Supporting Information

Korman et al. 10.1073/pnas.0913531107

SI Text

Materials and Methods. *Enzyme assays.* TE/CLC* hydrolytic assays were carried out for the *apo*-mutant and mutants of the catalytic triad residues as previously described and with minor adjustments as noted below (1, 2). Briefly, purified proteins were exchanged into 50 mM potassium phosphate buffer (pH 7.5) on EconoPac 10DG columns (BioRad) according to the manufacturer's instructions. A solution of benzoyl-SNAC (250 mM) in DMSO was prepared for substrate addition. The hydrolysis of a series of concentrations of benzoyl-SNAC was monitored in triplicate after the addition of enzyme (10 μ M final), and 100 μ l aliquots of the reaction were quenched in 100 μ l of 8 M urea at 2, 5, 10, and 15 min followed by incubation on ice. The resulting free thiols were reacted with 5,5'-dithiobis-2-nitrobenzoate (8 mg/ml in ethanol, 8 μ l) for 15 min at 20°C and quantified by UV-visible spectroscopy at 412 nm (1, 2). Reactions were also

 Udwary D, Merski M, Townsend CA (2002) A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. J Mol Bio 323:585–598. carried out using a 125 mM benzoyl-SNAC in DMSO solution for substrate addition leading to a 2x higher concentration of DMSO for the enzyme reactions: *apo*-mutant (*apo*-TA1): $k_{cat} = 0.043 \pm 0.0025 \text{ s}^{-1}$, $K_m = 2.61 \pm 0.37 \text{ mM}$, $k_{cat}/K_m = 0.017 \pm 0.003 \text{ s}^{-1}/\text{mM}^{-1}$; a p o - TA 1 (D 2 0 7 0 N): $k_{cat} = 0.0345 \pm 0.0017 \text{ s}^{-1}$, $K_m = 2.66 \pm 0.33 \text{ mM}$, $k_{cat}/K_m = 0.013 \pm 0.002 \text{ s}^{-1}/\text{mM}^{-1}$; all catalytic triad mutants gave undetectable activity. The added DMSO had little effect on enzyme activity.

CD analysis. Purified proteins from both the *apo-* and *holo-*series were prepared as for the enzyme assays. Concentrations were determined by Bradford assay with bovine albumin as a standard. Proteins were brought to 20 μ M final concentration and their CD spectra (JASCO J-710) were collected as the average of three scans.

 Gokhale RS, Hunziker D, Cane DE, Khosla C (1999) Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chem Biol* 6:117–125.



Fig. S1. CD spectra for apo- (Left) and holo-ACP-TE (Right) didomains and their mutants used in this study.

010050	β1	β2	η1	β3	η2
Q12053	1850 1860	1870	1880	1890 1900	1910
Q12053 AAS90022 AAS90047 EAA61613 AAZ95017 AAT69682 Q03149 AAU10633 CAB92399 EAA18956 AAD38786	KPYCRPSTSVVLQGL KPYCRPSTSVVLQGL VPECRPTTSVVLQGL TDSCRSTNSVILQGK KGRIPPAWSMYLQGS TSRHPPATSILLQGN SLNHPPATSILLQGN SLNUFPASTLLQGS VAEIPRSTSTILQG NYPHRKATSVLLQGN	PMVARKTLFMLP PMVARKTLFMLP PMVARKTLFMLP PMXRKTLFMLP PKTAAKTLFLLP QKRSKEILFLFP PHTATKKLFMFP PRTASKTLFLP PSKARSTLFLP TKHCSQTLFLP TRTATKNLWNP HRTATRHLFMIP	DGGGSAFSYASIPRI DGGGSAFSYASIPRI DGGGSAFSYASIPRI DGGGSASSYITIPRI DGGGSASSYITIPRI DGGGAATSYLSIPRI DGSGSASSYATIPAI DGSGSATSYASIPPI DGSGSATSYASIPPI DGSGSATSYTIPSI DGSGCATSYTEISOU DGSGSATSYTEISOU	KS.DTAVVGLNCPYA KS.DTAVVGLNCPYA KS.DTAVVGLNCPYA AS.DVAVVGLNCPYA QS.DVAVVGINCPYA QS.DVAVVGINCPYA SP.DVCVYGLNCPYM SP.DVCVYGLNCPYM SPDGVAVYGLNCPYM SS.DMRVIGLNSPYL SS.NWAVWGLFSPFM GS.DVAVWGMFSPFM	RDPENMNCT.HG RDPENMNCT.HG RDPENMNCT.HG RDPENMNCT.HQ KTPHKFADHTLP KTPQNLTCS.LD KAPEKLTCS.LD KAPEKLTCS.LD KDSSYLVEFGLK TKPHEFNCA.LQ KTPEEYKCG.VY RTPEEYKCG.VY
Q12053	α1 <u>000000000000000000000000000000000000</u>	β4 1930 <u>β4</u> 194	α2 0000000000000 1950	β5 1960	α3 <u>00000</u> 970 1980
Q12053 AAS90022 AAS90047 EAA61613 AAZ95017 AAT69682 AAC39471 Q03149 AAU10633 CAB92399 BAA18956 AAD38786	A MIESFCNEIRRQP AMIESFCNEIRRQP AMIESFCNEIRRQP AMISFCNEIKRQP DVIASVVEGIRGRQA ELTEPYLAEIRRQP GLTELVVNEILRRQP GITELVVNEILRRQP GITELVVNEILRRQP GMAAKFIEAMKARQS GMATKFIEMKRQP	RGPYHLGGWSSG QGPYHLGGWSSG EGPYHLGGWSSG QGPYHLGGWSSG QGPYHLGGWSSG KGPYSFGGWSAG KGPYNLGGWSAG QGPYNLGGWSAG QGPYNLGGWSAG QGPYHLAGWSAG QGPYSLAGWSAG EGPYAVSGWSAG	GAFAYVVAEALVN.Q GAFAYVVAEALVN.Q GAFAYVVAEALIN.A GAFAYVTAEALIN.A GIFAYVTAEKMIK.Q GILAYAVAQELILE GICAFDAARQLILE GICAYDAARKLVLQQ GICAYEAALIITRA GVSAFDAARQLVS.E GVIAYEIVNQLTK.A GVIAYEIVNQLTKA	GEEVHSLIIIDAPIP GEEVHSLIIIDAPIP GEVHSLIIIDAPVP GDEVHSLIIIDAPVP GDEVGSLFIFDAPVP GEEVSTLLIDSPFP GEEVETLLIDSPFP GEIVETLLIDSPNP GEIVESLILIDSPNP GEIVESLILIDAPCP GEIVESLILIDAPCP GEVVSHLIIDAPCP	QAME OLP RAFYE QAME OLPRAFYE QAME OLPRAFYE QVMEKLPRSFYE QVMEKLPRSFYE QVMEKLPRFYE IGLEKLPPRLYK IGLEKLPPRLYS VGLEKLPPRLYS VGLEKLPPRLYD VGLGKLPKRMYD VTIEPLPRSLHA ITIEPLPAGLHA
Q12053	2222	<u>ک</u>	α4 η3 000000000000000000000000000000000000	2030	β6
Q12053 AAS90022 AAS90047 EAA61613 AAZ95017 AAT69682 AAC394711 Q03149 AAU10633 CAB92399 EAA18956 AAD38786	HCNSIGLFATQPGA. HCNSIGLFATQPGA. HCNSIGLFATQPGA. YCNNLGLFSNQPGG. AVNFTES HCTNVGLFGTELSRG FFNSIGLFGGGK FFNSIGLFGGGK FLNSQNVFGSDNPHS FLEKSGIFGAFEMG. WFASIGLLGEGDDEA WFAEIGLLGEGDGE.	SPDGSTEPPSYL SPDGSTEPPSYL SPDGSTEPPPYL TTDGTAQPPYL TAVGTAEPPPYL S.GGPNKTPEWL AAPPDWL TAGTSVKAPEWL AKKIPSWL AKKIPSWL	TPHFIAVVDVMLDYK IPHFIAVVDVMLDYK IPHFTAVVDVMLDYK IPHFAXVDVMLDYK MPHFRASIELLHGYH LPHFLAFIDSLDAYK LAHFLAFIDSLDAYK IAHFLAFIDALDAYV FQHFCVFIEALDRYV LPHFAASVTALSNYT	LAPLHARRMP. LAPLHARRMP. LAPLKARRMP. VAPLKTNRMP. CKPLQTKKMP. APPMKLGNKT. AVPULPFNDSKWAKKMP AVPLPFNDSKWAKKMP AVPUDSGLVGLASPLP. PEPFEHGMA. AEPIPKEKCP. ADPIPKDKCP.	KVGIVWAAD KVGIVWAAD KVGIVWAAD KVGIWAAD KVGIWAAS KVGIWAAS KVMVIWAGE KTYLWAKD APPQTYMLWAED KTTIIWAAD KTTIIWAAD KTTIIWAED KTTIIWAED
		_	α5	β7	η4
Q12053	2050	T	T <u>2060</u> <u>2070</u>	Q → TT 2080	2090
Q12053 AAS90022 AAS90047 EAA61613 AAZ95017 AAT69682 AAC39471 Q03149 AAU10633 CAB92399 BAA18956 AAD38786	TVMDERDAPK TVMDERDAPK TVMDEDNAPK TVMKEDEAPK CAFDGVRYAHIPPSA GVCGKPGDPW.PEPA GVCKNSDSAR.PEYR GVCKNPDDPR.PEAQ GVCKLPTDPR.PDPY GVCKLPTDPR.PDPY	MKGMHF MKGMHF MKGMHF GDTDEDTEGMKF EDGSEDPREMQW EDGSKDPREMQW DDDPREMRW PDDPREMWW PTGHALF PTGHALF	MIQKRTEFGPDGWDT MIQKRTEFGPDGWDT MIQKRTNFGPDGWDV MIQKRTNFGPDGWDV LTEKRKDFGATEWAS LLNDRTDLGPNGWDE LLENRTDLGPNGWEA LLNNREDFGPNGWDE LLDNRTDFGPNRWDE LLDNRTDFGPNRWDE	IMPGASFD.IVRADGA VMPGASFD.IVRADGA VCPGAVFD.IVRADGA VCPGAVFD.IVRAUDA LFPGTDVD.ARVVESE LVGPQNIGGIHVMEDA LVGKENIGGITVIHDA LLGGKEGLFMDRIAEA FIGAGNIS.TMAIENA YLDVNKFR.TRHMPG YLDIEKMT.FHHMPG	NHFTLMQ NHFTLMQ NHFTLMSNCLPT NHFTLMSNCLPT NHFTMT NHFTMTK NHFTMTK NHFSMLKR NHFSMMHG NHFSMMHG
012053	0.0.0.0	α6			
Q12053 AAS90022 AAS90047 EAA61613 AAZ95017 AAT69682 AAC39471 Q03149 AAU10633 CAB92399 BAA18956 AAD38786	ZIOQ ZIOQ KEHVSIIS SHLNRKAEDSLELES KARVNYVS GQKAK.ELS GKAK.ELS GKAK.ELA GKAK.ELA GKAK.ELA GKAK.ELA GKAK.ELA 	DLIDRVMA DLIDRVMA SLVPISSR DLIDKVMG EHMRDGLG QFMATAMS TFMKNALG AFLARALD AKIRETMG GFLKEGIH			

Fig. S2. TE/CLC domain alignment. Catalytic triad residues are indicated with a black arrow. See text for critical structural residues involved in binding. Secondary structural elements are shown for PksA TE/CLC from *Aspergillus parasiticus*. TE/CLC sequences consist of *A. parasiticus* PksA (accession number, Q12053; metabolite, aflatoxin), *A. flavus* PksA (AAS90022; aflatoxin), *A. nomius* PksA (AAS90047; aflatoxin), *A. nidulans* STCA (EAA61613; sterigmatocystin), *Mycosphaerella pini* PksA (AAZ95017; dothistromin), *Cercospora nicotianae* CTB1 (AAT69682, cercosporin), *A. fumigatus* Alb1 (AAC39471; naphthopyrone YWA1), *A. nidulans* WA (Q03149; naphthopyrone YWA1), *Gibberella zeae* PKS12 (AAU10633; aurofusarin), *Gibberella fujikuroi* PKS4 [CAB92399 updated sequence (1); bikaverin], *Colletotrichum lagenarium* PKS1 (BAA18956; melanin), and *Nodulisporium sp.* ATCC74245 PKS1 (AAD38786; melanin).

1. Crawford JM, Vagstad AL, Whitworth KP, Ehrlich KC, Townsend CA (2008) Synthetic strategy of nonreducing iterative polyketide synthases and the origin of the classical "starter-unit effect." ChemBioChem 9:1019-1023.

SANG SANG



Fig. S3. SDS-PAGE image for the TE/CLC monodomain purification (His, His₆-tagged construct; Cleaved, protein after thrombin treatment; and Q-Seph, 8-fractions enriched after anion exchange chromatography). Image of TE/CLC monodomain crystal.

			SeMet TE	
	Native PksA TE	Peak	Inflection	Remote
Data Collection				
Wavelength (Å)	1.0	0.98	0.98	1.2
Space Group	P212121		P212121	
Cell Dimensions				
a (Å)	66.93		66.91	
b (Å)	66.94		67.13	
c (Å)	67.92		67.80	
Resolution Range (Å)	50-1.7	50-3.0	50-3.1	50-3.0
No. of Observations	402056	269924	269977	274209
No. of Unique Reflections	34357	6539	5937	6510
Redundancy	6.6	11.2	11.2	12.2
Completeness (%) (last shell)	99.9 (99.2)	98.7 (90)	98.7 (90.8)	99.8 (99)
Mean $I/\sigma(I)$ (last shell)	40.7 (5.7)	13 (4.5)	13.2 (4.1)	16.5 (6.7)
R _{merge} (last shell)	5.7 (29.5)	19.1 (28.1)	18.9 (32.1)	18.5 (29.6)
MAD phasing				
Resolution Range (Å)			50-3.5	
Number of Derivative Sites			13	
FOM			0.5	
Refinement				
Resolution Range (Å)	50-1.7			
No. of Reflections	33560			
No. of Protein Atoms	2042			
No. of Waters	170			
R _{crys} (%)	18			
R _{free} (%)	20			
Geometry				
RMSD for Bonds (A)	0.0058			
RMSD for Angles (deg)	1.31			
RMSD for B Main Chain	1.005			
RMSD for B Side Chain	1.55			
Ramachandran Plot (%)				
most favored	90.5			
tavored	8.5			
generously allowed	1.0			
disallowed	0			

DN AS

2 N A C