BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

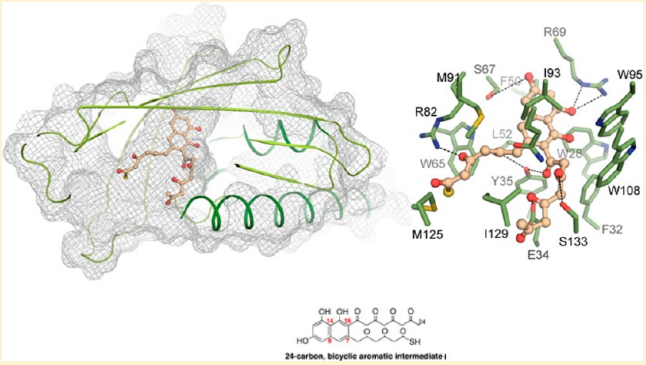
Insight into the Molecular Basis of Aromatic Polyketide Cyclization: Crystal Structure and in Vitro Characterization of WhiE-ORFVI

Ming-Yue Lee,[†] Brian D. Ames,[†] and Shiou-Chuan Tsai*[,†,‡,#]

[†]Department of Molecular Biology and Biochemistry, [‡]Department of Chemistry, and [#]Department of Pharmaceutical Sciences, University of California, Irvine, California 92697, United States

Supporting Information

ABSTRACT: Aromatic polyketides are biologically active natural products. Many important pharmaceuticals are derived from aromatic polyketides. Especially important in aromatic polyketide biosynthesis is the regiospecific cyclization of a linear, preassembled polyketide chain catalyzed by aromatase/cyclase (ARO/CYC), which serves as a key control point in aromatic ring formation. How different ARO/CYCs promote different cyclization patterns is not well understood. The *whiE* locus of *Streptomyces coelicolor* A3(2) is responsible for the biosynthesis of an aromatic polyketide precursor to the gray spore pigment. The WhiE ARO/CYC catalyzes the regiospecific C9−C14 and C7−C16 cyclization and aromatization of a 24-carbon polyketide chain. WhiE ARO/CYC shares a high degree of similarity to another nonreducing PKS ARO/CYC, TcmN ARO/CYC. This paper presents the apo crystal structure of WhiE ARO/CYC, and cocrystal structures of WhiE and TcmN ARO/CYCs bound with polycyclic aromatic compounds that mimic the respective ARO/CYC products. Site-directed mutagenesis coupled with in vitro PKS reconstitution assays was used to characterize the interior pocket residues of WhiE ARO/CYC. The results confirmed that the interior pocket of ARO/CYCs is a critical determinant of polyketide cyclization specificity. A unified ARO/CYC-mediated cyclization mechanism is proposed on the basis of these structural and functional results.



INTRODUCTION

The Gram-positive soil bacteria of genus *Streptomyces* utilize the dissociated type II polyketide synthases (PKSs) to produce aromatic polyketides.[1] Type II PKSs comprised a core set of enzymes termed the "minimal PKS" (MinPKS): the ketosynthase-chain length factor (KS-CLF) heterodimer is responsible for iterative decarboxylative condensations to generate a poly-β-ketone chain, whereas the acyl carrier protein (ACP)[2,3] shuttles both malonyl-CoA units to KS-CLF and subsequent polyketide intermediates to chain-modifying enzymes. The highly reactive linear poly-β-ketone intermediate is then modified by the ketoreductase (KR) and/or aromatase/cyclase (ARO/CYC) to the eventual formation of the aromatic core. Although the aromatic core is critical for the bioactivities of the aromatic polyketides, little is known regarding the molecular mechanisms controlling aromatic ring formation. The crystal structures of TcmN ARO/CYC from the tetracenomycin (tcm) C pathway of *Streptomyces glaucescens*[4] and ZhuI ARO/CYC from *Streptomyces* sp. R1128[5] have recently been published and allow insight into the ARO/CYC mechanism. The structure and functional analysis of TcmN ARO/CYC has provided the first insight suggesting that the interior pocket of ARO/CYC plays a direct role in binding, folding, and directing the regiospecific cyclization of an unreduced polyketide chain. Subsequently, the ZhuI ARO/CYC crystal structure provided rationalization of first-ring cyclization specificity and support for the role of the interior pocket in catalysis. However, much is still unknown about ARO/CYC, in particular, if the size of the interior pocket also controls the chain length of a given polyketide substrate.

The *whiE* gene cluster in *Streptomyces coelicolor* A3(2) is required for the developmentally regulated biosynthesis of a yet uncharacterized polyketide-derived gray spore pigment.[6,7] The *whiE* gene cluster contains genes homologous to other type II spore pigment PKSs[8] such as *sch* from *Streptomyces halstedii*[9] and *cur* from *Streptomyces curacoi*.[10] *whiE* genes also share homology with type II PKSs responsible for aromatic polyketide antibiotic production such as tetracenomycin C,[11] granaticin,[12] and actinorhodin.[13] Heterologous expression and genetic studies involving the *whiE* minimal PKS in the presence or absence of *whiE*-ORFVI (WhiE ARO/CYC) has identified the putative ARO/CYC substrate as a 24-carbon unreduced polyketide chain (Figure 1).[14,15] Expression of the WhiE MinPKS (KS-CLF and ACP) alone in vivo results in the production of over 30 polyketides, with 8 structurally characterized products exhibiting variable chain lengths (14,

Received: November 12, 2011
Revised: March 11, 2012
Published: March 20, 2012

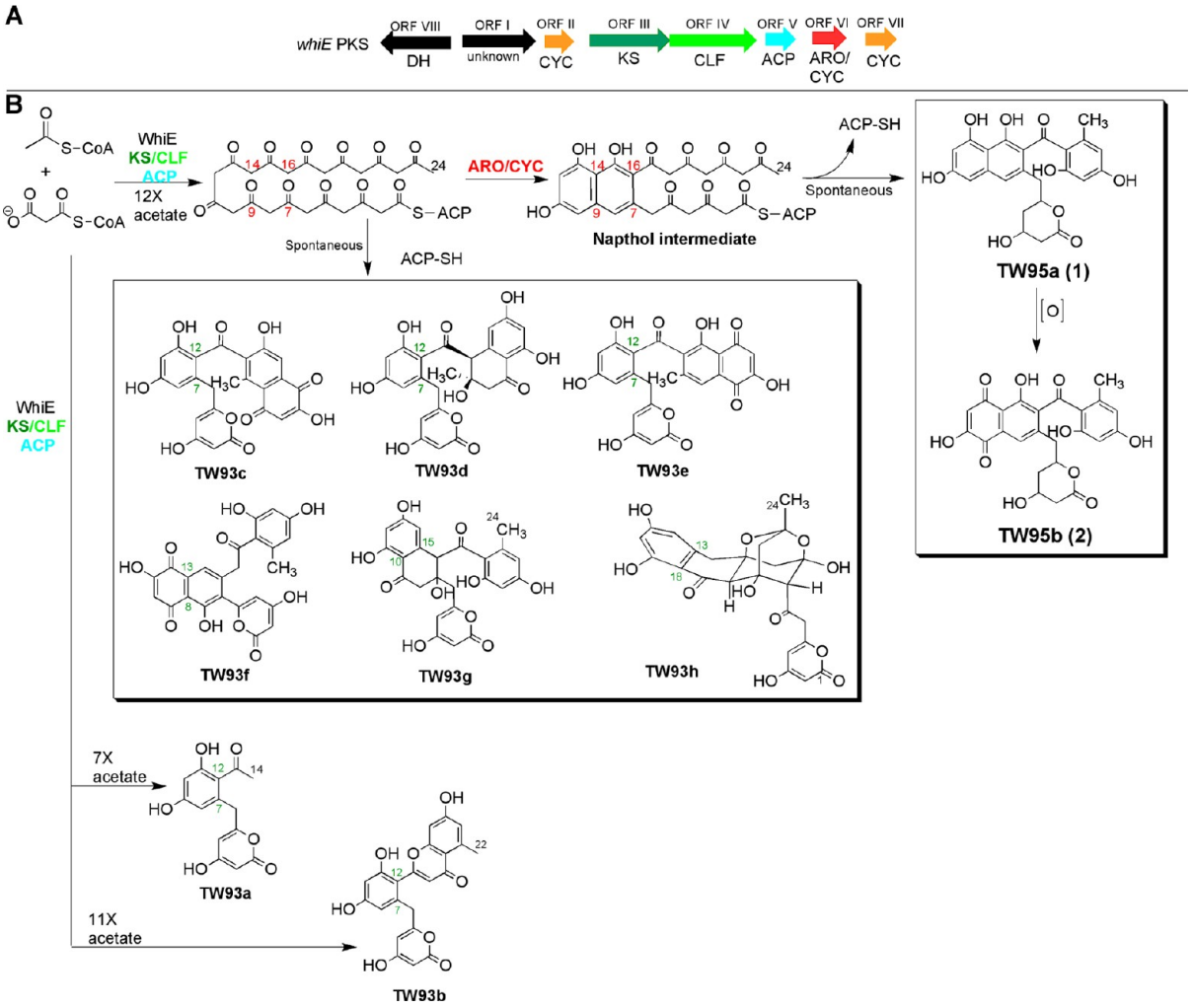


Figure 1. Proposed WhiE PKS biosynthetic pathway from genetic recombination studies:[7,14,15] (A) cartoon representation of the *whiE* gene cluster with each open-reading frame (ORF) annotated with the putative or proposed function; (B) proposed WhiE PKS biosynthetic pathway from genetic recombination studies showing the products in the presence or absence of WhiE ARO/CYC. In the presence of WhiE ARO/CYC there was exclusive production of a 24-carbon polyketide product with C9−C14 and C7−C16 cyclizations (TW95a (**1**) and TW95b (**2**)). When WhiE ARO/CYC is absent, there are many shunt products of various chain lengths and cyclization patterns.

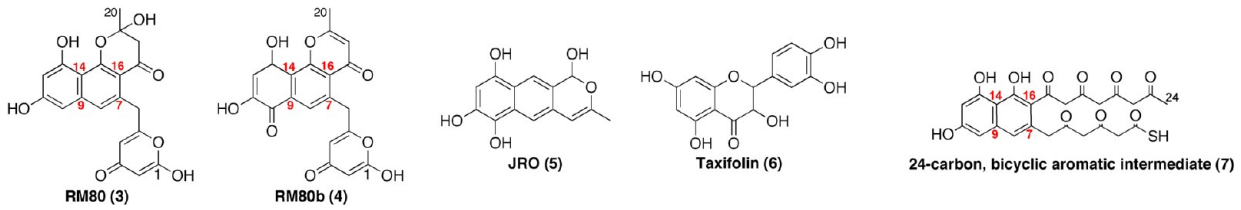


Figure 2. Additional polyketide compounds discussed in this study. RM80 (**3**) and RM80b (**4**) are both aromatic polyketides from engineered biosynthesis. JRO (**5**) and taxifolin (**6**) are both cocrystallization ligands. **7** is used for docking simulations for WhiE ARO/CYC.

22, or 24 carbons) and aberrant first-ring cyclization patterns (C7−C12, C8−C13, C10−C15, and C13−C18).[14] The addition of WhiE ARO/CYC led to the exclusive biosynthesis of the 24-carbon naphthyl dodecaketide TW95a (**1** in Figure 1) as a result of C9−C14 first-ring and C7−C16 second-ring cyclizations.[15] These studies demonstrated the ability of WhiE ARO/CYC to promote chain length specificity, as well as efficiently suppressing spontaneous aldol chemistry to direct the specific folding and cyclization of the polyketide.

WhiE ARO/CYC produces a C9−C14, C7−C16 cyclized polyketide intermediate, similar to that of TcmN ARO/CYC. However, an important distinction is that the endogenous WhiE ARO/CYC substrate is a 24-carbon, rather than a 20-carbon (*tcm*-encoded), polyketide chain. Engineered biosynthesis involving different combinations of PKS subunits has shown that WhiE ARO/CYC can functionally replace TcmN ARO/CYC, because expression of the Tcm MinPKS with WhiE[16] or TcmN ARO/CYC[17,18] yields the same 20-carbon, C9−C14 first-ring (and C7−C16 second-ring) cyclized polyketides, RM80 (**3**) and RM80b (**4**) (Figure 2). Amino acid sequence comparison reveals that WhiE ARO/CYC shares an overall identity/similarity of 54/65% to TcmN ARO/CYC with 80/93% identity/similarity for interior pocket-defining residues (Figure 3). Because the interior pocket of ARO/CYC

dx.doi.org/10.1021/bi201705q | *Biochemistry* 2012, 51, 3079−3091

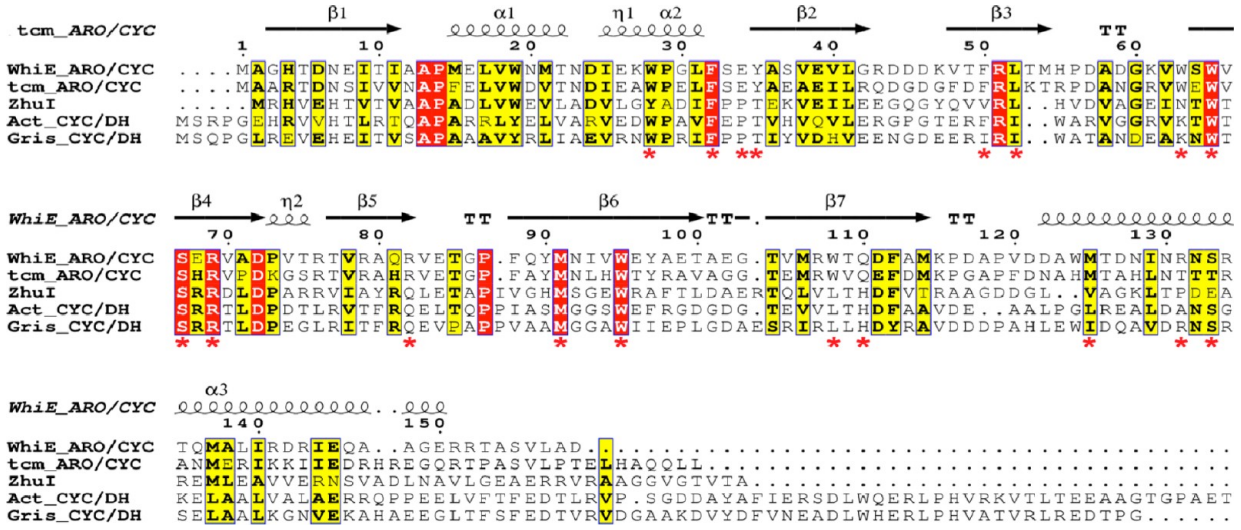


Figure 3. Sequence alignment of WhiE, Tcm, ZhuI, Act, and Gris ARO/CYCs rendered using ESPript (Risler similarity matrix with global score set to 0.7). The top line illustrates secondary structure elements according to the WhiE ARO/CYC structure. The asterisks found below the aligment are used to indicate those residues that define the interior pocket.

has been shown to be a determinant of cyclization specificity,[18,19] the high similarity observed between the pocket-defining residues of WhiE and TcmN ARO/CYCs follows logically in that both enzymes promote the same first- and second-ring cyclization specificity, and the only difference is the chain lengths of the respective natural polyketide substrates (24 and 20 carbons, respectively). Structural insight into WhiE ARO/CYC pocket composition and conformation will therefore aid in understanding how WhiE ARO/CYC accommodates a 24-carbon polyketide intermediate (4 carbons longer than that naturally acted on by TcmN ARO/CYC) to effect similar chain folding and cyclization patterns. However, in the past, such a comparative effort has been hampered by the lack of cocrystal structures of ARO/CYC bound with substrate or product analogues.

We present the first report of an ARO/CYC crystal structure with a bound aromatic polyketide analogue solved to 1.8 Å along with the 1.8 Å native crystal structure. Additionally, we present the results of polyketide docking simulations and in vitro functional characterization of WhiE ARO/CYC. The cocrystal structure of WhiE ARO/CYC with a bound synthetic aromatic polyketide analogue is critically compared to a cocrystal structure of TcmN ARO/CYC bound to the plant polyketide taxifolin; together, these cocrystal structures enable us to highlight key features of the interior pocket in its capacity to bind multicyclic aromatic compounds. Taken together, this work provides insight into the ability of ARO/CYC to accommodate a variety of polyketide chain lengths and enables us to propose a generalized mechanism of ARO/CYC-mediated polyketide cyclizations/aromatizations, in which the interior pocket is an active participant in binding, folding, and cyclization of an unreduced polyketide

MATERIALS AND METHODS

Construction of pWhiE. The plasmid pYT198, a pCR blunt plasmid (Invitrogen) construct containing the *whiE* ORFVI (*whiE ARO/CYC*) gene, was kindly provided by Prof. Yi Tang. All cloning steps were performed in *Escherichia coli* strain Nova-Blue (Novagen). The *whiE ARO/CYC* gene was amplified via PCR using Pfu Turbo DNA polymerase (Stratagene) along with the following primers (underlined portions denote noncomplementary region; restriction sites, in bold, are inserted to the 5′ portion of the primer sequence):

forward: GCGCTC**CATATG**GCAGGGCACACCGAC

reverse: GCGC**GAATTC**TTAGTCGGCGAGCACC-GAC

The amplified gene product was double-digested with *Eco*RI and *Nde*I, gel-purified, and ligated onto pET28a vector (Novagen). DNA sequencing ([UC]Sequencing Facility, UC Davis) was used to confirm the final construct.

WhiE ARO/CYC Protein Overproduction and Purification Protocol. *E. coli* strain BL21 (DE3) transformed with pWhiE was cultured in 2 × 1 L LB containing kanamycin (50 μg/mL) at 37 °C until A_{600} reached ∼0.6; the temperature was then reduced to 18 °C, and 0.1 mM IPTG was added to induce expression. After approximately 18 h of additional incubation, the cells were harvested at 4000 rpm for 20 min, yielding approximately 5 g of wet cell mass per liter of culture, and the pellets were stored at −80 °C.

The cell pellets containing expressed protein were resuspended in lysis buffer (50 mM Tris·HCl, pH 7.5, 300 mM NaCl, 20 mM imidazole, and 10% glycerol), sonicated, and clarified (21000 rpm for 60 min). The clarified lysate was mixed with Ni-IMAC resin (Bio-Rad Laboratories), and batch-bound at 4 °C for 1 h. The mixture was then applied to a gravity-flow column, washed once with 25 mL of wash buffer 1 (lysis buffer + 20 mM imidazole), and again with 25 mL of wash buffer 2 (lysis buffer + 30 mM imidazole). The protein was then eluted with six 5 mL fractions of lysis buffer containing increasing concentrations of imidazole (50, 75, 100, 150, 250, and 500 mM). The wash and elution fractions were visualized by SDS-PAGE, with ≥90% pure protein observed at 30 mM imidazole fraction and onward. Eluted protein fractions (30−250 mM imidazole) were pooled, concentrated to 5−10 mg/mL using an iCON 9000 molecular weight cutoff centrifugal filter device (Pierce), and buffer-exchanged into 20 mM HEPES, pH 7.0, using a PD-10 desalting column (GE Healthcare). The protein concentration was determined using the Bradford method with BSA as a standard[20] (Bio-Rad Laboratories).

dx.doi.org/10.1021/bi201705q | *Biochemistry* 2012, 51, 3079−3091

Site-Directed Mutagenesis. The QuikChange II Site-Directed Mutagenesis Kit (Stratagene) was used to introduce WhiE ARO/CYC pocket-residue mutations, which were confirmed through automated DNA sequencing. The primers used for mutagenesis are as follows (mutated nucleotides are in bold):

F32A: GTGGCCCGGCCTG**GCG**AGCGAGTACGCC
F32Y: GGCCCGGCCTGT**AT**AGCGAGTACGCC
E34A: GCCTGTTCAGCG**C**GTACGCCTCCGTG
Y35A: GCCTGTTCAGCGAG**GCG**GCCTCCGTG-GAGG
Y35F: CCTGTTCAGCGAGT**TT**GCCTCCGTG-GAGGTG
Y35T: CTGTTCAGCGAG**AC**CGCCTCCGTGG
R69A: GGGTCTCCGAG**GC**GGTCGCCGACCC
R69K: GCTGGGTCTCCGAG**AAA**GTCGCC-GACCCCGTC
R69Q: GGGTCTCCGAGC**A**GGTCGCCGACCC
R82A: GTCCGCGCCCAG**GC**GGTCGAGACCG
R82E: CACCGTCCGCGCCCAG**GAA**GTCGA-GACCGG
R82K: CACCGTCCGCGCCCAG**AAA**GTCGA-GACCGG
R82Q: GTCCGCGCCCAGC**A**GGTCGAGACCGG

WhiE ARO/CYC Crystallization. Two microliters of 5 mg/mL WT protein was mixed with 2 μL of well solution (0.1 M Bis-Tris, pH 5.0−6.0, 0.2 M $MgCl_2 \cdot 6H_2O$, 24−30% PEG3350) and equilibrated over a well volume of 500 μL at 4 °C using the sitting drop vapor diffusion method. Clustered crystals would grow within 1 week (space group $P2(1)$, with one molecule per asymmetric unit).

For cocrystallization, 5 mM JRO (100 mM stock solution dissolved in DMSO; Prof. Townsend, Johns Hopkins University) was mixed with 5 mg/mL WT WhiE ARO/CYC. The mixture was then passed through a 0.2 μm centrifugal filter device (Pall Life Sciences) to remove precipitate due to the low solubility of JRO in aqueous solutions. Two microliters of this mixture was then mixed with 2 μL of well solution (0.1 M Bis-Tris, pH 5.0−6.0, 0.2 M $MgCl_2 \cdot 6H_2O$, 28−34% PEG3350) and equilibrated over a well volume of 500 μL at 4 °C using the hanging drop vapor diffusion method. Colored crystals appeared within 1 week (space group $P2(1)$, with two molecules per asymmetric unit). The unit cell dimensions of either crystal used for structural determination are shown in Table 1. The crystals were cryopreserved in well solution containing 20% glycerol and then flash frozen in liquid nitrogen.

Cocrystallization of TcmN ARO/CYC with Taxifolin. The expression and purification of TcmN ARO/CYC (for crystal structure solution with bound taxifolin) was performed as previously described.[18] Crystals of TcmN ARO/CYC were grown as previously described.[18] Soaks with the flavonoid polyketide taxifolin (*trans*-dihydroquercetin) were performed with (±)-taxifolin hydrate (Sigma); a stock solution of 45 mM taxifolin was made in well solution (20% PEG 8000, 0.1 M NaOAc, pH 5.0) plus 10% DMSO; crystals were transferred to a 10 μL drop containing 5 mM taxifolin in well solution and allowed to stand for 1 h at 4 °C and then transferred to cryoprotectant solution (well solution plus 5 mM taxifolin and 20% glycerol) prior to being flash frozen in liquid nitrogen.

Diffraction Data Collection. Monochromatic X-ray diffraction data for native WhiE ARO/CYC were collected to 1.8 Å on beamline 9-2 of the Stanford Synchrotron Radiation Laboratory (SSRL). For WhiE ARO/CYC with bound JRO, diffraction data to 1.8 Å resolution were collected at beamline 8.2.2 of the Advanced Light Source (ALS). Diffraction data for TcmN ARO/CYC with bound taxifolin were collected at ALS beamline 8.3.1. Single crystals were used for all data sets used for structure solution, and all data were processed using HKL2000.[21] Table 1 lists the data collection statistics.

Table 1. Data Collection and Refinement Statistics

	WhiE-ORFVI		TcmN ARO/CYC
	native	(+) JRO	(+) taxifolin
crystallographic data			
wavelength (Å)	1.0000	0.9998	1.0332
resolution (Å)	50−1.8	50−1.8	50−1.7
space group	$P2_1$	$P2_1$	$C222_1$
a, b, c (Å)	47.8, 42.1, 48.0	64.4, 42.1, 69.3	72.5, 105.0, 51.1
α, β, γ (deg)	90, 91.9, 90	90, 102.0, 90	90, 90, 90
total observations	65592	125588	157778
unique reflections	17725	31213	22974
high-resolution shell (Å)	1.84−1.80	1.85−1.80	1.73−1.67
average redundancy[a]	3.7 (3.4)	4.0 (3.0)	6.9 (5.6)
completeness (%)	99.2 (93.6)	98.9 (92.2)	99.9 (100)
mean $I\sigma/(I)$	18.3 (6.1)	14.9 (2.8)	46.9 (5.0)
R_{sym} (%)	7.6 (19.1)	8.5 (33.8)	3.8 (30.9)
refinement			
resolution (Å)	48−1.80	50−1.80	32−1.67
no. of reflections	17705	28074	22482
no. of protein atoms	1376	2536	1245
no. of water atoms	199	217	154
no. of heteroatoms	6	51	22
R_{crys}, R_{free} (%)	18.7, 21.4	19.0, 25.4	20.8, 23.3
geometry			
rmsd bonds (Å)	0.013	0.015	0.005
rmsd angles (deg)	1.987	1.587	1.20
average *B* factors (Å^2)	27.0	28.7	24.3
Ramachandran (%)			
most favored	98.2	94.4	94.8
additional allowed	1.8	5.6	5.2
generously allowed	0	0	0

[a]Values in parentheses represent the highest resolution shell.

Phasing, Model Building, and Refinement. The initial phasing for WhiE ARO/CYC was performed by molecular replacement using CNS[22] using a polyalanine version of the TcmN ARO/CYC structure[18] as the search model. Rigid-body refinement using CNS was applied to the molecular replacement solution followed by manual model building of main-chain and side-chain atoms in COOT.[23] Iterative rounds of model building and refinement (combined with density modification) in CNS were performed until an R_{crys} of 27.5% and an R_{free} of 29.4% were reached. At this point, crystallographic waters and a bound glycerol molecule were added to the model with COOT, and the data were transferred for refinement in REFMAC (CCP4)[24] with the final few rounds of refinement done using Phenix.refine[25,26] in Phenix. Further building, improvement, and refinement of the model gave a final R_{crys} of 18.7% and an R_{free} of 21.4%, with the model

dx.doi.org/10.1021/bi201705q | *Biochemistry* 2012, 51, 3079−3091

Table 2. In Vitro Assay Results for WhiE ARO/CYC Mutants

	products isolated (% peak area)				products isolated (% peak area)			
PKS4 minPKS + WhiE ARO/CYC	PK8 (**8**)	naphthopyrone (**9**)	ratio of **8**:**9**	Act MinPKS + WhiE ARO/CYC	SEK4 (**10**)	SEK4b (**11**)	UWM1 (**12**)	RM77 (**13**)
	52.2	47.8	1.1: 1		26.6	71.8		
WT	87.4	12.6	6.9: 1	WT	11.4	44.6	15.3	28.7
R82A	54.1	45.9	1.2: 1	R82A	32.6	74.7		
R82E	66.2	33.8	2.0: 1	R82E	28.9	71.1		
R82Q	53.1	46.9	1.1: 1	R82Q	24.3	73.8		
Y35A	58.8	41.2	1.4: 1	R82K	24.4	71.6		
Y35T	50.5	49.	1.0: 1	Y35A	28.8	70.9		
Y35F	79.9	20.1	4.0: 1	Y35F	10.8	45.6	23.8	19.7
R69A	54.0	46.0	1.2: 1	F32Y/Y35F	20.5	67.9	3.4	8.2
R69Q	55.4	44.6	1.2: 1	R69A	26.4	72.9		
R69K	87.8	12.2	7.2: 1	R69Q	26.5	72.9		
F32A	61.4	38.6	1.6: 1	R69K	10.3	39.3	25.8	24.6
E34A	81.5	18.5	4.4: 1	TcmN ARO/CYC	11.5	29.4	29.3	29.8
TcmN ARO/CYC	75.9	24.1	3.1: 1	Zhul ARO/CYC	68.4	30.7		

consisting of residues 1−159 (with an additional 14 residues at the N-terminal end present as cloning artifacts), 199 waters, and 3 molecules of glycerol.

The initial phases for WhiE ARO/CYC + JRO cocrystal structure were determined by molecular replacement using Phaser (CCP4)[27] and the refined WhiE ARO/CYC structure as the search model. In contrast to the native structure, there are two WhiE molecules present in the asymmetric unit. Following one round of refinement in REFMAC, inspection of the $2F_o - F_c$ (contoured to 1.0σ), and $F_o - F_c$ (at 3.0σ) electron density maps revealed clear and complete density in the pocket defining the ligand JRO. The coordinate and topology files of JRO were prepared using the Dundee PRODRG server,[28] and automatic ligand fitting in COOT was used to place the compound into the experimental density. Iterative rounds of model building in COOT and refinement in REFMAC were used to complete the structure solution

The TcmN ARO/CYC−taxifolin cocrystal structure was solved by molecular replacement in CNS using the WT TcmN ARO/CYC structure as the search model. Following one round of combined refinement in CNS, there was well-defined and complete electron density defining the bound ligand taxifolin. The ligand coordinates were generated using the PRODRG server and the molecule placed into the corresponding electron density using Quanta (Accelrys Software Inc.). Further rounds of model building and improvement in Quanta and refinement in CNS produced the final model (R_{crys} of 20.8% and an R_{free} of 23.3%). Data refinement statistics for the WhiE and TcmN ARO/CYC structures described herein are listed in Table 1.

ActKS/CLF Expression and Purification. The expression of *act*KS/CLF from *S. coelicolor* CH999/pRJC006 has been described previously.[29] Briefly, spores were grown in 50 mL Super YEME containing 50 μg/mL kanamycin for 3 days at 30 °C with shaking at 250 rpm. The mycelia were then transferred to 500 mL of Super YEME containing 50 μg/mL kanamycin and grown as before for 2 days. Protein expression was induced by the addition of 5 μg/mL thiostrepton and cell growth continued as before for 1 more day. Cells were harvested by centrifugation (5000 rpm × 30 min), resuspended in 40 mL of lysis buffer (100 mM KP_i, pH 7.5, 0.1% Triton X-100, 5 mM TCEP, 1.5 mM benzamidine, 1 tablet EDTA-free protease inhibitor cocktail (Roche), and 10% glycerol), and lysed on ice by sonication (8 × 1 min pulses). Cell debris was removed by centrifugation (21000 rpm × 30 min) followed by binding of lysate supernatant to 3 mL of Ni-IMAC resin (Bio-Rad) in batch mode by spinning at 4 °C for 2 h. Protein was eluted with increasing concentrations of imidazole in 100 mM KP_i, pH 7.5, 500 mM NaCl, and 10% glycerol. Fractions containing *act*KS/CLF were collected at 150 and 500 mM imidazole, pooled, and buffer-exchanged to 100 mM KP_i (pH 7.2) and 20% glycerol.

Holo-FrenN and Holo-Act ACP and MAT Expression and Purification. *E. coli* BAP1 cells[30] expressing pET21c-*fren*ACP or pET28a-*act*ACP were grown at 37 °C in 1 L of LB medium supplemented with 100 μg/mL ampicillin to an OD_{600} of 0.5. At that point, protein expression was induced by the addition of 1 mM IPTG at 18 °C overnight. The initial purification of holo-ACP was conducted as described previously.[31] Following Phenyl-Sepharose chromatography, ACP was buffer-exchanged to buffer A (50 mM Tris-Cl, pH 7.5, 2 mM DTT, 2 mM EDTA, and 10% glycerol). Protein was applied to a HiTrap Q FF column (GE Healthcare) and eluted with a linear gradient from 0 to 100% buffer B (buffer A plus 1 M NaCl). Fractions containing pure holo-ACP were pooled and concentrated. *S. coelicolor* MAT was expressed and purified from *E. coli* BL21(DE3)/pGFL16 by Ni-IMAC as described previously.[32]

PKS4 minPKS Protein Expression and Purification. The expression and purification of the PKS4 KS-AT didomain construct and the standalone PKS4 ACP (together constituting the PKS4 "MinPKS" for use in the in vitro reconstitution assay) was performed as previously described.[33]

In Vitro Reconstitution Assay. Ten micromolar PKS4 KS-MAT or *act*KS-CLF, 50 μM *holo*-ACP, 5 mM malonyl-CoA, and 50 μM WhiE ARO/CYC (WT or mutant) were mixed together in a 250 μL reaction tube. An additional component required for the Act MinPKS reconstitution assay is 1 μM MAT (*S. coelicolor*). The reaction buffer used was 0.1 M Tris·HCl, pH 7.0. The mixture was incubated at room temperature overnight and extracted with ethyl acetate:methanol:formic acid (95:4.9:0.1). The organic phase of the extraction was then dried via a SpeedVac vacuum concentrator (Thermo Scientific) and reconstituted in 100% DMSO. The extracts were then run through analytical reverse-phase HPLC using a Synergi Hydro-RP C_{18} column (Phenomenex), 150 × 4.60 mm. The mobile phases for RP-HPLC: buffer A, water containing 0.1% formic acid; buffer B, acetonitrile containing 0.1% formic acid. The

dx.doi.org/10.1021/bi201705q | *Biochemistry* 2012, 51, 3079−3091

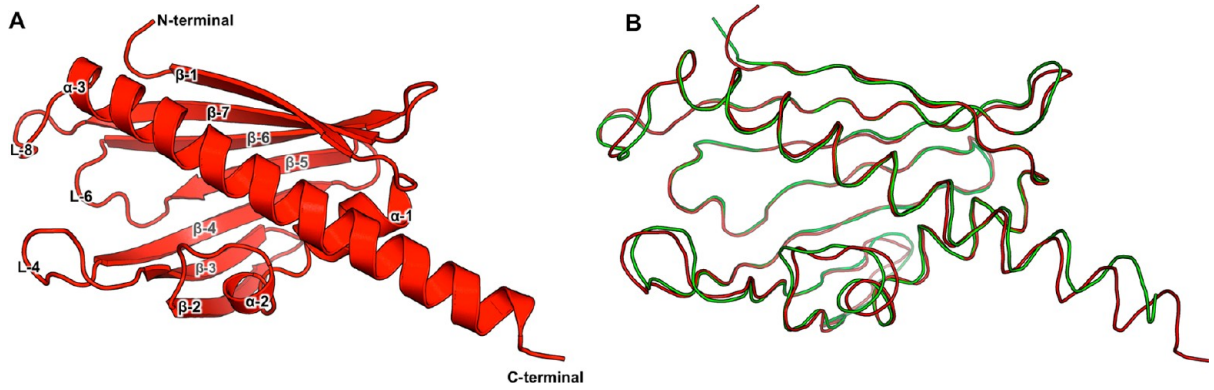


Figure 4. Overall WhiE ARO/CYC structure characterized by a helix-grip fold (A); structural alignment of WhiE ARO/CYC (red) to TcmN ARO/CYC (green) (B) to show that, overall, the topology of these two enzymes is highly similar.

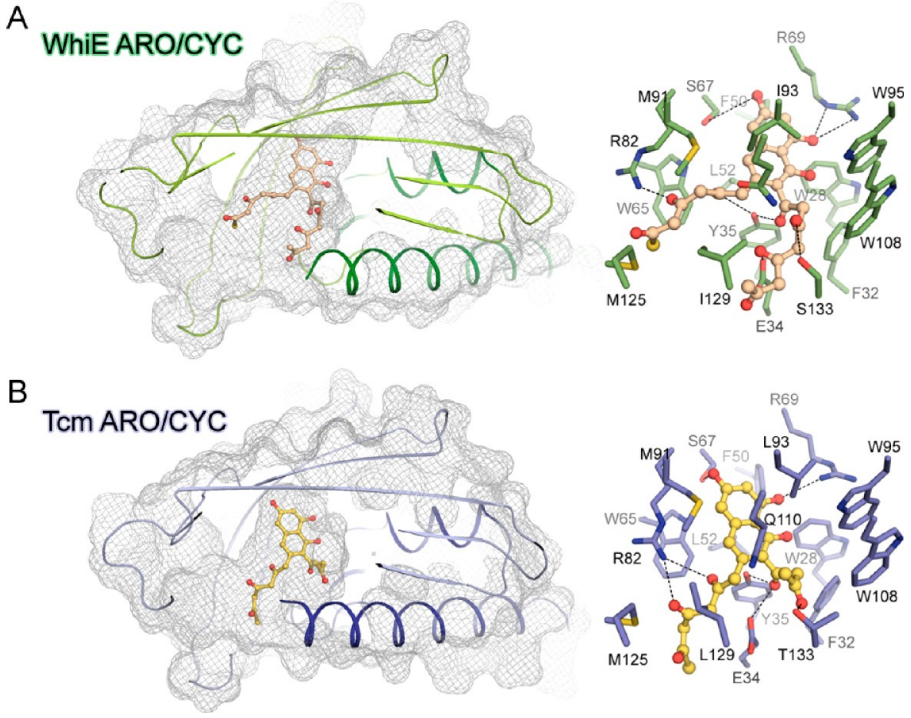


Figure 5. Docking simulations of polyketide substrates into the interior pocket of WhiE and TcmN ARO/CYCs: (A) WhiE ARO/CYC docked with the 24-carbon bicyclic aromatic intermediate 7 (Figure 2); (B) TcmN ARO/CYC docked with a 20-carbon bicyclic aromatic intermediate (for both panels, (left) cutaway mesh surface view showing the positions of the docked molecules; (right) view of pocket residue interactions with the docked polyketides).

gradient for the HPLC run consisted of the following steps: equilibration (0–5 min, 5% B); linear gradient (5–15 min, 5–50% B); linear gradient (15–20 min, 50–95% B); wash (20–25 min, 95% B); linear gradient (25–28 min, 95–5% B); and re-equilibration (3 min, 5% B). Peak integration ratios of the HPLC chromatograms are listed in Table 2.

Small-Molecule Docking. Docking simulations between WhiE ARO/CYC and the polyketide intermediates were performed using GOLD.[34] The WhiE ARO/CYC structure was defined as static, with the binding pocket being defined as 15 Å around the side-chain hydroxyl of tyrosine 35, atom 384 in the mol2 file. The default settings were for the genetic algorithm (GA) parameters, and 20 docking trials were performed each run.

RESULTS AND DISCUSSION

Overall Fold: WhiE and TcmN ARO/CYCs Are Highly Similar. The crystal structure of apo WhiE ARO/CYC was solved to 1.8 Å using molecular replacement with TcmN ARO/CYC (PDB code 2RER[18]) as the search model. The overall structure is characterized by a helix-grip fold[35,36] consisting of a seven-stranded antiparallel β-sheet flanking a long C-terminal α-helix and two small helices to one side of the protein (Figure 4A). Structural superimposition between WhiE and TcmN ARO/CYC reveals that both structures are highly similar with an rmsd of 0.75 Å for residues 1–151 (Figure 4B). The only notable difference in backbone structure is the C-terminal helix (α-3), where this region extends for an additional 1.5 helical turns at the C-terminal end for WhiE ARO/CYC but for TcmN ARO/CYC the equivalent region is random coil. The ordering of this region in the WhiE ARO/CYC structure may be due to

dx.doi.org/10.1021/bi201705q | *Biochemistry* 2012, 51, 3079–3091

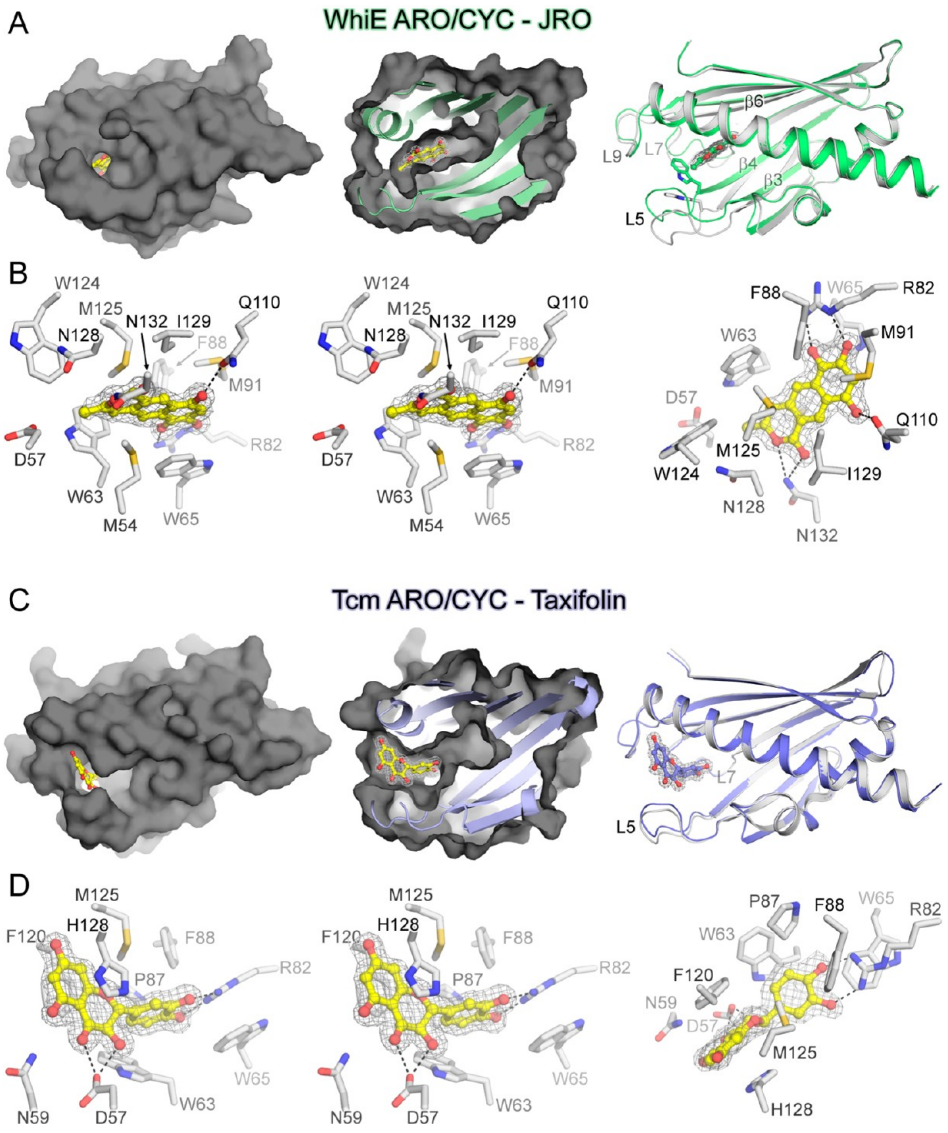


Figure 6. Comparison of the cocrystal structures WhiE ARO/CYC with bound JRO and TcmN ARO/CYC with bound taxifolin. (A) Left, molecular surface view of WhiE ARO/CYC showing the bound ligand as yellow sticks; middle, a cutaway view of the surface reveals the position of the JRO in the interior pocket; right, overlay of apo (gray) and ligand-bound (green) forms of WhiE ARO/CYC illustrating the conformational changes that occur to accommodate ligand binding (the residue shown is W63). The gray mesh surrounding the ligand is from the 1.8 Å experimental $2F_o - F_c$ electron density map contoured to 1.5σ. (B) Left, stereoview of ligand-binding interactions (potential hydrogen bonding is indicated by a dashed line); right, an alternate view of the same. Panels C and D illustrate the cocrystal structure of taxifolin bound to TcmN ARO/CYC (cartoon representation of ligand-bound structure colored blue) and are represented as described for panels A and B above. The electron density map contoured around taxifolin is from the 1.7 Å $2F_o - F_c$ electron density map contoured to 1.0σ.

crystal packing, because the C-terminal end of α-3 contacts adjacent protein molecules in the crystal lattice. Similar to TcmN ARO/CYC, WhiE ARO/CYC crystallized as a monomer. The similarity between the overall fold of WhiE and TcmN ARO/CYCs offers an excellent opportunity to critically compare the molecular features of their corresponding interior pockets.

Cocrystal Structures Highlight Important Features in the Interior Pocket. An important structural element of the helix-grip fold is the presence of an interior pocket formed between the β-sheets and C-terminal helix. The lipid/sterol-binding StAR-related lipid transfer (START) protein domains[37,38] and the PR-10 family of plant pathogenesis-related proteins[36,39] have both demonstrated the ability of the interior pocket to mediate polycyclic ligand binding. Related to this function of binding multicyclic ligands, the interior pocket of ARO/CYC is predicted to bind the putative linear polyketide substrate, as well as ARO/CYC-mediated cyclization products.[18] The pocket-defining residues of WhiE ARO/CYC include hydrophobic (3 Phe, 4 Trp, 3 Met, and 8 with aliphatic side chains), polar (2 Asn, 2 Gln, 1 Tyr, and 1 Ser), and charged (2 Arg, 1 Asp, and 1 Glu) residues (Figure 3). The amphipathic composition of residues would provide appropriate hydrophobic and hydrogen bond interaction for binding both the linear and cyclized polyketide intermediates. The pocket residues between WhiE and TcmN ARO/CYC are highly conserved both at the sequence level (Figure 3) and in three-dimensional structures (Figure 5; Supporting Information, Figure S1).

dx.doi.org/10.1021/bi201705q | *Biochemistry* 2012, 51, 3079–3091

The polyketide binding potential of the ARO/CYC interior pocket was confirmed experimentally by two cocrystal structures: WhiE ARO/CYC with bound JRO (**5**) (Figure 2, a synthetic tricyclic aromatic polyketide mimic) and TcmN ARO/CYC with bound taxifolin (**6**) (Figure 2, an aromatic plant polyketide). These two aromatic polyketide analogues were found after extensive searching and crystallization screening of aromatic polyketides with WhiE and TcmN ARO/CYCs; only two cocrystals afforded well-defined electron density of the bound polyketide analogues (Figure 6). The WhiE ARO/CYC cocrystal structure with bound JRO was solved to 1.8 Å with clear electron density of JRO bound approximately halfway into the pocket of WhiE ARO/CYC (Figure 6A). Upon JRO binding, a conformational change in the N-terminal region of β-strands 3, 4, and 6, and loops 5, 7, and 9 (and associated residue side chains) is observed (Figure 6A). There are substantial interactions between the active-site residue side chains and JRO, including (Figure 6B) (1) edge-face and offset-stacked $\pi-\pi$ interactions mediated by three Trp (W63, W65, W124) and one Phe (F88) (of particular note is the dramatic shift observed for W63, which moves ~4 Å with ~50° rotation of the indole ring to set up perpendicular $\pi-\pi$ stacking interactions with the JRO aromatic rings); (2) hydrophobic interaction above or below the plane of the ligand ring system by three Met (M54, M91, M125) and one Ile (I129), with additional interactions provided by the β-CH_2 side chains of D57 and N128; (3) hydrogen bond interactions provided by R82, Q110, and N132. The nature of the interactions described is relevant to understanding how an aromatic polyketide intermediate, such as the bicyclic product of WhiE ARO/CYC-mediated cyclizations (**1**, Figure 1), may bind to the pocket of WhiE ARO/CYC. Because JRO is a tricyclic aromatic compound that mimics the ARO/CYC product (**5**, Figure 2), the fact that it is not within interaction distance from the conserved and functionally essential pocket arginine (R69, Figure 3) may reflect the product-exit mode, as opposed to the substrate-binding mode. Note that the JRO–WhiE complex is a snapshot of binding, which may represent the product exiting the active site or a nonproductive mode. Nevertheless, to the best of our knowledge, this is the first product analogue-bound ARO/CYC structure that helps visualize how a polyketide product may interact with the ARO/CYC pocket residues.

Similarly, the TcmN ARO/CYC–taxifolin cocrystal structure shows taxifolin bound at the entrance to the interior pocket, where minimal protein conformation changes occur at R82, H128, F120, and loops 4 and 6 to accommodate taxifolin binding (Figure 6C). Similar to the binding of JRO to WhiE ARO/CYC, there are substantial aromatic $\pi-\pi$ interactions observed between taxifolin and two Trp (W63 and W65), two Phe (F88 and F120), and a His (H128) residue of TcmN ARO/CYC. There are also hydrophobic interactions with aliphatic side chains of M125 and P87 and four potential hydrogen bond interactions provided by R82 and D57 (Figure 6D). Similar to JRO, taxifolin binds the entrance of the pocket and is distant from conserved and functionally critical residues such as Y35 and R69. Specifically, the 2-(3,4-dihydroxyphenyl) substituent forms a hydrogen bond with R82 and may preclude the ligand from fitting deeper into the pocket. Interestingly, attempts to crystallize WhiE ARO/CYC with taxifolin were unsuccessful, suggesting that small changes to pocket residue composition can affect ligand-binding potential. In addition, although the location and binding modes of ligands (JRO to WhiE ARO/CYC and taxifolin to TcmN ARO/CYC) are different due to their distinct chemical structures, they share many common residues that may govern putative substrate-binding and product-exit modes, including D57, W65, R82, F88, and M125. Of these residues, W65 is highly conserved among mono- and didomain ARO/CYC, and R82 is found as either Arg or Gln, whereas the hydrophobic character of F88 and M125 is conserved (Figure 3). In summary, the two cocrystal structures show that the interior pocket of ARO/CYC can bind multicyclic molecules that mimic natural aromatic polyketide products, and they highlight key residues that may interact directly with the putative substrate and product.

Docking Simulations Visualize Chain Length Control of ARO/CYCs. To visualize how the putative bicyclic aromatic product of WhiE ARO/CYC (**7**, Figure 2) may bind to the interior pocket, and for comparison to the experimentally observed cocrystal structures, we performed docking simulations utilizing the program GOLD[40] (Figure 6A). Over 100 independent docking rounds, the bicyclic intermediate **7** was consistently docked in a wide U-shaped geometry. A cutaway view of a mesh surface representation of WhiE ARO/CYC reveals that its interior pocket (Figure 5A) is larger than the pocket of TcmN ARO/CYC (Figure 5B). Therefore, WhiE ARO/CYC is capable of accepting a 24-carbon polyketide intermediate and contains sufficient interior pocket space to mediate the C9–C14 first-ring and C7–C16 second-ring cyclizations.

To generate the larger interior pocket to accommodate a C24 (vs C20) polyketide substrate, WhiE ARO/CYC undergoes several subtle changes in pocket residue composition and conformation compared to TcmN ARO/CYC (Figure 6). There are seven changes to the residues that define the WhiE ARO/CYC pocket, as compared to those of TcmN ARO/CYC (TcmN → WhiE ARO/CYC residue: T54 → M54, L93 → I93, H128 → N128, T132 → N132, T133 → S133, and N136 → Q136) (Figure 3; Supporting Information, Figure S1). As noted from this list, there are three pocket-defining residues in WhiE ARO/CYC (T54, T132, and N136) with side chains larger than the corresponding residues in TcmN ARO/CYC (M54, N132, and Q136). Therefore, pocket residue switching alone does not explain the increased pocket size of WhiE ARO/CYC. Rather, the larger pocket for WhiE ARO/CYC can be explained by a combination of several factors: (1) the switch of H128 and T133 in TcmN ARO/CYC to the smaller N128 and S133, respectively, in WhiE ARO/CYC; (2) the change in conformation of the guanidinium group of R82 (flipped 180° and forming a 2.9 Å hydrogen bond with the backbone carbonyl of E84) (Figure 5A); (3) the slightly extended conformation of the loop connecting α3-β7 (Figure 4); and (4) the observed conformation of the side chains N132 and Q136 (both oriented away from the pocket interior). The extent to which protein crystallization may account for these changes is unknown; nevertheless, WhiE ARO/CYC crystal structure exemplifies that the interior pocket can accommodate a 24-carbon polyketide intermediate as compared to the pocket of TcmN ARO/CYC in binding a 20-carbon polyketide intermediate.

Pocket Residue Mutagenesis and in Vitro Activity Assay. To explore the importance of pocket residues for ARO/CYC activity, we generated the following WhiE ARO/CYC mutants: F32A, E34A, Y35A/F/T, R69A/K/Q, and R82A/G/E/K/Q. Y35 and R69 were previously proposed as the active site residues that were critical for TcmN ARO/CYC activity,

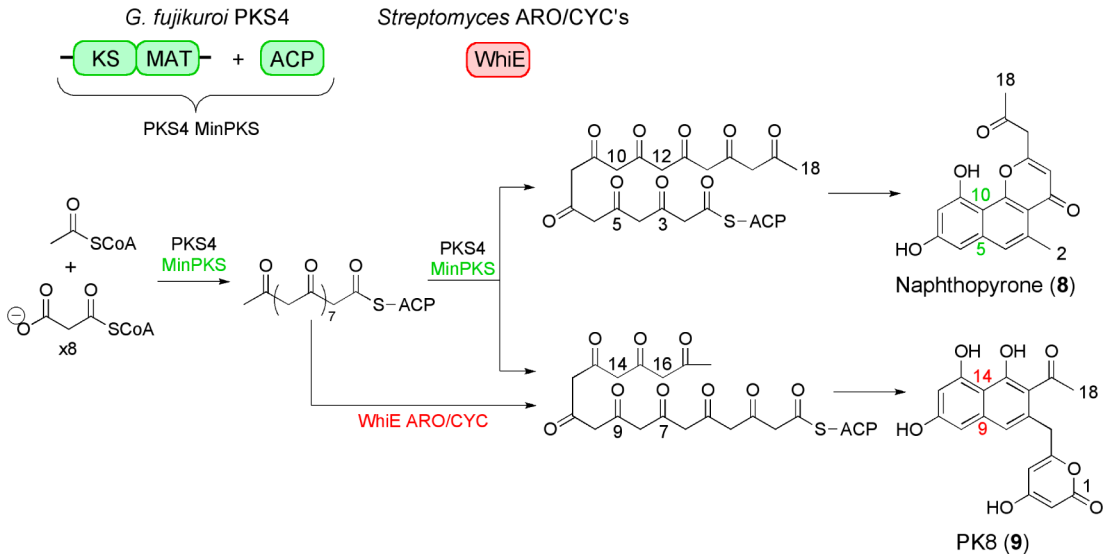


Figure 7. PKS4 minPKS-based in vitro assay for WhiE ARO/CYC mutational analyses. PKS4 minPKS alone produces two products, **8** and **9** (in ~1:1 ratio), each with distinct cyclization patterns. WhiE ARO/CYC directs the cyclizations of nascent polyketide chain down a single path for C9–C14 first-ring cyclization in the production of PK8 (**9**).

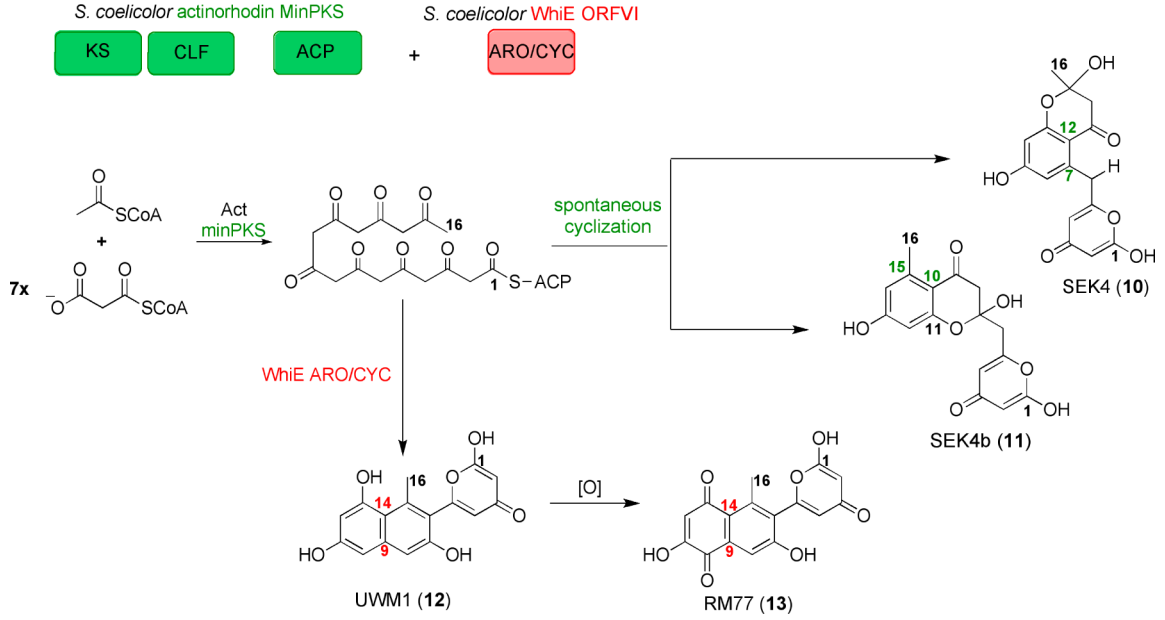


Figure 8. Actinorhodin MinPKS-based in vitro assay for WhiE ARO/CYC mutational analyses. Actinorhodin MinPKS alone produces two products, **10** and **11**, each with distinct cyclization patterns. WhiE ARO/CYC directs the cyclizations of nascent polyketide chain down a single path for C9–C14 first-ring cyclization in the production of RM77 (**13**).

and R82 had been observed to directly interact with the ligand from the two cocrystal structures (Figures 6 and 9B). The other residues were identified to be in close interaction distance from the putative ligand in our docking trials.

To assay WT and mutant WhiE ARO/CYC activity, we performed in vitro reconstitution assays utilizing either the fungal PKS4 minPKS (KS-MAT and ACP) from *Gibberella fujikuroi* (Figure 7, the PKS4 assay)[33] or the bacterial type II actinorhodin MinPKS (Act KS-CLF and ACP) from *S. coelicolor* (Figure 8, the *Act* assay).[29] Previous studies have established that the PKS4 minPKS alone produces a 1:1 mixture of **8** and **9** (Figure 7).[33] In contrast, when the C9–C14 (first ring) and C7–C16 (second ring) specific WhiE or TcmN ARO/CYCs were combined in vitro with the PKS4 minPKS, there was a dramatic increase in the production of the C9–C14 cyclized compound **9** (**9**:**8** = 10: 1).[33]

In the second assay, Act MinPKS alone produces the octaketides SEK4 **10** (C7–C12 first-ring cyclization) and SEK4b **11** (C10–C15 first-ring cyclization).[41] These in vitro results correspond well with the in vivo expression results.[41] In contrast, when WhiE or TcmN ARO/CYCs were incubated with the Act MinPKS, the resulting polyketides are the octaketide RM77 **13**[42] and its reduced product, UWM1 **12**,[43] both of which are C9–C14 first-ring cyclized products (Figure 8). ZhuI ARO/CYC was included as a control for the *Act* assay, as it catalyzes C7–C12 first-ring cyclization to produce SEK4 **10** as the major product along with the minor product, SEK4b **11**.[44] On the basis of the HPLC UV absorption, the yields are

dx.doi.org/10.1021/bi201705q | *Biochemistry* 2012, 51, 3079–3091

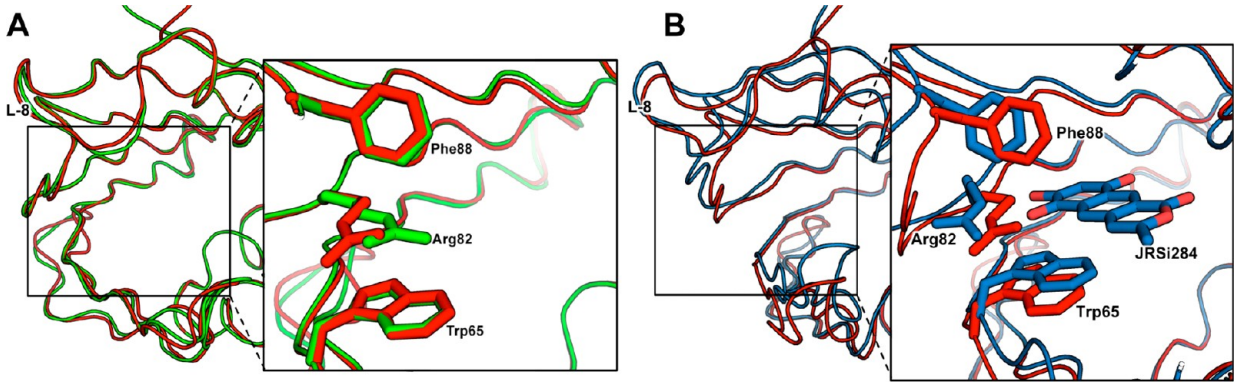


Figure 9. $\pi-\pi$ interactions of Trp65−Arg82−Phe88: (A) WhiE (red) and TcmN (green) ARO/CYC overlays showing that the positions of Trp and Phe are almost identical, whereas Arg82 of TcmN is in a different rotameric state than WhiE Arg82; (B) WhiE ARO/CYC apo (red) and JRO-bound (blue) overlays, with the ligand (JRO) showing significant movement for all three residues upon ligand binding.

similar between the wild type and mutant enzymes. Therefore, the mutation affects the product distribution (in terms of cyclization pattern) but not the catalytic efficiency. On the basis of previous studies,[45,46] the rate-limiting step for aromatic PKS should be the upstream chain elongation catalyzed by the KS domain; thus, the mutation significantly changed the cyclization outcome, but not the efficiency. Below, we present the assay results of WhiE mutants.

1. Y35 Mutants. On the basis of work with TcmN ARO/CYC,[18] the conserved Tyr35 was proposed to be the general base of ARO/CYC. To test this hypothesis, we generated Y35A, Y35T, and Y35F. Consistent with the previously proposed mechanism, Y35A and Y35T were inactive (Table 2). However, to our surprise, Y35F remains partially active (**9**:**8** = 4.0:1 for the PKS4 assay; **12** and **13** are produced to WT levels in the *Act* assay). Further examination of WhiE ARO/CYC crystal structure reveals that Y35 interacts with F32 through edge−face $\pi-\pi$ stacking. Y35F thus preserves this interaction. To probe the importance of this edge−face $\pi-\pi$ interaction, F32A and F32Y/Y35F mutants were generated. F32A, similar to Y35A and Y35T, is inactive (Table 2). F32Y/Y35F is active, albeit to a lesser degree when compared to WT (**12** and **13** are produced at approximately 4 times less than WT levels in the *Act* assay, Table 2). Taken together, these assay results show that the hydroxyl moiety at residue 35 need not be present for catalysis, but the preservation of the edge−face $\pi-\pi$ interaction between Phe32 and Tyr35 is critical for ARO/CYC activity. Therefore, the tyrosinate may not act as the sole active site base. Rather, because the crystal structure shows a series of ordered water molecules interacting with Y35 and neighboring E34, the active site base can be a polarized water molecule. The inactive E34A mutant further supports this hypothesis. On the other hand, the loss of activity of F32A, Y35A, and Y35T may be related to a change of the local protein conformation. In comparison, Y35F and F32Y/Y35F remain active because the $\pi-\pi$ interaction between Y35F and F32 is preserved; thus, the ARO/CYC pocket remains unperturbed.

2. R69 Mutants. On the basis of our TcmN ARO/CYC work, the conserved R69 was proposed to be a general acid important for first-ring cyclization by ARO/CYC.[18] To test this hypothesis, R69A, R69K, and R69Q mutants were generated. Consistent with the TcmN ARO/CYC mutation result, R69A and R69Q were both inactive (Table 2). The inactive R69Q supports the hypothesis that residue 69 needs to be more than just a polar functionality capable of anchoring the polyketide substrate through the formation of hydrogen bonds; rather, this position requires a positively charged residue side chain. The fact that R69K remains active corroborates well with this hypothesis (Table 2). In contrast, our recently published work on ZhuI ARO/CYC showed that the mutation of Arg66 (equivalent to Arg69 of WhiE and TcmN ARO/CYCs) to A, E, Q, or K abolished activity, suggesting that this guanidinium functionality is critical in controlling first ring cyclization and aromatization.[5]

The slightly enhanced polyketide production by WhiE ARO/CYC R69K could be the result of the lowered pK_a of the Lys ε-amine group compared to the Arg guanidinium moiety, thus allowing the Lys side chain to serve as a better proton donor. Furthermore, because the Lys side chain is shorter than that of Arg, Lys may be better positioned to orient and interact with carbonyl-13 for protonation to form an enol intermediate. In support of this hypothesis, docking simulation and the two cocrystal structures both support the different geometries of these two residues resulting in different positioning to interact with the substrate carbonyl groups. The residual activity was not observed in ZhuI ARO/CYC R66K mutant, as it remained inactive. Taken together, the above results suggest that the positive charge of residue 69 is critical for its role as the active site acid and aids in intermediate stabilization, as previously proposed for TcmN ARO/CYC.[18]

3. Arg82 Mutants. The conserved Arg82 was proposed previously as a proton donor that facilitates the second-ring aromatization.[18] To determine the role of R82 in catalysis, we generated R82A, R82G, R82E, R82Q, and R82K. On the basis of the above result for R69 mutants, we expect that R82K should remain active, whereas the other mutants would be inactive. Additionally, if residue 82 requires only a side chain that can anchor the polyketide substrate, then the E, Q, and K mutants should all remain active. To our surprise, all five R82 mutants are inactive (Table 2). Therefore, the presence of an Arg at residue 82 is critical for enzyme activity. The WhiE ARO/CYC crystal structure provides a structural role for Arg82: there is a $\pi-\pi$ stacking interaction between the delocalized π-electrons from the guanidinium moiety of Arg82 and the Trp65 indole ring. This is additionally stabilized by a perpendicular $\pi-\pi$ stacking interaction from the side chain of Phe88 (Figure 9). These $\pi-\pi$ stacking interactions help orient the WhiE ARO/CYC entrance loop (L-8, Figures 2 and 9) that controls ligand binding and product egress. Structural superimposition of WhiE with TcmN ARO/CYC shows that

dx.doi.org/10.1021/bi201705q | *Biochemistry* 2012, 51, 3079−3091

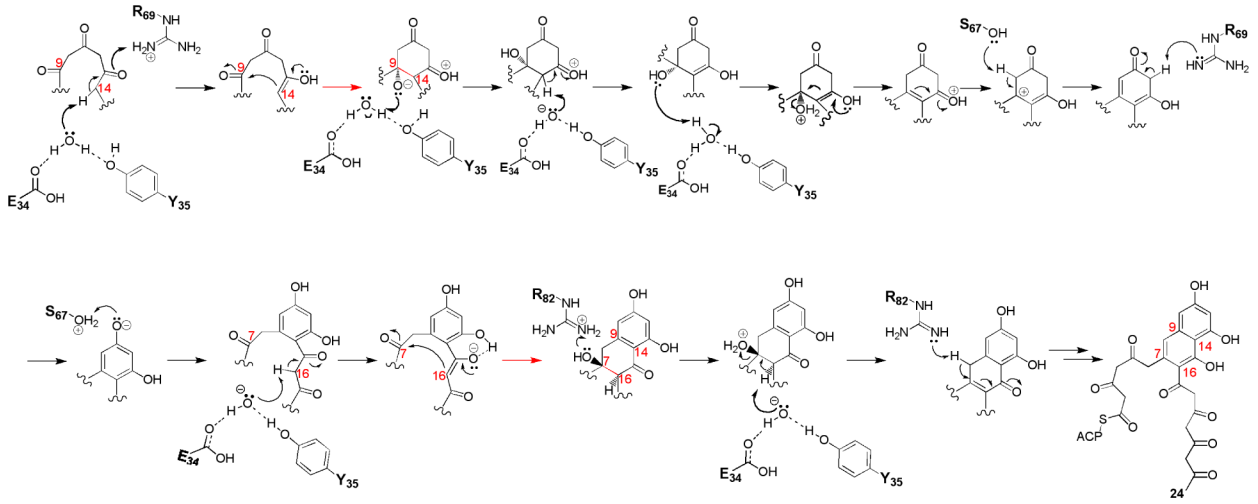


Figure 10. Revised ARO/CYC mechanism based on this work. A water molecule coordinated by Tyr35 and Glu34 acts as the active-site base to abstract a proton from carbon 14 (C14) to create the enolate at C13. The guanidinium moiety of Arg69, in close proximity to C13 oxygen, donates a proton to form the enol, which then collapses to attack the C9 carbonyl to form the first ring. Tyr35, Ser67, and Arg69 then are involved in multiple rounds of dehydrations en route to an aromatized first ring. The second ring cyclization between C16 and C7, as well as the aromatization events, follows in a similar fashion. The absence of ARO/CYC leads to compound **8** in Figure 7 and compounds **10** and **11** in Figure 8, whereas the presence of WhiE ARO/CYC promotes C9−C14 cyclization and leads to the formation of compound **9** in Figure 7 and **13** in Figure 8.

the positions of Phe88 and Trp65 overlay well, whereas Arg82 possesses different rotameric conformations (Figure 9A). Despite the different conformations of the R82 guanidinium group, this series of $\pi-\pi$ stacking interactions is preserved (Figure 9A).

In the WhiE ARO/CYC cocrystal structure, these side-chain $\pi-\pi$ stacking interactions are disturbed (Figure 9B) as a result of JRO binding. JRO binding essentially replaces Arg82 spatially, causing Trp65 to shift 1.2 Å to allow for Trp65−ligand $\pi-\pi$ stacking (Figure 9B). An additional edge−face $\pi-\pi$ interaction is observed between the ligand and Phe88 as Phe88 shifts 1.4 Å relative to the apo structure (Figure 9B). Finally, Arg82 shifts 0.9 Å away from its original position in the apo structure, but forms additional hydrogen bonds with JRO (Figures 5 and 9B). From these interactions with the product analogue JRO, we postulate that during the binding of the putative 24-carbon polyketide substrate Arg82 may modulate and facilitate the substrate binding or product release event via W65−R82−F88 interactions that orient loop L-8. The Trp−Arg−Phe stacking interactions would be disrupted and L-8 could be misoriented when Arg82 is mutated to other residues, resulting in a loss of enzyme activity as observed (Table 2). We conclude that Arg82 serves both a critical structural role in maintaining L-8 conformation and the previously proposed role in polyketide substrate stabilization of the enol form.[18,19]

Proposed Gereral Mechanism of C9−C14 ARO/CYC. On the basis of the apo WhiE ARO/CYC structures, cocrystal structures of WhiE and TcmN ARO/CYCs, docking simulations, and mutational analyses, we proposed the following general mechanism for ARO/CYC-catalyzed C9−C14 and C7−C16 cyclizations/aromatizations of type II polyketides. During our previous study of TcmN ARO/CYC, R69 and Y35 were proposed as the active site acid and base.[18] From the proposed TcmN ARO/CYC mechanism,[18] Arg69 is the active site acid, whereas the active site base can be an E34−Y35-coordinated water molecule (Figure 10). For this reason, we mutated the corresponding Arg69, Arg82, and Tyr35 of WhiE ARO/CYC (Table 2), which supported the importance of these residues for ARO/CYC functions. Therefore, the first step is the abstraction of the α-proton from carbon 14 to form the enolate of carbonyl-13. Following the generation of the negative charge on carbonyl-13 oxygen, Arg69 is oriented to donate a proton to form an enol intermediate, which then collapses for C9−C14 first-ring cyclization. The coordinated water molecule then donates a proton to the negatively charged oxygen linked to carbon 9 as a result of the ring formation to stabilize the oxyanion. Docking simulation supports the −OH of carbon-9 and the proton of carbon-14 being in the *anti* conformation, allowing for spontaneous dehydration to occur, forming a C−C double bond between C9 and C14. Alternatively, the coordinated water can abstract the proton from C14 to generate the charge, which results in a C−C double bond between C14 and C13 (Figure 10). First-ring aromatization occurs before the same set of reactions proceeds for the C7−C16 second-ring cyclization. In the second-ring cyclization step, the active site base is again a replenished, coordinated water molecule; however, the proton donor for second-ring cyclization and dehydration is hypothesized then to be the critical Arg82 residue (Figure 10).

Biological Significance. The crystal structure of WhiE ARO/CYC reveals that the overall structure and residue composition are similar to those of TcmN ARO/CYC, but subtle conformation and residue changes in the pocket of WhiE ARO/CYC effectively increase the pocket space (compared to TcmN ARO/CYC) to accommodate the binding of a 24-carbon polyketide substrate that cyclized at C9−C14 first ring and C6−C16 second ring. Findings from this work show that pocket residue variability and side-chain conformation are important for chain length determination. The two cocrystal structures with product analogues bound lay the foundation for future studies on substrate/product binding of ARO/CYC. These findings can be leveraged to understand the large diversity of polyketides with varied chain lengths and cyclization patterns that are affected by the ARO/CYCs from various type II PKS pathways. These examples include actinorhodin[13] (octaketide), tetracenomycin C[47] and daunorubicin[48] (decaketides), *whiE* spore pigment[15] and pradimicin[49]

dx.doi.org/10.1021/bi201705q | *Biochemistry* 2012, 51, 3079−3091

(dodecaketides), griseorhodin A[50] (tridecaketide), and the pentadecaketide fredericamycin[51] ARO/CYCs.

The two enzyme assays show that WhiE ARO/CYC productively interacts with both bacterial type II and fungal aromatic type I MinPKS assemblies to direct the C9−C14 first-ring and C7−C16 second-ring cyclizations. These findings are significant in utilizing ARO/CYCs in engineered biosynthesis as they show that ARO/CYCs control the initial cyclization regiospecificity. Furthermore, ARO/CYCs not only control the cyclization pattern but also control chain lengths when reconstituted with different MinPKS systems.

The WhiE ARO/CYC interior pocket residues R69 and R82 are critical for its cyclization activity. The observation that the charge property of Arg69 is critical corroborates well with the TcmN ARO/CYC mutagenesis studies.[18] Despite the larger interior pocket size and four extra carbons for its natural substrate, WhiE ARO/CYC promotes the same first- and second-ring cyclization pattern as TcmN ARO/CYC, suggesting that the residues at the entrance of the pocket may be as important as the interior pocket residues for directing cyclization events. The conserved Arg82 near the pocket entrance may mediate these interactions, as we demonstrated that the guanidinium moiety contributes critical structural and physical properties to the ARO/CYC entrance local environment. We conclude that the docking of ACP and the extent to which the polyketide chain is introduced into the ARO/CYC pocket, possibly controlled by the conserved Arg82 along with the loop L-8 local environment, are critical determinants of cyclization specificity.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Stereoview of protein models. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

The atomic coordinates have been deposited in the Protein Data Bank (accession codes 3TVR, 3TVQ, and 3TL1).

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (949) 824-4486. E-mail: sctsai@uci.edu. Fax: (949) 824-8552.

Funding

[†]S.-C.T., B.D.A., and M.-Y.L. are supported by the Pew Foundation and National Institute of General Medicinal Sciences (NIGMS R01GM076330).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Prof. Yi Tang for the generous provision of pYT198 and Pks4 min PKS constructs, Prof. Craig Townsend for the generous gift of JRO, and Prof. John Crosby for the generous gift of pRJC006. We are grateful for the assistance of Dr. John Greaves, Director of the UCI Mass Spectrometry Facility, in product characterization. Diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) and the Advanced Light Source (ALS).

■ ABBREVIATIONS USED

ARO/CYC, aromatase/cyclase; PKS, polyketide synthase; ACP, acyl carrier protein; KS, ketosynthase; CLF, chain length factor; DH, dehydratase; KR, ketorductase; PPT, phosphopantetheine; FAS, fatty acid synthase

■ REFERENCES

(1) Rawlings, B. J. (1999) Biosynthesis of polyketides (other than actinomycete macrolides). *Nat. Prod. Rep. 16*, 425−484.

(2) Byers, D. M., and Gong, H. (2007) Acyl carrier protein: structure-function relationships in a conserved multifunctional protein family. *Biochem. Cell. Biol. 85*, 649−662.

(3) Zhou, P., Florova, G., and Reynolds, K. A. (1999) Polyketide synthase acyl carrier protein (ACP) as a substrate and a catalyst for malonyl ACP biosynthesis. *Chem. Biol. 6*, 577−584.

(4) Ames, B. D., Korman, T. P., Zhang, W., Smith, P., Vu, T., Tang, Y., and Tsai, S. C. (2008) Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides. *Proc. Natl. Acad. Sci. U.S.A. 105*, 5349−5354.

(5) Ames, B. D., Lee, M. Y., Moody, C., Zhang, W., Tang, Y., and Tsai, S. C. (2011) Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway. *Biochemistry 50*, 8392−8406.

(6) Davis, N. K., and Chater, K. F. (1990) Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. *Mol. Microbiol. 4*, 1679−1691.

(7) Yu, T.-W., and Hopwood., D. A. (1995) Ectopic expression of the *Streptomyces coelicolor whiE* genes for polyketide spore pigment synthesis and their interaction with the *act* genes for actinorhodin biosynthesis. *Microbiology 141*, 2779−2791.

(8) Blanco, G., Brianb, P., Pereda, A., Méndez, C., Salas, J. A., and Chater, K. F. (1993) Hybridization and DNA sequence analyses suggest an early evolutionary divergence of related biosynthetic gene sets encoding polyketide antibiotics and spore pigments in Streptomyces spp. *Gene 130*, 107−116.

(9) Blanco, G., Pereda, A., Méndez, C., and Salas, J. (1992) Cloning and disruption of a fragment of *Streptomyces halstedii* DNA involved in the biosynthesis of a spore pigment. *Gene 112*, 59−65.

(10) Bergh, S., and Uhlén, M. (1992) Analysis of a polyketide synthesis-encoding gene cluster of *Streptomyces curacoi*. *Gene 117*, 131−136.

(11) Motamedi, H., and Hutchinson, C. R. (1987) Cloning and heterologous expression of a gene cluster for the biosynthesis of tetracenomycin C, the anthracycline antitumor antibiotic of *Streptomyces glaucescens*. *Proc. Natl. Acad. Sci. U.S.A. 84*, 4445−4449.

(12) Ichinose, K., Bedford, D. J., Tornus, D., Bechthold, A., Bibb, M. J., Peter Revill, W., Floss, H. G., and Hopwood, D. A. (1998) The granaticin biosynthetic gene cluster of *Streptomyces violaceoruber* Tü22: sequence analysis and expression in a heterologous host. *Chem. Biol. 5*, 647−659.

(13) McDaniel, R., Ebert-Khosla, S., Hopwood, D. A., and Khosla, C. (1994) Engineered biosynthesis of novel polyketides: *act*VII and *act*IV genes encode aromatase and cyclase enzymes, respectively. *J. Am. Chem. Soc. 116*, 10855−10859.

(14) Shen, Y., Yoon, P., Yu, T. W., Floss, H. G., Hopwood, D., and Moore, B. S. (1999) Ectopic expression of the minimal whiE polyketide synthase generates a library of aromatic polyketides of diverse sizes and shapes. *Proc. Natl. Acad. Sci. U.S.A. 96*, 3622−3627.

(15) Yu, T.-W., Shen, Y., McDaniel, R., Floss, H. G., Khosla, C., Hopwood, D. A., and Moore, B. S. (1998) Engineered biosynthesis of novel polyketides from *Streptomyces* spore pigment polyketide synthases. *J. Am. Chem. Soc. 120*, 7749−7759.

(16) Alvarez, M. A., Fu, H., Khosla, C., Hopwood, D. A., and Bailey, J. E. (1996) Engineered biosynthesis of novel polyketides: properties of the whiE aromatase/cyclase. *Nat. Biotechnol. 14*, 335−338.

(17) McDaniel, R., Hutchinson, C. R., and Khosla, C. (1995) Engineered biosynthesis of novel polyketides: analysis of tcmN function in tetracenomycin biosynthesis. *J. Am. Chem. Soc. 117*, 6805−6810.

(18) Ames, B. D., Korman, T. P., Zhang, W., Smith, P., Vu, T., Tang, Y., and Tsai, S.-C. (2008) Crystal structure and functional analysis of

tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides. *Proc. Natl. Acad. Sci. U.S.A. 105*, 5349−5354.
(19) Ames, B. D., Lee, M.-Y., Moody, C., Zhang, W., Tang, Y., and Tsai, S.-C. (2011) Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway. *Biochemistry 50*, 8392−8406.
(20) Bradford, M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem. 72*, 248−254.
(21) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307−326.
(22) Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D 54* (Part 5), 905−921.
(23) Emsley, P., and Cowtan, K. (2004) COOT: model-building tools for molecular graphics. *Acta Crystallogr. D 60*, 2126−2132.
(24) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D 53*, 240−255.
(25) Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P., and Adams, P. D. (2010) Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr. D 66*, 1153−1163.
(26) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D 66*, 213−221.
(27) Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., and Terwilliger, T. C. (2004) Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Radiat. 11*, 53−55.
(28) Schuttelkopf, A. W., and van Aalten, D. M. F. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D 60*, 1355−1363.
(29) Matharu, A. L., Cox, R. J., Crosby, J., Byrom, K. J., and Simpson, T. J. (1998) MCAT is not required for in vitro polyketide synthesis in a minimal actinorhodin polyketide synthase from *Streptomyces coelicolor*. *Chem. Biol. 5*, 699−711.
(30) Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., and Khosla, C. (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science 291*, 1790−1792.
(31) Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Persson, B., and Jornvall, H. (2003) Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chem.−Biol. Interact. 143−144*, 247−253.
(32) Kumar, P., Koppisch, A. T., Cane, D. E., and Khosla, C. (2003) Enhancing the modularity of the modular polyketide synthases: transacylation in modular polyketide synthases catalyzed by malonyl-CoA:ACP transacylase. *J. Am. Chem. Soc. 125*, 14307−14312.
(33) Zhang, W., Li, Y., and Tang, Y. (2008) Engineered biosynthesis of bacterial aromatic polyketides in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A. 105*, 20683−20688.
(34) Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol. 267*, 727−748.
(35) Iyer, L. M., Koonin, E. V., and Aravind, L. (2001) Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. *Proteins 43*, 134−144.
(36) Radauer, C., Lackner, P., and Breiteneder, H. (2008) The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol. Biol. 8*, 286.
(37) Tsujishita, Y., and Hurley, J. H. (2000) Structure and lipid transport mechanism of a StAR-related domain. *Nat. Struct. Biol. 7*, 408−414.
(38) Schrick, K., Nguyen, D., Karlowski, W., and Mayer, K. (2004) START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biol. 5*, R41.
(39) Liu, J.-J., and Ekramoddoullah, A. K. M. (2006) The family 10 of plant pathogenesis-related proteins: their structure, regulation, and function in response to biotic and abiotic stresses. *Physiol. Mol. Plant Pathol. 68*, 3−13.
(40) Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Improved protein-ligand docking using GOLD. *Proteins 52*, 609−623.
(41) Carreras, C. W., and Khosla, C. (1998) Purification and in vitro reconstitution of the essential protein components of an aromatic polyketide synthase. *Biochemistry 37*, 2084−2088.
(42) Kramer, P. J., Zawada, R. J. X., McDaniel, R., Hutchinson, C. R., Hopwood, D. A., and Khosla, C. (1997) Rational design and engineered biosynthesis of a novel 18-carbon aromatic polyketide. *J. Am. Chem. Soc. 119*, 635−639.
(43) Seow, K. T., Meurer, G., Gerlitz, M., Wendt-Pienkowski, E., Hutchinson, C. R., and Davies, J. (1997) A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. *J. Bacteriol. 179*, 7360−7368.
(44) Tang, Y., Lee, T. S., and Khosla, C. (2004) Engineered biosynthesis of regioselectively modified aromatic polyketides using bimodular polyketide synthases. *PLoS Biol. 2*, E31.
(45) Beltran-Alvarez, P., Arthur, C. J., Cox, R. J., Crosby, J., Crump, M. P., and Simpson, T. J. (2009) Preliminary kinetic analysis of acyl carrier protein-ketoacylsynthase interactions in the actinorhodin minimal polyketide synthase. *Mol. Biosyst. 5*, 511−518.
(46) Beltran-Alvarez, P., Cox, R. J., Crosby, J., and Simpson, T. J. (2007) Dissecting the component reactions catalyzed by the actinorhodin minimal polyketide synthase. *Biochemistry 46*, 14672−14681.
(47) Shen, B., and Hutchinson, C. R. (1996) Deciphering the mechanism for the assembly of aromatic polyketides by a bacterial polyketide synthase. *Proc. Natl. Acad. Sci. U.S.A. 93*, 6600−6604.
(48) Hutchinson, C. R. (1997) Biosynthetic studies of daunorubicin and tetracenomycin C. *Chem. Rev. 97*, 2525−2536.
(49) Kim, B. C., Lee, J. M., Ahn, J. S., and Kim, B. S. (2007) Cloning, sequencing, and characterization of the pradimicin biosynthetic gene cluster of *Actinomadura hibisca* P157−2. *J. Microbiol. Biotechnol. 17*, 830−839.
(50) Li, A., and Piel, J. (2002) A gene cluster from a marine *Streptomyces* encoding the biosynthesis of the aromatic spiroketal polyketide griseorhodin A. *Chem. Biol. 9*, 1017−1026.
(51) Wendt-Pienkowski, E., Huang, Y., Zhang, J., Li, B., Jiang, H., Kwon, H., Hutchinson, C. R., and Shen, B. (2005) Cloning, sequencing, analysis, and heterologous expression of the fredericamycin biosynthetic gene cluster from *Streptomyces griseus*. *J. Am. Chem. Soc. 127*, 16442−16452.