Biochemistry Supporting Information

STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF ZHUI AROMATASE/CYCLASE FROM THE R1128 POLYKETIDE PATHWAY

Brian D. Ames¹, Ming-Yue Lee¹, Colleen Moody¹, Wenjun Zhang⁴, Yi Tang⁴, and Shiou-Chuan Tsai^{1,2,3}*

¹Department of Molecular Biology and Biochemistry, ²Department of Chemistry, and ³Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697

⁴Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90096

Figure S1. General route for Type II PKS polyketide chain formation and subsequent ketoreduction and/or cyclization leading to aromatic ring formation. Step 1 of the process involves priming KS with an acyl-CoA followed by decarboxylative condensation of malonyl units to produce a polyketide chain. Step 2 involves optional ketoreduction (in most cases specific for the C-9 carbonyl group), cyclization and aromatization. The three main classes of aromatase/cyclase (ARO/CYC) enzymes associated with a particular cyclization pattern and enzyme activity are illustrated, including TcmN¹⁻³, WhiE-ORFVI⁴, RemI^{5.6}, ZhuI^{7.8}, MtmQ^{9,10}, ActVII¹¹, and Gris-ORF4^{12,13}. In general, class 1 are mono-domain proteins that promote dual cyclizations/aromatizations of an unreduced polyketide chain to yield regiospecific C9-C14 and C7-C16 bonds. Class 2 are mono- or di-domain proteins that also act on an unreduced polyketide chain but catalyze C7-C12 cyclization and aromatization) to catalyze dehydration/aromatization of the first ring.



Figure S2. Multiple sequence alignment of select PKS ARO/CYC proteins. Sequences 1-11 are mono-domain proteins and sequences 12-22 are di-domain proteins. The sequence alignment was generated using ClustalW and colored using GeneDoc (red, green, and blue shading represent 100%, 80%, and 50% identity respectively). Numbering is according to the ZhuI sequence. Annotations included as part of the consensus line represent a conserved property at this position within a 3-residue exception limit (polar [%], hydrophobic [!], charged [#], aromatic [\$]). An asterisk below the consensus line marks residues defining the interior pocket of ZhuI. Arrows indicate the three putative catalytic residues of ZhuI.

												NCBI
			1	10	20	3	0 40		50	60	70	Accession
Monodomain	S.glaucescens TcmN ARO/CYC	:	MAARTDN	ISIVVNAPFEI	WDVTND	IEAWPEL	ESEYAEAEILR	QDGDGFD	FRLKTRPDAM	IGRVWEW	VSHRVPDKGSRTVRAH	P16559
	S.coelicolor_WhiE-ORFVI	:	MAGHTDN	IE <mark>ITIAAP</mark> MEI	WNMTND	IEK <mark>WP</mark> GL	FSEYAS <mark>VE</mark> VLG	RDDDKVT	FRL TMHPDAI	GKVWSW	VSERVADPVTRTVRAQ	P23154
	S.griseus_FdmI	:	MAASTDN	ISVVVDAPME1	LWDMTND	IES <mark>WP</mark> GL	FSEYAKAEILG	RDGATVT	F <mark>RL</mark> TLHPDES	G <mark>GAVWD</mark> W	VSERTPDPATRTVRAR	AAQ08919
	S.sp.SANK-61196_SanI	:	MALTDN	ITIVIDAPMDI	WEMTND	VESWPTL	FSEYAAAEILG	RTGATVR	F <mark>RL</mark> TLHPDEN	I <mark>G</mark> AVWSW	V <mark>SERTPDP</mark> ETRTVRAR	ADG86318
	S.cyaneus_CurF	:	MAGHTEN	IE <mark>ITIAAP</mark> VDI	WDMTND	L <mark>E</mark> RWPEL	FSEYASCEVLS	REANTVT	FRL TMHPDEN	I <mark>GKV</mark> WSW	V <mark>SERTAD</mark> REKLVVRAR	CAA44383
	S.resistomycificus_RemI	:	MPRIEN	IS <mark>IVIG</mark> ADPRI	V FDVTND	IARWSEI	FDEYSHAKVLS	EERDGRWTEIV	FELTNEEGA0	WRSWRI	LDHRELVAVAER	CAE51178
	S.olivaceus_ElmNI	:	MTTTDN	TITIRAPFDI	WRRTND	VE <mark>SWP</mark> TL	FAEYAS <mark>V</mark> DVLR	RDGDAVE	F <mark>RL</mark> TTRPDAI	0 <mark>GKV</mark> WSW	VSRREPDLASRTVSAR	CAP12603
	S.collinus_RubF(1-155)	:	MAGHTDN	IG <mark>IVVE</mark> APMDI	WTMTND	VAA <mark>WP</mark> RL	FNEYASA <mark>E</mark> I <mark>L</mark> G	RDGDTVT	FRLALHPDEN	I <mark>G</mark> TVWSW	VSERTPDPATRTVVAR	AAG03070
	A.mediterranei_ActV	:	MAAHTDN	ISVFIE <mark>AP</mark> MEI	LVWQRTND	VE SWTEL	FDEYGEATVLD	RRDDRIT	FRLALHPDEN	I <mark>GKV</mark> WAW	VSERRLDEAARTTLSH	ADJ47192
	S.tendae_LlpCI	:	MSGHTDN	ISTVIDAPLDI	LVWDMTND	VAS <mark>WP</mark> DL	FSEYAEATVLE	RDGNRIV	FRLAMHPDA (GTVWSW	VSE <mark>RILDP</mark> VART <mark>V</mark> HAR	CAM34342
	S.sp.R1128_ZhuI	:	MRHVEH	ITVTVAAPADI	WEVLAD	VLGYADI	FPPTEKVEILE	EGQGYQV	VRLHVDV-A-	GEINTW	TSRRDLDPARRVIAYR	AAG30197
Didomain	S.coelicolor_ACTVII	•	MSRPGEHRVVH	TLRTQAPARE	RLYELVAR	VEDWPAV	FEPTVH <mark>V</mark> QVLE	RGPGTER	FRIWARV-G-	GRVKTW	TSRRTLDPDTLRVTFR	Q02055
	S.griseus_gris-ORF4	:	MSQPGLREVEH	EITVS <mark>APA</mark> AA	AVYRLIAE	VRNWPRI	FPPTIYVDHVE	ENGDEER	I <mark>RIWA</mark> TA-N-	DEAKNW	TSRRTLDPEGLRITFR	CAA54862
	S.peucetius_DpsF	:	MSELPLQQTEH	EIHTSAAPDA	AVFAVLAD.	ARAWPAV	FPPSVHVEQVE	HTGSSER	I <mark>RIWA</mark> TA-N-	GSLRTW	TSRRELDERARRIRFR	AAA65203
	S.nogalater_SnoaE	:	MTAGDTTLTTHRTEH	IRV <mark>TV</mark> S <mark>APA</mark> RE	VFDLVAD	ITGWPHT	PPTVHAE YAE	RGASEER	I <mark>RIWA</mark> TA-N-	-GEVKAW	TSHRSLDREGLRVRFR	CAA12012
	S.rimosus_OxyK	:	MPAPTSHRAVH	RTEIDAPADE	VYALIRD	AAEWPRH	FTPTVHVERAE	LDARSER	L <mark>RIWA</mark> TA-N-	-GEVKHW	TSHRALDPEGQSVRFR	AAZ78334
	S.fradiae_UrdL	:	MTTRSVEH	IE I TVE APA AF	AVYRLIAE	VENWPRI	FPPTIYVEHLE	RGEGEER	I <mark>RIWA</mark> TA-N-	-GKAKNW	TSRRTLDADNLRITFR	AAF00205
	S.venezuelae_JadD	:	MTTREVEH	IE I TVA APA QA	VYRLIAE	VENWPRI	FPPTLYVDHVE	RGEREER	I <mark>RIWA</mark> TA-N-	-GAPKNW	TSKRTLDPDNLRITFR	AAB36566
	S.sp.AM-7161_med-ORF19	:	MTQPTPRETAH	IE I TVRATAEF	RLYELIAD	VGGWPSI	FPPSVHADHLE	RGDKEER	I <mark>RLWA</mark> TV-D-	-GQVKHW	TSRRTLDRAGLRVDFR	BAC79027
	S.albaduncus_ChryK	:	MSGPKVHEAEH	IHITVAASAKA	AVHQLLVE	VENWPLL	FPPSVHVERLE	QDGEHER	IRIWATA-N-	-GAAKTW	TSRRRIDAAEGRIEFR	CBH32086
	S.aureofaciens_AurH1	:	MTQPGLREVEH	IE ITVSAPAAA	VYRLIAE	VVNWPRI	FPPTIYVDQVP	EGPGEER	IRIWATAP	IGEAKNW	TSRRTLDAEALRITFR	AAX57195
	_S.cinnamonensis_gra-ORF4	:	MTTRQVEH	EYTIGAPAA	VYRLLAD	VSHWPQI	PPTIHVERQA	TGAHQER	IHIWATAN	GOAKNW	TSRETLDPEALRIDSA	P41178
			*	**!-!-Ap!	-v\$%!#	!wp-!	E'%!%!#-!- * **	·#***	!r!*!*! * * *	-g#!-*w	-s#R-!d!-!r	
			1.12			10.00	575 200(83.0	1.5. 2	10 (10) M	an), an	Ť	
			80 90	100	1	10	120	130	140	150	160	
Monodomain	S.glaucescens_TcmN_ARO/CYC	:	RVETG-PFAYMNLH	TYRAVAGG-1	TEMRWVQE	FDMKPGA		P FDNAHMTAHLI	NTTT <mark>R</mark> ANMEI	RIKKIIE	DRHREGQRTP	
	S.coelicolor_WhiE-ORFVI	:	RVETG-PFQYMNIV	EYAETAE <mark>G</mark> -1	TVMRWTQD	<mark>F</mark> AMKP <mark>D</mark> A		PVDDAWMTDNII	N <mark>RNSR</mark> TQMAI	IRDRIE	QA AGERRTASVLAD	
	S.griseus_FdmI	:	RIE <mark>T</mark> G- P FVHMNILW	TYASTDAG-1	/EMRWRQE	FEVKPGL		P FGDAEMTERLI	NT <mark>NTRREMA</mark> H	RIKELVE	KAAAERRAAR	
	S.sp.SANK-61196_SanI	:	RIETG-PFEHMDIHW	TYADTDG <mark>G</mark> -V	/EMRWRQE	FTVRPGL		PFGDTEMTERL1	NT <mark>NTRREMA</mark> H	RIKGLVE	RAAAVAR	
	S.cyaneus_CurF	:	RVETG-PFEYMNIV	EYEETPDG-1	RMHWTQD	<mark>F</mark> AMKP <mark>D</mark> A		PVDDAGMTDII	N <mark>RNS</mark> PIQMA	IRDRIE	EVSCTTP	
	S.resistomycificus_RemI	:	RDPLY-PFAYMHLR	SYQEVPE <mark>G</mark> -7	CLMTWIQD	FELDDRF		EVPLATVLERM	NTHTRHNQA	<mark>IK</mark> QKI <mark>E</mark>	SGAVRP	
	S.olivaceus_ElmNI	:	RVETG-PFRFMHLTW	TYQETPDG-V	/VLRWIQE	FEVADEA		PFDDAAMAERI	DRNTRLNMGH	RIKQLIE	ADATKETV	
	S.collinus_RubF(1-155)	:	RVETG-PFAHMNIR	EYEEVPG <mark>G</mark> -V	/RMRWIQD	FAMKPDA		PIDDAGMTARLI	N <mark>RN</mark> TAVQMRI	L <mark>IK</mark> AKVE	TAARAGRQLQ	
	A.mediterranei_ActV	:	RVETG-PFKYMTLFW	EYTEVAG <mark>G</mark> -V	RLRWVQD	FEMKPGA		PLDDHAMAARII	NG <mark>NS</mark> VRQQ <mark>A</mark> I	IKQLLE	QAAREAEVPV	
	S.tendae_LlpCI	:	RVETG-NFKYMWLFW	EYTTEDDG-V	RLRWVQD	FELKPGL		PMDDAAMTDRLI	NA <mark>NS</mark> VAQ <mark>L</mark> EI	IKEKIE	AVARATAATR	
	S.sp.R1128_ZhuI	:	QLETAPIVGHMSGEW	RAFTLDAER	rqlv <mark>l</mark> thd	FVTRAAG	DDGLVAGKLTP	DEAREMLEAVVI	ERNSVADLNA	VLGEAE	RRVRAAGGVGTVTA	
Didomain	S.coelicolor_ACTVII	:	QELTQPPIASMGGSW	EFRGDGDG-1	revvlthd	FAAVD		EAALPGLREAL	DANSGKELA	LVALAE	RRQPPEELV	
	S.griseus_gris-ORF4	:	QEVPAPPVAAMGGAW	IIEPLGDAES	SRIRLLHD	YRAVDDD		PAHLEWIDQAV	DRNSRSELA	ALKGNVE	KAHAEE	
	S.peucetius_DpsF	:	QEVSAHPVAAMGGE W	IVEEAGDGG	TRVRLTHD	FRAVDDD		PETIGWIHRAV	DRNSEAELAS	SLRTAL	RPDGTA	
	S.nogalater_SnoaE	:	QEKSQHPVAGMGGE W	IIRPVDGERS	SEVI <mark>LTHD</mark>	FQAVEDD		PAHVDWIHRAV	DRNSGAELAA	ALKPRRN	AR	
	S.rimosus_OxyK	:	QEVCSPPVAAMSGEW	VLRDLPGGRO	CELTLHHT	FAAVDDR		PEDVEWITTAT	DRNSRTELAN	IKALAE	AAGSDA	
	S.fradiae_UrdL	:	QEVSTPPVAAMGGT	IIEPLSGDSS	SRIRLLHD	YRAVDDD		PQGLKWIDEAV	DRNSRSELA	ALKTNVE	LAHASEEIT	
	S.venezuelae_JadD	:	QEVSTPPVAAMA-TW	IIEPVSATES	GSALLHD	YRAIDDD		PEGLKWVDEAV	DRNSRSELA	ALKTNVE	LAHASEEITF	
	S.sp.AM-7161_med-ORF19	:	QEVPSPPMAAMGGAW	IIEPTGPGEC	RVRLLHD	FRAVGDD		PGTLDWIDRAV	DANSRSELA	LKRHAE	RGADGP	
	S.albaduncus_ChryK		QEVSSPPVAAVGGT	TVEADGEER	RVRLLHD	YRAVDDD		PAALEWIAAAV	DRNSRAELDZ	LKTHAE	AAADGS	
	S.aureofaciens_AurH1	•	QEVTAPPVAAVGGT	TIEELSAQES	SRVRLLHD	YRAVDDD		ADSLAWIDEAV	DRNSRSELA		AAHAAE	
	S.cinnamonensis_gra-ORF4	:	SEVTTAPVAAVGGT	IVEPLGADSS	SRIRLLHD	SGPSPR-		PODLOWIDRAV	DTTRHLELA	LATTSN	HAHAAEEREL	
			* ** * * *		-15!8!-8# * *	\$-i		* **	551155! - ** *	*		
					Ť				Ť			

Figure S3. Size-exclusion chromatography studies illustrating that ZhuI ARO/CYC is dimeric in solution. (A) Elution profile of proteins used as standards to calibrate a Superdex 200 10/300 GL column (GE Healthcare): 1, ferritin (440 kDa); 2, catalase (232 kDa); 3, aldolase (158 kDa); 4, albumin (67 kDa); 5, ovalbumin (45 kDa); 6, chymotrypsinogen A (25 kDa); 7, ribonuclease A (13.7 kDa). Standard proteins were reconstituted to give a final concentration between 2.0-7.5 mg/mL then 20 µL of each sample was injected onto the column. The mobile phase for each run consisted of 20 mM sodium phosphate buffer, pH 7.0. The flow rate was 0.5 mL/min. (B) Elution profile of ZhuI ARO/CYC when 1 mL of protein at ~10 mg/mL (solid line), or ~2 mg/mL (dashed line), was injected and ran under the same conditions as described for (A). An asterisk marks aggregated protein eluting near the column void volume. (C) Calibration curve prepared by plotting K_{av} (the gel-phase distribution coefficient) vs. log molecular weight (in kDa). $K_{av} = (V_e - V_0)/(V_c - V_0)$, where V_e is the elution volume, V_0 is the column void volume (determined using blue dextran), and V_c is the column volume (taken as 24.5 mL). Linear regression analysis of the plotted standards was used to estimate the molecular weight of ZhuI to be 48 kDa (theoretical for dimer is 41 kDa).



Figure S4. Dimeric structure of ZhuI ARO/CYC and comparison to other helix-grip fold (panels B and D) or hotdog fold (panel C) dimers. (A) ZhuI ARO/CYC from *Streptomyces* sp. R1128 (PDB ID, 3TFZ). (B) The Tcm ARO/CYC dimer resulting from the addition iodide (represented as purple spheres) to the crystallization drop (PDB ID, 2REZ). (C) The *Pseudomonas aeruginosa* fatty acid dehydratase FabZ 'double-hotdog' dimer (PDB ID, 1U1Z). (D) Dimeric structure of SMU.440 from *Streptococcus mutans* (PBD ID, 2B79).





Figure S5. Stereoview images of the docking results described in the main text Figure 6 for ZhuI and TcmN ARO/CYC.

Figure S6. Circular dichroism (CD) data for WT and mutant ZhuI. All protein samples were prepared for CD by diluting the purified ZhuI (prepared for crystallization trials as described in the Materials and Methods section) to 15 μ M with water and filtering (0.2 μ m). Data were collected using a Jasco J-715 CD spectropolarimeter at 25 °C in a 0.1 cm path length quartz cuvette.

