to understand how the polyketide chain is folded, cyclized, and aromatized. Taken together, this evidence firmly establishes that Tcm ARO/CYC is an active participant in the actual cyclization reaction, and it identifies residues important for cyclization specificity and catalysis. These findings will enhance future endeavors into engineered biosynthesis of cyclic polyketide natural products.

Results and Discussion

Overall Structure: Helix–Grip Fold and Hotdog Fold. The 1.9 Å crystal structure of Tcm ARO/CYC was solved by MAD. It possesses a helix–grip fold (7) consisting of a seven-stranded antiparallel
β-sheet, a long C-terminal α-helix that runs down the center of the beta-sandwich, and two small helices between β1 and β2 forming a helix–loop–helix motif that seals one end of the β-sandwich (Fig. 2A). The topology of Tcm ARO/CYC is remarkably similar to members of the Bet v1-like superfamily (7, 8), such as the birch pollen allergen Bet v1 and the MLN64-START domain (Fig. 2B) despite their low sequence homology (11–14% identity) to Tcm ARO/CYC. Importantly, and similar to other Bet v1-like superfamily members, Tcm ARO/CYC has a deep interior pocket formed between the β-sheet interior and C-terminal α-helix. Interestingly, the Tcm ARO/CYC structure looks similar to the “hotdog-in-a-bun” fold (9), whose members include the fatty acid dehydratase (DH) and thioesterase II (TE II). It is likely that the topology shared by the helix–grip fold and hotdog fold is a conserved feature of PKS and FAS. However, detailed structural comparison of Tcm ARO/CYC versus FabZ [the Pseudomonas aeruginosa DH domain (10)] reveals three major differences [supporting information (SI) Fig. S1A]: (i) the β-sheet in FabZ, with strand order 1, 2, 3, 7, 6, 5, 4 and directions +−−−−++, is markedly different from Tcm ARO/CYC.
The central helix of Tcm ARO/CYC is offset from strand 1, compared with FabZ, thus allowing interior pocket formation (3). FabZ does not have the conserved Tcm ARO/CYC α1 and α2 that are important for interior pocket formation. This interior pocket has been proposed for all Bet v1-like superfamily members to bind small molecules, such as phytosteroid hormones, lipids, enediyne, and cholesterol (7, 8). Consistent with this role, we propose that the interior pocket of Tcm ARO/CYC may bind an ACP-bound polyketide intermediate preassembled by the upstream minimal PKS.

**Interior Pocket.** We show the ARO/CYC domain to have an interior pocket, which strongly suggests that Tcm ARO/CYC can bind nascent polyketide intermediates in the pocket and influence their regiospecificity during polyketide cyclization and aromatization. The pocket dimensions (20 Å deep, cross section 14 Å × 10 Å) can easily accommodate an ACP-bound decaketide (20 carbons), such as intermediates 8-12 (Fig. 1B). The distribution of hydrophobic and polar residues in the pocket creates an ideal environment for substrate binding. The sequence alignment reveals that pocket residues Trp-28, Phe-32, Trp-65, Ser-67, Arg-69, Met-91, and Trp-95 are highly conserved among monodomain and the N-terminal half of didomain ARO/CYCs (Fig. S2 and Fig. S3A). Tcm ARO/CYC also includes unique pocket residues, such as: Thr-54, Leu-93, His-128, Thr-132, Thr-133, and Asn-136. Similar to the Bet v1-like proteins, residue variability in the ARO/CYC interior pocket will change the pocket shape and lead to distinctly different substrate specificity observed in different ARO/CYCs. The pocket also has a narrow “neck” region defined by conserved Trp-65, Phe-88, Leu-129, Tyr-35, Thr-132, Thr-133, Glu-34, and Arg-82; its narrow dimensions could limit the positional and conformational freedom of a bound polyketide intermediate (Fig. 3 and Fig. S3A).

Therefore, the size and shape of the interior pocket, and the distribution of the hydrophobic and polar/charged residues throughout the pocket (Fig. S3A) is consistent with its proposed function in polyketide binding and catalysis.

By analogy to fatty acid synthase (FAS), it has been suggested that PKS ARO/CYCs may act as DHs to catalyze aromatic ring formation via multiple dehydration reactions. Past studies on DH domains, such as FabA and FabZ (10), show that the conserved His and Asp serve as the active site acid and base for dehydration. The Tcm ARO/CYC pocket does contain His-128, Asp-57, and Glu-34. However, His-128 is not conserved in PKS ARO/CYCs. Repeated attempts to incubate known DH substrates with Tcm ARO/CYC in vitro also did not result in dehydration. Docking studies (discussed below) also indicate that His-128 and Asp-57 are not appropriately positioned in the pocket to catalyze first- or second-ring dehydrations. Other charged pocket residues, such as Tyr-35, Arg-69, Glu-34, and Arg 82 are highly conserved and may serve as the active site acid/base. Therefore, if Tcm ARO/CYC catalyzes dehydrations, the residues involved are distinctly different from the fatty acid DH.

**Substrate Docking Simulation Elucidates Possible Role of Pocket Residues.** To explore substrate binding in the interior pocket, we docked all possible putative substrates (polyketide moieties) or products (Fig. 1, products 2-3 and intermediates 8-12) into the pocket, using GOLD (11) in 10 independent runs. We found that the linear intermediate 8 has many possible docking motifs in the interior pocket, consistent with the flexibility of the linear polyketide chain. In comparison, polyketides with one or more rings that impose torsional and steric constraints are much more consistently docked into the Tcm ARO/CYC pocket with the following findings (Fig. 3A and Fig. S3B):
1. First-ring cyclization. If the polyketide is cyclized at C9-C14 as in 9-12 (Fig. 1B), the C11- and C13-carbonyl or hydroxyl groups will be anchored by hydrogen bonds to the side chains of highly conserved Ser-67 and Arg-69 at the pocket bottom. This is similar to the Ser and Asn proposed to bind the C6- and C8-OH groups of scytalone (20, Fig. S4) in scytalone dehydratase studies (12). This conformation also docks C1-C8 and C15-C20 of the polyketide chain in the pocket and positions the C1 thiol at the pocket entrance, where ACP would be covalently linked to C1 via a thioester bond. 9-12 also contact many pocket residues (Phe-32, Tyr-35, Ser-67, Arg-69, Met-91, and Trp-95) strictly conserved among mono- and di-domain ARO/CYCs (Fig. S2). Therefore, the interior pocket can effectively accommodate a C9-C14 first-ring cyclized intermediate. In comparison, if the polyketide is cyclized at C7-C12 (15, Fig. 1C), the C9- and C11-OH can still form hydrogen bonds with Ser-67 and Arg-69, but the now-longer C13-C20 fragment creates steric clashes with pocket residues (Fig. 3B). The unfavorable docking of 15 supports experimental observations that Tcm ARO/CYC promotes chain folding and C9-C14 (rather than C7-C12) first-ring cyclization. ACP may deliver intermediate 8 to the Tcm ARO/CYC pocket, where Ser-67 and Arg-69 are well positioned to anchor the polyketide, and the size and shape of the pocket can promote chain folding for C9-C14 first-ring cyclization and subsequent aromatization.

2. Second-ring cyclization and aromatization. The C9-C14 first-ring cyclization and residues in the narrow neck region (Fig. 3) place C7 and C16 in close proximity, favoring C7-C16 second-ring cyclization. Specifically, the C7 carboxyl could hydrogen-bond to Arg-82, and the C16 methylene is near Glu-34 and Tyr-35 (Fig. S3B, intermediate 10). Similar to the mechanism of the fourth-ring cyclases SnoaL (13) and AKnH (14), abstraction of the C16 proton by Glu-34 or Tyr-35 can lead to intramolecular aldol condensation that cyclizes the second ring, where the enolate intermediate can be stabilized by charge delocalization (Fig. 4).

3. Subsequent cyclizations. Docking shows that the Tcm ARO/CYC pocket can accommodate up to three linearly fused rings but not the fourth, because the intramolecular C19-C2 aldol condensation (3 to 2) (Fig. 1A) results in clashes with pocket residues of helix 3 (Fig. S3B). Therefore, after third ring cyclization (likely spontaneous), the polyketide must be transferred to TcmI (6), where fourth-ring cyclization occurs to produce TcmF1 (2).

Note that the above discussion applies to the highly conserved mono-domain ARO/CYCs that promote C9-C14 first-ring cyclization (Fig. S2B). In the case of didomain ARO/CYCs, the first-ring cyclization has been proposed to occur before ARO/CYC action, either in upstream KS/CLF or KR domains (15). Therefore, in the di-domain ARO/CYCs, the pocket residues likely orient the intermediate for first-ring aromatization but do not direct the first-ring cyclization.

Mutations of Pocket Residues Change the Cyclization Outcome. Based on the crystal structure and docking results, point mutations were generated for Glu-34, Tyr-35, Arg-69, Trp-63, Trp-108, Asn-136, and Gln-110. Multiple attempts to generate Ser-67 and Arg-82 mutants were unsuccessful. Tcm MinPKS with or without the WT or mutant Tcm ARO/CYC was expressed in a Streptomyces host. As a positive control, the plasmid pBR3 (expressing the Tcm MinPKS with WT Tcm ARO/CYC) is used to represent catalytically active Tcm ARO/CYC that produces the 20-carbon C9-C14 cyclized products 4 and 5 (Fig. 1A). As a negative control, the plasmid pSEK33 (expressing the Tcm MinPKS alone) was used, which produces predominantly the 20-carbon C7-C12 cyclized product 6 (>80%) and minor quantities of 7 (<20%) (Fig. 1A). The polyketides produced in each case were purified and analyzed by HPLC, and product profile comparison led to the identification of three major classes of pocket mutants (Fig. 5 and Table S2).

1. Significantly, R69A, R69D, and Y35A abolished the production of the C9-C14 cyclized product RM80 and instead produced the C7-C12 product 6 (>80%). Docking simulations show that the highly conserved residues R69 and Y35, anchor polyketides in the correct orientation for C9-C14 cyclization. Therefore, the loss of these anchoring points through mutations would abolish the C9-C14 cyclization; rather, the product in the absence of Tcm ARO/CYC, SEK15 (6), is obtained. To eliminate the possibility that the loss of RM80 production is due to protein conformational change, we solved the R69A crystal structure and show that the R69A structure possesses native conformation and that the mutation does not change side chain packing in the pocket (Fig. S7). This control experiment also supports the dual role of Arg-69 and Tyr-35 as the anchoring points and catalytic residues that promote cyclization (detailed in Fig. 4).

2. Strongly diminished RM80 production (ratio of [4 + 5]/6 < 0.5), represented by Y35T, W108A, W108L, and Q110H. The rescued yet largely diminished activity of Y35T suggests that Y35 serves both as an anchor point and a shape-defining residue. The residual activity of Y35T may implicate the hydroxyl functionality at this position for substrate orienting and/or enolate stabilization, with the large reduction in
activity resulting from the short length and different conformation of the Thr side chain compared with Tyr-35. The detrimental effect of the Trp-108 mutants suggest that the bulky and aromatic nature of Trp-108 is important for the appropriate folding, cyclization, and aromatization of the polyketide intermediates; the mutation would also disrupt the π-stacking interaction with Trp-95, which likely contributes to the conformational and structural stability of the monodomain ARO/CYCs. Finally, Gln-110 could be important for hydrogen bonding to intermediates 8-12 or for positioning active site water(s) for catalysis (Fig. 4), and its mutation may cause a misalignment of the hydrogen bonding network.

3. Decreased RM80 production (ratio of [4 + 5]/[6] > 1), represented by E34A, E34Q, W63N, and N136A. These results show that Glu-34, Trp-63, and Asn-136 are not essential for catalysis or substrate docking.

The above analyses show that mutations of Tcm ARO/CYC pocket residues can significantly alter polyketide product distribution and strongly suggest that the polyketide chain binds to the interior pocket, where Arg-69 and Tyr-35 orient the polyketide chain, along with other pocket-defining residues such as Trp-108 and Gln-110, so that C9-C14 first-ring and C7-C16 second-ring cyclizations occur.

**Biological Significance and Proposed Mechanism.** The aromatic rings are crucial to the bioactivities of many aromatic polyketides (16). For example, the conjugated aromatic ring system in the anthracycline polyketides (such as doxorubicin) is essential in intercalating at the CpG steps of the DNA helix, resulting in the inhibition of DNA replication and transcription. Similarly, the flat fused-ring system of the polyketide tetracycline is crucial to the antibiotic activity of tetracycline by binding to the A-site of the ribosomal 30S subunit via stacking with RNA bases and orienting the hydrophilic functional groups of tetracycline for interaction with the RNA sugar phosphate backbone. Despite the importance of the polyketide aromatic rings, little is known about how the cyclizations are precisely controlled. Although the inclusion of Tcm ARO/CYC promotes C9-C14 cyclized products in vitro and in vivo (3, 4), studies on other ARO/CYC domains have been inconclusive (1): It is never clear whether the polyketide chain is ever bound by ARO/CYC or whether this domain is only an auxiliary protein that promotes protein–protein interactions.

In this work, the 1.9 Å crystal structure of Tcm ARO/CYC reveals that the ARO/CYC has an interior pocket capable of binding a 20-carbon polyketide substrate. Further, docking simulations of polyketide substrates (Fig. 3 and Fig. S3) in this interior pocket help rationalize why Tcm ARO/CYC promotes C9-C14 but not C7-C12 first-ring cyclization: When the polyketide chain is anchored in the pocket via hydrogen-bonding with Ser-67, Arg-69, and Tyr-35, C7-C12 is not preferred because of steric clashes of the polyketide chain with the neck region of the interior pocket (Fig. 3B). Therefore, we hypothesized that both the anchoring residues and pocket-forming residues are important for the cyclization specificity. Confirming this hypothesis, we found that mutations of Arg-69 and Tyr-35 abolish the C9-C14 cyclization activity, whereas mutations of other pocket-defining residues, such as Trp-108 and Gln-110, largely diminished the C9-C14 activity. As a control experiment, the R69A crystal structure also confirmed that mutations did not change the backbone or side chain packing inside the pocket. The above results offer a strong support that the polyketide chain enters the ARO/CYC interior pocket and that the pocket residues actively participate in the cyclization events.

Based on the above results, combined with extensive docking simulations (Fig. 3 and Fig. S3B), we propose the following catalytic mechanism of Tcm ARO/CYC (Fig. 4; step-by-step in SI Text) in which the Tyr-35 and Arg-69 serve as the active site acid/base to promote first-ring cyclization and aromatization, whereas Tyr-35, Arg-82, and an active site water serve as the active site acid/base for the second-ring cyclization and aromatization. Tyr-35, Ser-67, and Arg-69 also serve as the polyketide anchors, which, combined with pocket-defining residues at the neck region (such as Trp-108 and Gln-110) (Fig. 3A), help fold the polyketide chain that results in C9-C14 first-ring and C7-C16 second-ring cyclization. In summary, the results presented here provide evidence that the polyketide cyclization/aromatization events can happen inside the ARO/CYC interior pocket. The chemical insight gleaned from this work is a step toward defining the molecular rules for the ARO/CYC cyclization specificity, whose rational control will be an important parameter for engineered polyketide biosynthesis.
Materials and Methods

Materials. pET28a (+), Escherichia coli strain BL21 (DE3), and restriction grade thrombin were purchased from Novagen. Restriction endonucleases, T4 DNA ligase, and custom oligonucleotides were purchased from New England Biolabs, Promega, and Operon Biotechnologies, respectively.

Gene Expression and Protein Purification. The gene encoding Tcm ARO/CYC was cloned into pET-28a by restriction digest of the plasmid pRZ106 (4) with NdeI/EcoRI followed by ligation. The resulting plasmid, pBA101, was then transformed into E. coli strain BL21 (DE3) for protein expression. Selenomethionine-substituted protein was produced by using metabolic inhibition of the methionine pathway in M9 minimal medium with selenomethionine as described in ref. 28. Harvested cells were suspended in lysis buffer [50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 20 mM imidazole, and 10% glycerol], sonicated, and clarified. This was followed by Ni-IMAC purification and removal of the His-tag, using thrombin. Cleaved protein was dialyzed against SP buffer [50 mM malonic acid (pH 5.0)/NaOH, 2 mM EDTA, and 10% glycerol] and loaded onto an Hi-Trap SP column (Amersham). A linear gradient from 0 to 0.2 M NaCl in SP buffer was run spanning 100 ml, and the cleaved Tcm ARO/CYC was eluted at 0.1 M NaCl. The protein was then passed through Superdex 200 10/300 GL column (Amersham) in 10 mM Hepes (pH 7.0)/NaOH to afford >99% pure protein.

Site-Directed Mutagenesis. Site-directed mutagenesis was performed by using the QuikChange site-directed mutagenesis kit (Stratagene) with pBA101 as the template. The mutagenic oligonucleotides are listed in SI Materials and Methods.

Crystallization and Data Collection. The native, mutant, and selenomethionine-substituted Tcm ARO/CYC were crystallized by the sitting drop vapor diffusion method at 4°C. Two microliters of 5 mg/ml protein were mixed with 2 μl of well solution [0.1 M NaAcO (pH 4.0–5.0) and 20–26% polyethylene glycol 8K] and equilibrated over a 500-μl well (form 1 crystals). Crystals were cryopreserved in well solution with 20% glycerol, then flash-frozen in liquid nitrogen. Both the native and multiwavelength anomalous dispersion (MAD) data were collected at the Advanced Light Source (Berkeley, CA). All data were processed using Denzo/Scalepack (17), and the statistics are listed in Table S3. Crystallization and Data Collection.

The native, mutant, and selenomethionine-substituted Tcm ARO/CYC were crystallized by the sitting drop vapor diffusion method at 4°C. Two microliters of 5 mg/ml protein were mixed with 2 μl of well solution [0.1 M NaAcO (pH 4.0–5.0) and 20–26% polyethylene glycol 8K] and equilibrated over a 500-μl well (form 1 crystals). Crystals were cryopreserved in well solution with 20% glycerol, then flash-frozen in liquid nitrogen. Both the native and multiwavelength anomalous dispersion (MAD) data were collected at the Advanced Light Source (Berkeley, CA). All data were processed using Denzo/Scalepack (17), and the statistics are listed in Table S3. Crystallization and Data Collection.

Phasing, Model Building, and Refinement. Solve (18) was used to find five of the six possible selenium sites for MAD phasing to 3.0 Å, with a mean figure of 37%. Iterative rounds of model building and refinement were continued until the Rmerge reached 26% (Rfree = 29%). Waters were then added automatically, using CNS to appropriate peaks at >3.0σ in Fo–Fc maps, followed by manual inspection to validate the water positions. Final refinement gave an Rmerge of 22% and an Rfree of 25% to 1.9 Å. See Table S3 for data refinement statistics.

Small-Molecule Docking. The program GOLD (11) was used for docking between WT Tcm ARO/CYC and the polyketide-SH portion of intermediates 8-12, 15, Tcm F2 (3), or Tcm F2 (2) (Fig. 1). The protein input file was prepared by removing waters and adding hydrogens, then saved as mol2 file format; ligands were generated using eFigit (19) and Tcm F2 were generated using ChemOffice (Cambridge soft) and saved as mol2. The binding pocket was defined as all residues within 12 Å of atom 285 (phenolic oxygen of Tyr-35), and docking was performed by using the default settings for the GA parameters with 10–25 docking trials performed for each ligand.

In Vivo Assay. Shuttle vectors for the heterologous expression of the MinPKS (tcm K5/CLF and act ACP) and WT or mutant Tcm ARO/CYC were derived from pSEK33 (4) and used to transform Streptomyces coelicolor strain CH999 or Streptomyces lividans K4-114 for polyketide biosynthesis. The cloning procedures that result in the final plasmid construct pBR3, containing both MinPKS and Tcm ARO/CYC, are detailed in SI Materials and Methods. See Table S4 for a description of all PKS plasmids used. Plasmids were passed through E. coli K12 ER2925 (New England Biolabs) to generate unmethylated DNA before their transformation into CH999. Growth on R5 agar, extraction, NMR, and HPLC product analysis follows procedures published in ref. 4. Electrospray ionization mass spectroscopy in negative ion mode (ESI−) was used to confirm the product mass. Details regarding the extraction, purification, ESI characterization procedures, and minor product identifications are discussed in SI Text.

ACKNOWLEDGMENTS. We thank M. Cocco, J. Nowick, and O. Khakshoor for their assistance with HPLC purification; J. Greaves at the University of California, Irvine Mass Spectrometry Facility; and P. Dennison at the University of California, Irvine NMR facility for their invaluable assistance in compounds characterization. S. coelicolor CH999, pRZ106, pSEK33, and purified SEK4 were kindly provided by C. Khosla. S. lividans K4-114 was generously provided by Kosan Biosciences. Diffraction data were collected at the Advanced Light Source and Stanford Synchrotron Radiation Laboratory. This work was supported by the Pew Foundation and National Institutes of Health Grant R01GM076330.