Supporting Information for

De Novo Design and Implementation of a Tandem Acyl Carrier Protein Domain in a Type I Modular Polyketide Synthase

Zilong Wang,[†,§,⊥] Saket R. Bagde,[†] Gerardo Zavala,[†] Tsutomu Matsui,[▽] Xi Chen,[#] and Chu-Young Kim[,†,‡]*

[†]Department of Chemistry and Biochemistry, The University of Texas at El Paso, 500 West University Avenue, El Paso, Texas 79968, United States

[‡]Border Biomedical Research Center, The University of Texas at El Paso, 500 West University Avenue, El Paso, Texas 79968, United States

[§] Department of Biological Sciences, National University of Singapore, 117543 Singapore

[⊥]NUS Synthetic Biology for Clinical and Technological Innovation (SynCTI), Centre for Life Sciences, National University of Singapore, 119077 Singapore

[▽]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Stanford University, 14 2575 Sand Hill Road, MS69,Menlo Park, California 94025, United States

[#]College of Chemistry and Materials Science, Northwest University, 1 Xue Fu Avenue, Xi'an 710127, People's Republic of China

**To whom all correspondence should be addressed. The University of Texas at El Paso*
1852 Hawthorne Street, El Paso, Texas 79902, Tel: (915) 747-6935, E-mail: ckim7@utep.edu

Table of Contents

Materials and methods

Chemicals, plasmids, and strains. All reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless stated otherwise. DEBS3-pET21 plasmid, SFP-pET21 plasmid, NDK-SNAc, and *E. Coli* BAP1 strain were gifts from Dr. Chaitan Khosla (Stanford University).

Construction of Module 6-TE-pET28 and Module 6-3ACP-TE-pET28 expression plasmids and overproduction of the target proteins. Primers used to clone different gene fragments and restriction digestion-ligation sites are shown in Table S1. Module 6-3ACP-TE-pET28 plasmid was custom synthesized by GenScript (Piscataway, NJ) with codon optimization and then sub-cloned into pET28 vector with *NdeI*-*EcoRI* restriction sites. Module 6-TE-pET28 was constructed by inserting Module 6-TE into pET-28. Module 6-3ACP-TE-pET28 was constructed by sequentially inserting KS-AT-KR, 3×ACP, and TE into pET-28 (Table S2).

Expression and purification of module 6-TE and module 6-3ACP-TE. Cell stock was inoculated into 10 mL of LB broth containing 50 μg/mL kanamycin, incubated at 37 ℃ overnight at 225 rpm. This seed culture was used to inoculate 1 L of LB broth containing 50 μg/mL kanamycin and then incubated at 37 ℃ for 2 to 3 hours, at 225 rpm. When the OD_{600} reached 0.6, the culture was cooled on ice for 15~20 min. 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression and incubated at 18 ℃ to allow the protein expression. Cells were harvested by centrifugation at 7,000 rpm at 4 ℃. Harvested cells were transferred to a 50 mL Falcon tube kept at -80℃ until needed. Pellet from 1 L of cell culture was thawed at room temperature and resuspended in 10 mL cell lysis buffer (50 mM NaH2PO4: NaOH, pH 7.8, 300 mM NaCl, 10 mM PMSF). Cells were lysed by sonication for 15 min (500-Watt, 30% amplitude, pulse 5s on and 10s off). Cell debris and other insoluble material were removed by centrifugation for 45 min at 13,000 rpm at 4 ℃. The soluble fraction was filtered using a 0.22 μm filter, mixed with 1 mL of Ni-NTA resin (Thermo Scientific, USA), and incubated at 4 ℃ for 1 hour under constant mixing. Resulting resin was washed with wash buffer (50mM sodium phosphate, pH 7.8, 0.3 M NaCl, 20 mM imidazole) and the target protein was eluted using the elution buffer (50 mM sodium phosphate, pH 7.8, 0.3 M NaCl, 200mM imidazole).

Expression and purification of SFP. Recombinant SFP protein was prepared as described previously [1]. *E. Coli* BL21(DE3) transformed with the SFP-pET21 plasmid was grown at 37 ℃ until OD_{600} of 0.6. Protein expression was induced with 0.5 mM IPTG and the cell culture was maintained at 37 ℃ for 4 h. SFP was purified using Ni-NTA agarose (Thermo Scientific, USA), and mixed with 2 mM EDTA and 5 mM DTT at 4 ℃ overnight to chelate Mg^{2+}. Size exclusion chromatography was performed with the Superdex 200 10/300 GL column (GE healthcare, USA) in 10 mM HEPES-NaOH, pH 8.0, 5 mM DTT, 120 mM NaCl buffer to further improve the protein purity and change the buffer. Glycerol was added to 50% final concentration, protein solution was flash frozen, and kept at -80℃ until needed.

***In vitro* conversion of ACP from *apo*- to *holo*- form by recombinant SFP protein.** Reaction mixture containing 1 mM SFP, 20 mM module 6-TE or module 6-3ACP-TE, 30 mM coenzyme A or 90 mM coenzyme A, respectively, 50 mM Tris-HCl, pH 8.5, 0.3 M NaCl, 10 mM $MgCl_2$, and 5 mM DTT was incubated at 37 ℃ for 1 hour.

***In vitro* production of triketide lactone and ketolactone by Module 6-TE and Module 6-3ACP-TE.** This experiment is performed as described elsewhere with minor modifications [2]. The 200 μL reaction mixture contained 5 μM module 6-TE or module 6-3ACP-TE, 1 mM NDK-SNAc, 2.5 mM TCEP, 1 mM EDTA, 0.2 mM methylmalonyl CoA and 4 mM NADPH, 400 mM phosphate adjusted to pH 7.2, and 20% glycerol. The reaction mixture was incubated at room temperature for different time periods. 20 μL samples were analyzed by LC-MS/ESI.

***In vitro* production of triketide ketolactone by Module 6-TE without NADPH.** The 200 μL reaction mixture contained 5 μM module 6-TE or module 6-3ACP-TE, 1 mM NDK-SNAc, 2.5 mM TCEP, 1 mM EDTA and 0.2 mM methylmalonyl CoA, 400 mM phosphate adjusted to pH 7.2, and 20% glycerol. The reaction mixture was incubated at room temperature for different time periods. 10 μL samples were analyzed by LC-MS/ESI.

LC-MS/ESI analysis of triketide lactone and ketolactone production. LC analysis was carried out on a Poroshell 300SB-C18 column (5 μm, 300 Å, 21 × 75 mm, Agilent, USA), and eluted with a gradient of CH_3CN from 2% to 95% in 0.1% formic acid-H_2O in 18 min at a flow rate of 0.2 mL/min. 1 mM β-ketoacyl-SNAc was added as standard before every sample injection to normalize variations in chromatography, ionization and detection. MS/ESI was performed on Agilent 6550 iFunnel Q-TOF LC/MS System and the data was analyzed by MassHunter software.

Small-angle X-ray scattering (SAXS) analysis of Module 6-TE and Module 6/3×ACP6-TE. SAXS data was collected on the Bio-SAXS beam line BL4-2 at Stanford Synchrotron Radiation Lightsource (SLAC National Accelerator Laboratory, Menlo Park, CA) using a Pilatus X 1M detector with a 2.5 m sample-to-detector distance and 12.4 keV (wavelength, $\lambda = 1.00$ Å) beam energy [3]. A 1.5-mm quartz capillary cell was maintained at 20 °C that kept sample aliquot in the X-ray beam. The momentum transfer (scattering vector) q was defined as $q = 4\pi\sin(\theta)/\lambda$, where 2θ is the scattering angle. The q scale was calibrated by silver behenate powder diffraction [4]. All data were collected by the data acquisition program Blu-ICE using SEC-SAXS mode (see more details below) [3, 5]. For each image, an integrated transmission beam intensity value was recorded by a photodiode mounted inside the beamstop.

Online size-exclusion chromatography solution X-ray scattering (SEC-SAXS) was performed using Thermo Fisher Scientific UltiMate 3000 system with two Superose 6 Increase 3.2/300 columns. The Blu-ICE SEC-SAXS mode was employed for high throughput tandem SEC-SAXS data collection and subsequent initial analyses; 500 images were recorded with 1-sec exposure every 5 sec at 0.04 ml/min flow rate. The data processing program SasTool (http://ssrl.slac.stanford.edu/~saxs/analysis/sastool.htm) was used for scaling using the transmission intensity, azimuthal integration, averaging of individual scattering images and background subtraction. The first 100 images at the early part of the void volume were averaged and it is assigned as a buffer-scattering profile for the background. After taking the background images, the X-ray shutter was closed until the elution peak of interest in order to keep the sample cell clean from radiation-damaged sample. The automatic SEC-SAXS data processing and analysis pipeline “SECPipe”, which includes real-time consecutive Guinier analysis implemented in the program AUTORG [6], was used for data quality assessment and averaging scattering profiles every 5 images. Details of the experimental setup are summarized in Table S4.

Table S1. Primers used for the cloning of gene fragments for constructing module 6-TE and module 6-3ACP-TE.

Gene	Template plasmid	Restriction sites	Primer (forward/reverse)
module 6-te	DEBS3-pET21	*NcoI-HindIII*	5'-C ATG CCA TGG GCG ACC CGA TCG CGA TCG TCG-3'
			5'-CCC AAG CTT ACT ATT CCC TCC GCC CAG CCA G-3'
ks-at-kr	DEBS3-pET21	*NcoI-NdeI*	5'-C ATG CCA TGG GCG ACC CGA TCG CGA TCG TCG-3'
			5'-A TTC CAT ATG CGC CCA GGC GCC CCA C-3'
acp	DEBS3-pET21	*NdeI-EcoRI*	5'-A TTC CAT ATG GGC GAG GGC ATG GCC AC-3'
			5'-G GAA TTC GGC GGG AGT CCC GCT GT-3'
3×acp	3×ACP6-pET28	*NdeI-EcoRI*	5'-A TTC CAT ATG GGT GAA GGC ATG GCA ACC G-3'
			5'-G GAA TTC GGC CGG GGT GCC TGA GT-3'
te	DEBS3-pET21	*EcoRI-HindIII*	5'-G GAA TTC CGG GAA GCG AGC AGC GCT C-3'
			5'-CCC AAG CTT ACT ATT CCC TCC GCC CAG CCA G-3'

Table S2. DNA sequence of chemically synthesized *3×ACP6* gene.

Sequence
GGT GAA GGC ATG GCA ACC GGT GAT CTG GAA GGT CTG ACG CGC CGT GGT CTG CGC CCG ATG GCC CCG GAA CGT GCG ATT CGT GCT CTG CAT CAG GCG CTG GAT AAC GGC GAC ACC TGC GTC AGC ATT GCG GAT GTG GAC TGG GAA CGT TTT GCC GTT GGT TTT ACG GCG GCC CGT CCG CGC CCG CTG CTG GAT GAA CTG GTG ACC CCG GCA GTG GGT GCT GTT CCG GCG GTC CAG GCA GCT CCG GCG CGC GAA ATG ACG TCA CAA GAA CTG CTG GAA TTT ACC CAT TCG CAC GTT GCG GCC ATT CTG GGT CAC AGC TCT CCG GAT GCG GTC GGT CAG GAC CAA CCG TTT ACG GAA CTG GGC TTC GAT TCT CTG ACC GCA GTT GGT CTG CGT AAT CAG CTG CAA CAA GCG ACC GGT CTG GCC CTG CCG GCA ACC CTG GTT TTT GAA CAT CCG ACC GTC CGT CGC CTG GCG GAC CAC ATT GGT CAG CAA CTG ACC CCG GCT GTG GGT GCG GTT CCG GCC GTC CAG GCA GCT CCG GCC CGT GAA ATG ACG AGC CAG GAA CTG CTG GAA TTT ACG CAT TCT CAC GTG GCG GCC ATC CTG GGC CAT AGT TCC CCG GAT GCG GTT GGC CAA GAT CAG CCG TTC ACC GAA CTG GGC TTC GAT AGC CTG ACC GCG GTG GGT CTG CGT AAC CAA CTG CAA CAG GCG ACC GGC CTG GCA CTG CCG GCT ACC CTG GTG TTT GAA CAT CCG ACG GTT CGT CGC CTG GCA GAT CAT ATC GGT CAA CAG CTG ACC CCG GCG GTG GGT GCC GTT CCG GCA GTC CAA GCC GCC CCT GCC CGT GAA ATG ACG AGT CAG GAA CTG CTG GAA TTT ACA CAC TCC CAT GTT GCC GCA ATT CTG GGT CAC TCA TCG CCG GAT GCC GTG GGC CAA GAT CAG CCG TTT ACC GAA CTG GGT TTC GAT AGC CTG ACC GCA GTT GGC CTG CGC AAC CAA CTG CAA CAA GCG ACC GGC CTG GCT CTG CCG GCC ACC CTG GTC TTT GAA CAT CCG ACC GTG CGT CGC CTG GCG GAT CAT ATC GGT CAG CAA CTG GAC TCA GGC ACC CCG GCC CGT GAA GCA AGC TCT GCT CTG CGT GAT GGT TAT CGC CAG GCC GGC GTG AGT GGT CGT GTT CGC TCC TAC CTG GAC CTG CTG GCA GGC CTG TCG GAT TTT CGC GAA CAT TTC GAC GGT AGC GAT GGT TTT TCC CTG GAT CTG GTT GAT ATG GCC GAC GGT CCG GGT GAA GTT ACG GTC TGA

Table S3. Amino acid sequence analysis of select tandem ACP domains found in nature. NRPS: nonribosomal peptide synthetase, PKS: polyketide synthase, PUFA: polyunsaturated fatty acid synthase.

Organism	**Protein**	**Tandem ACP domain**			
		Constituent ACP		ACP-to-ACP linker	
		Number of ACPs	Seq. identity (%)	Length (residues)	Seq. identity (%)
Pseudomonas fluorescens	PKS	2	37.2	26	-
Pseudomonas fluorescens	PKS	3	20.0	13-18	38.5
Aspergillus nidulans	PKS	2	68.3	67	-
Lyngbya majuscule	PKS/NRPS	3	94.9	12-25	66.7
Moritella marina	PUFA	5	88.4	33-37	33.3
Schizochytrium sp.	PUFA	9	97.0	38-50	48.9
Aureispira marina	PUFA	4	100.0	44-46	79.6
Desulfuromonas acetoxidans	PUFA	5	88.6	19-30	26.3
Pelobacter propionicus	PUFA	4	95.5	27-31	74.1
Photobacterium profundum	PUFA	5	80.0	21-31	55.6
Shwanella japonica	PUFA	6	86.0	15-29	26.7
Vibrio sp.	PUFA	5	85.5	29-34	21.9

Table S4. SAXS data collection and analysis.

	Module 6-TE	**Module 6-3ACP-TE**
Data collection parameters		
Instrument	SSRL BL4-2	SSRL BL4-2
Type of Experiment	SEC-SAXS	SEC-SAXS
Defining slits size (H mm × V mm)	0.25× 0.25	0.25× 0.25
Detector distance (m)	2.5	2.5
Detector	Pilatus3 X 1M	Pilatus3 X 1M
Beam energy (keV)	12.4	12.4
Beam current (mA)	500, 5min top-off	500, 5min top-off
q range ($Å^{-1}$)	0.051–0.39	0.051–0.39
Exposure time per frame (s)	1	1
Frames per data set	500	500
Frames used for averaging	475^{th}-489^{th}	460^{th}-489^{th}
Sample cell size (quartz capillary) (mm)	1.5	1.5
Temperature (K)	293	293
SEC parameters		
SEC column	Superose 6 Increase 3.2/300	Superose 6 Increase 3.2/300
Sample concentration (mg/mL)	17.6	14.9
Injection volume (ul)	20	30
HPLC flow rate (mL/min)	0.04	0.04
Structural parameters		
$I(0)$ from Guinier	93.73 ± 0.57	81.39 ± 0.52
R_g (Å) from Guinier	58.8 ± 0.51	69.5 ± 0.61
$I(0)$ from $P(r)$	94.66	81.92
R_g (Å) from $P(r)$	60.34	71.87
D_{max} (Å) from $P(r)$	214.3	254
q range used for P(r) estimation ($Å^{-1}$)	0.2	0.2
Porod volume estimate ($Å^3$)	665000	796000
Calculated Mr (kDa) from sequence	394	426
Software employed		
Primary data reduction	SasTool	SasTool
Data processing	PRIMUS	PRIMUS

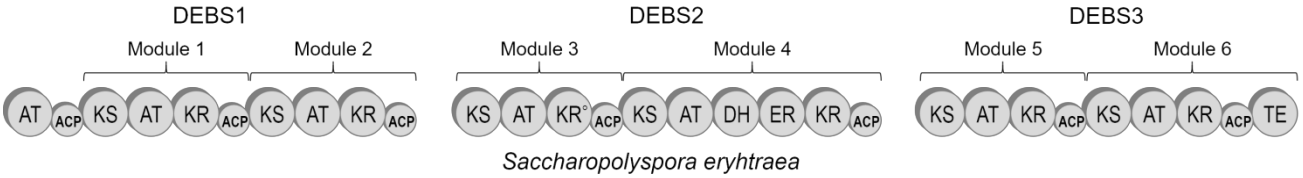

Figure S1. Domain organization of 6-deoxyerythronolide B synthase. ACP: acyl carrier protein domain, AT: acyltransferase domain, DH: dehydratase domain, ER: enoylreductase domain, KR: ketoreductase domain, KR°: inactivated ketoreductase domain, KS: ketosynthase domain, TE: thioesterase domain.

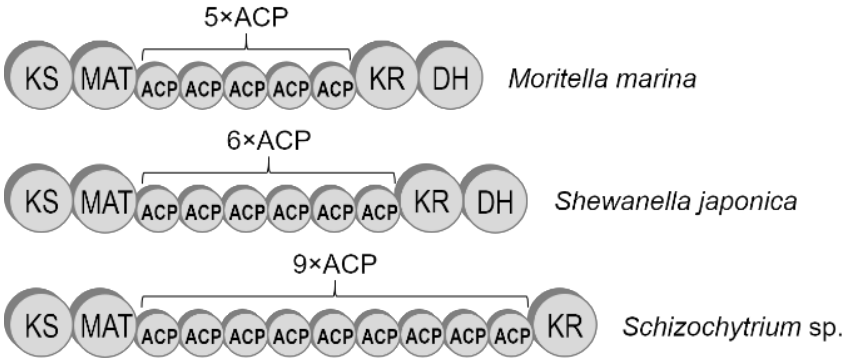

Figure S2. Domain organization of polyunsaturated fatty acid synthases from deep-sea bacteria. ACP: acyl carrier protein, DH: dehydratase, KR: ketoreductase, MAT: malonyl-CoA transferase.

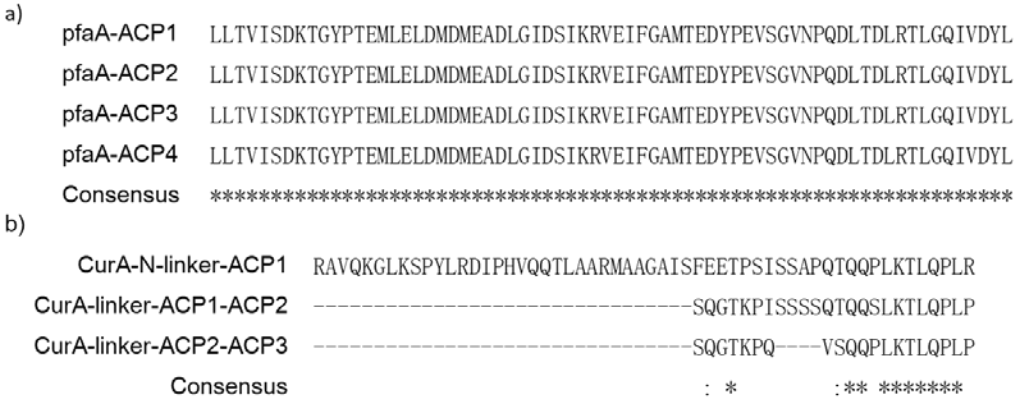

Figure S3. Amino acids sequence alignment of tandem ACP. a) alignment of six ACPs in PfaA from *Aureispira marina*. It shows a 100% identity among the constituent ACPs of the 4×ACP domain. b) alignment of linkers at the N-terminus of ACPs in CurA from *Lyngbya majuscula*. It shows that there are 12 residues are identical between intralinkers of constituent ACPs and linker upstream tandem ACP domain.

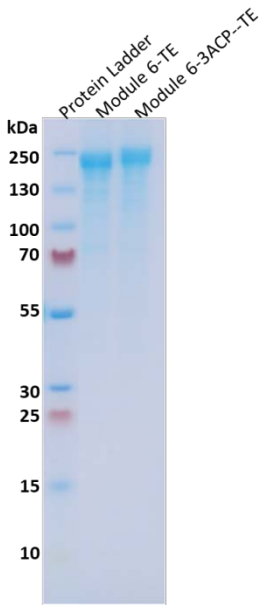

Figure S4. SDS-PAGE of purified DEBS module 6-TE (177 kDa) and module 6-3ACP-TE (198 kDa) proteins.

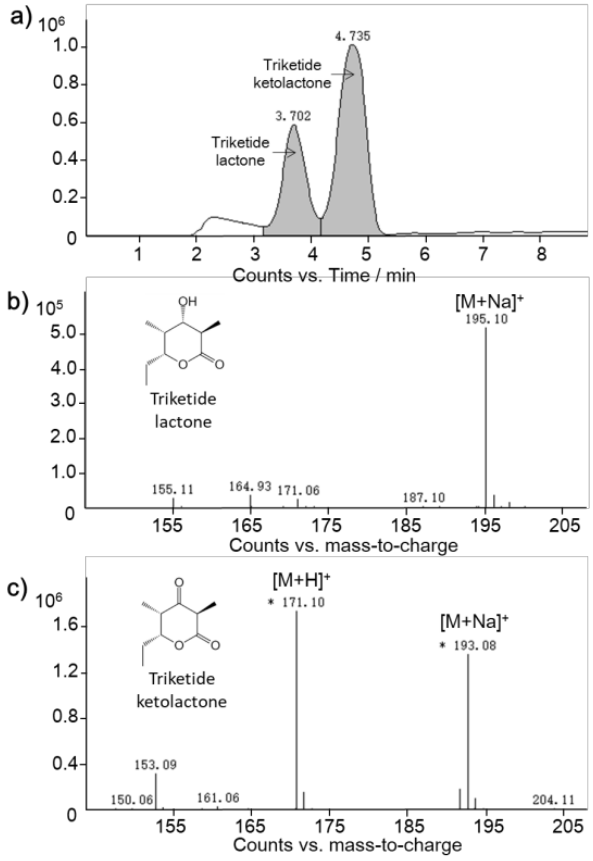

Figure S5. LC-MS analysis of macrolactone products formed by module 6-TE. a) Liquid chromatography elution profile of triketide lactone and triketide ketolactone. b) Mass spectrum for triketide lactone. c) Mass spectrum for triketide ketolactone.

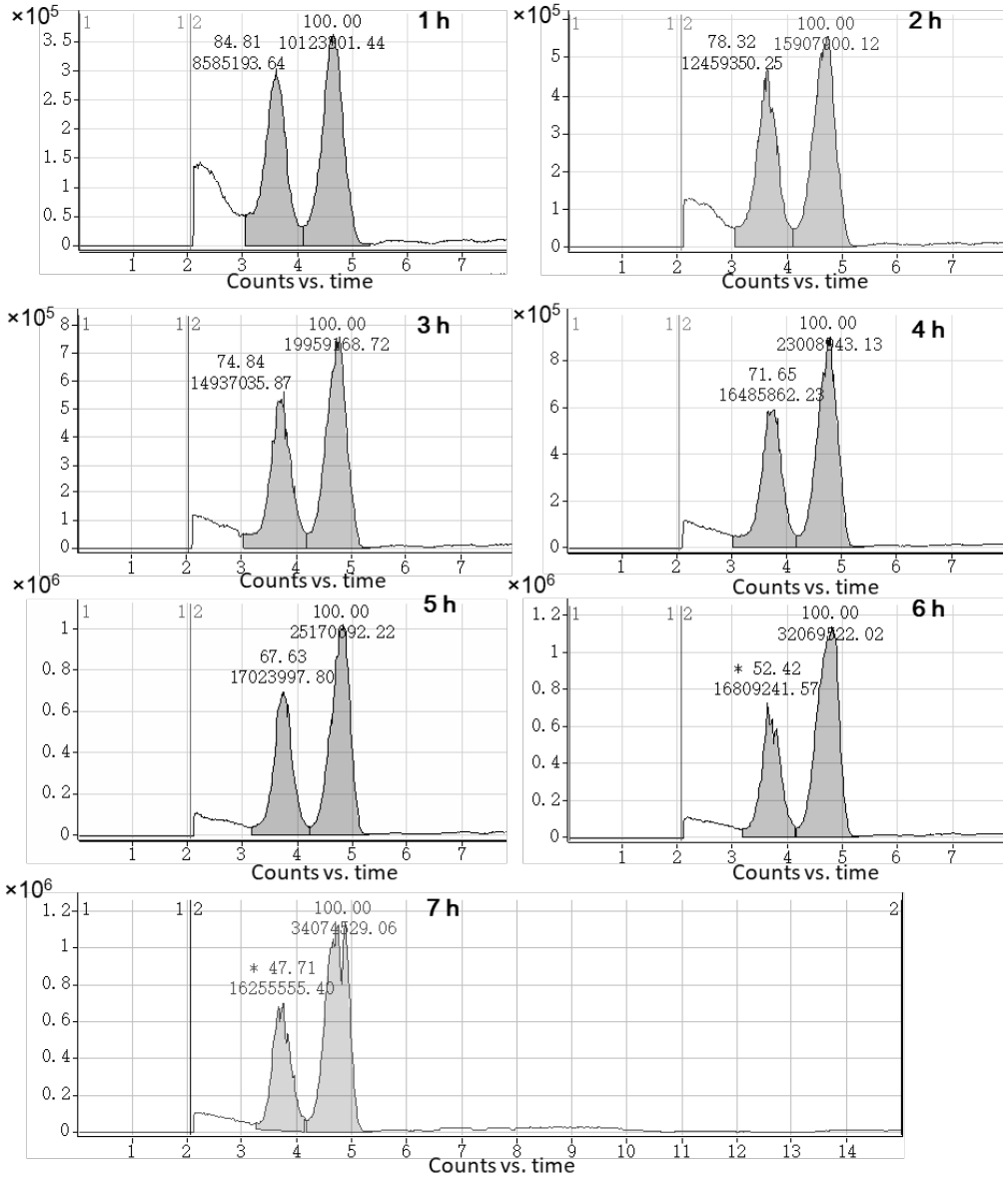

Figure S6. LC-MS/ESI analysis of triketide lactone and ketolactone production by module 6-TE. Agilent MassHunter software was used for data analysis. Triketide lactone eluted at 3.702 min and ketolactone is eluted at 4.735 min. Annotation at the top of the peak is the percent area of that particular peak compared to the largest peak and the below annotation is the actual peak area. Products were located by searching for the theoretical m/z for the $[M+H]^+$- and $[M+Na]^+$- ion. Triketide lactone: $[M+H]^+$=173.10 and $[M+Na]^+$=195.10, triketide ketolactone: $[M+H]^+$=171.10 and $[M+Na]^+$=193.10.

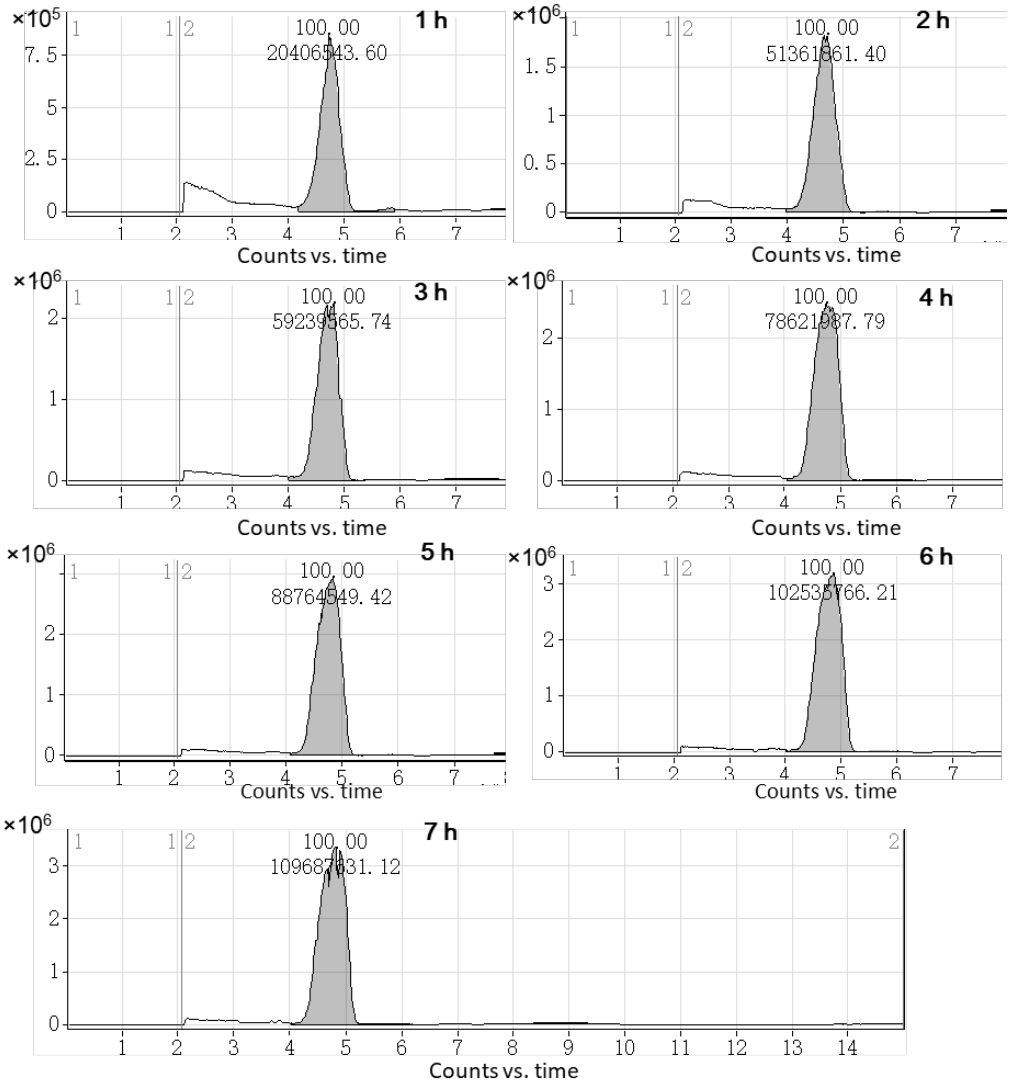

Figure S7. LC-MS/ESI analysis of triketide lactone and ketolactone production by module 6-3ACP-TE. Agilent MassHunter software was used for data analysis. Triketide ketolactone is eluted at 4.735 min. Annotation at the top of the peak is the percent area of that particular peak compared to the largest peak and the below annotation is the actual peak area. Products were located by searching for the theoretical m/z for the $[M+H]^+$- and $[M+Na]^+$- ion. Triketide lactone: $[M+H]^+$=173.10 and $[M+Na]^+$=195.10, triketide ketolactone: $[M+H]^+$=171.10 and $[M+Na]^+$=193.10. No triketide lactone was detected.

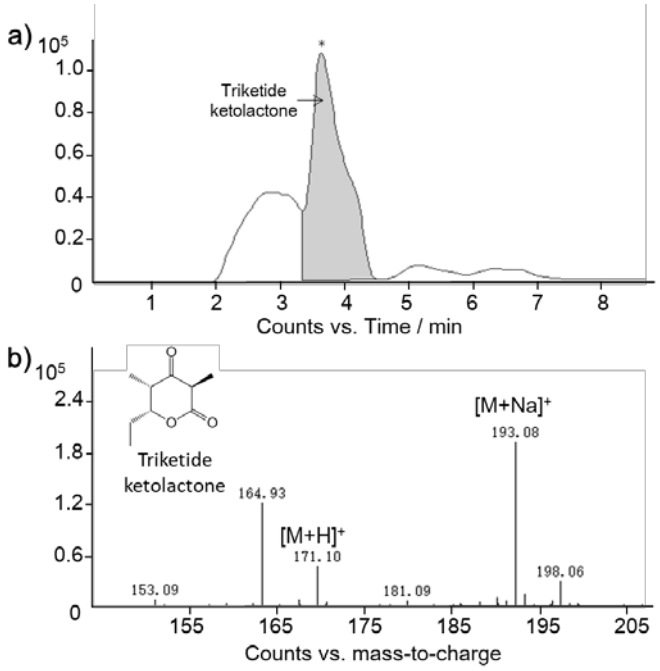

Figure S8. LC-MS analysis of macrolactone products formed by module 6-TE without NADPH. a) Liquid chromatography elution profile of triketide ketolactone. b) Mass spectrum for triketide ketolactone.

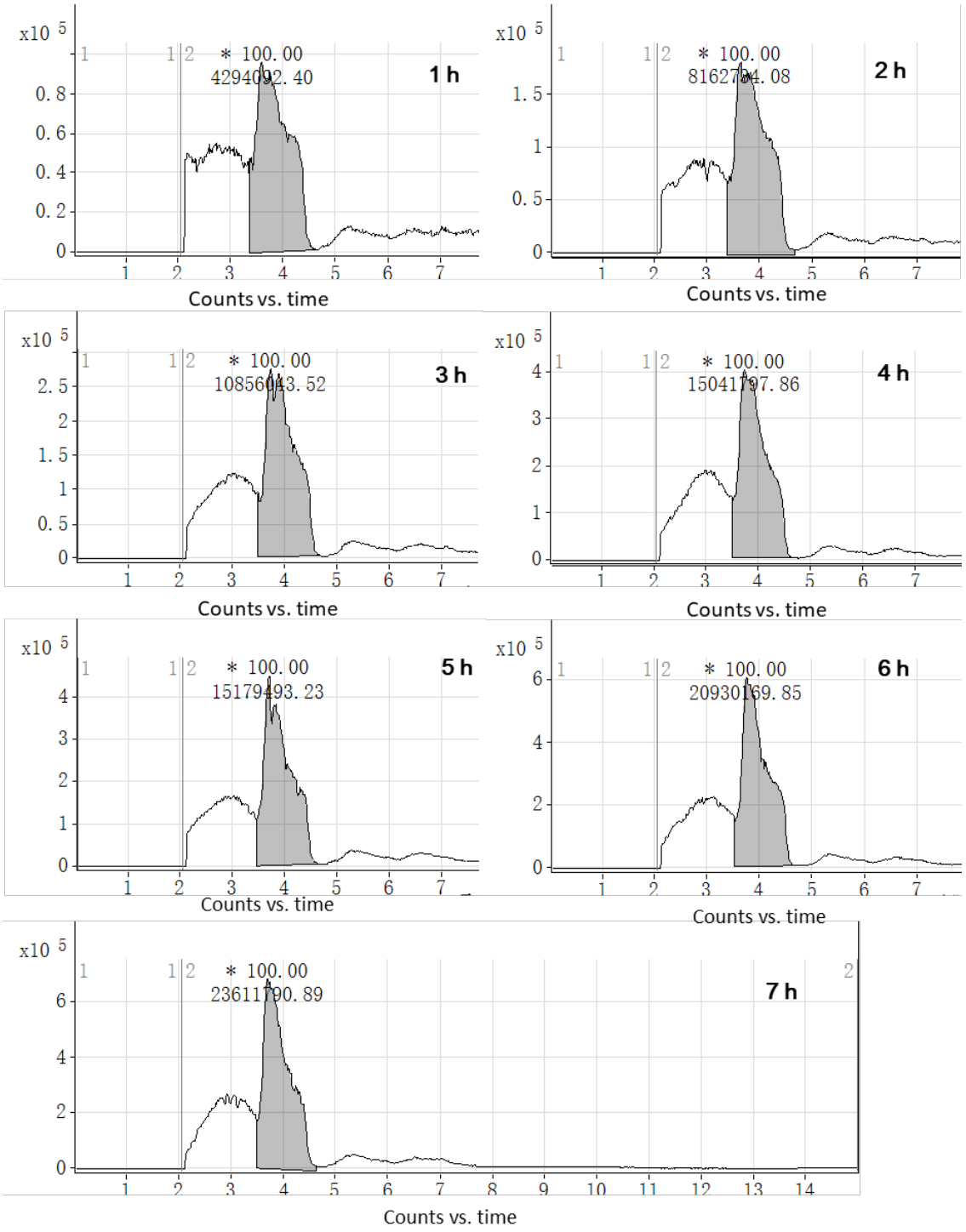

Figure S9. LC-MS/ESI analysis of triketide ketolactone production by module 6-TE without NADPH. Agilent MassHunter software was used for data analysis. Triketide ketolactone is eluted at 3.750 min. Annotation at the top of the peak is the percent area of that particular peak compared to the largest peak and the below annotation is the actual peak area. Products were located by searching for the theoretical m/z for the $[M+H]^+$- and $[M+Na]^+$- ion. Triketide ketolactone: $[M+H]^+$=171.10 and $[M+Na]^+$=193.10.

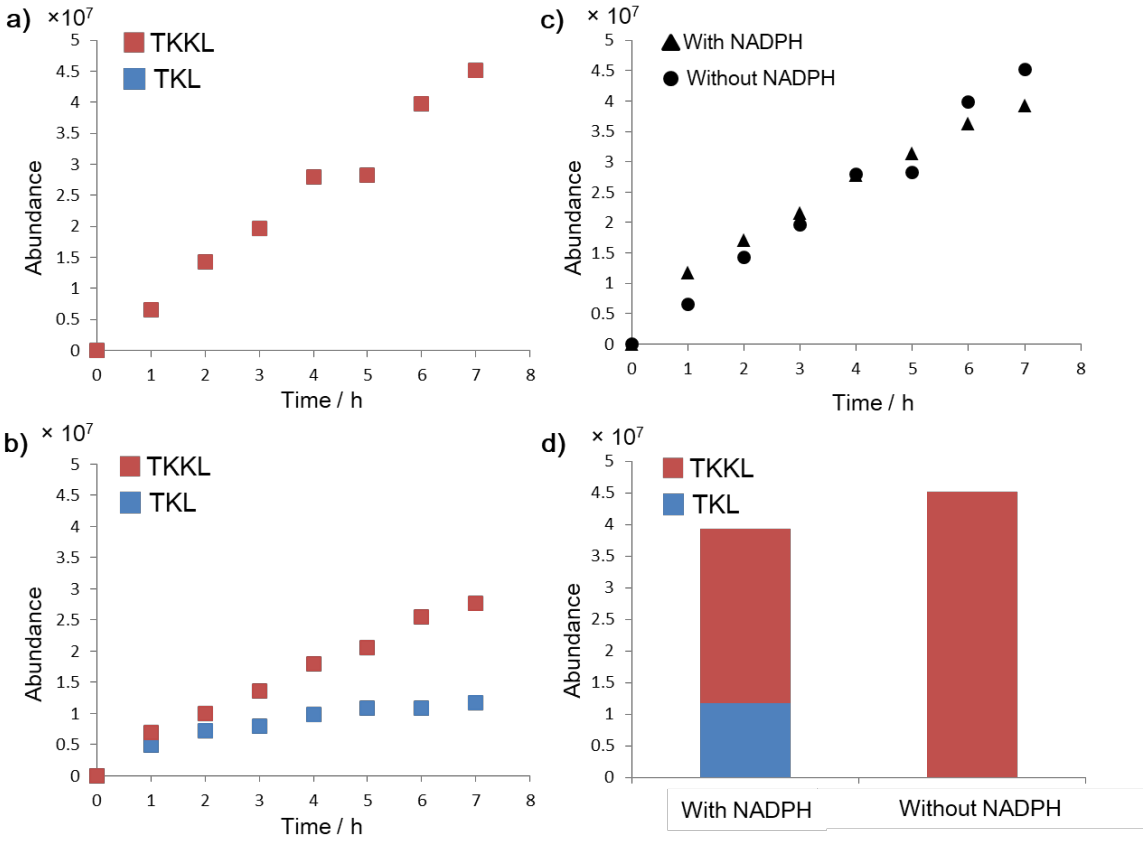

Figure S10. Comparison of the polyketide biosynthesis rate of module 6-TE with and without NADPH. a) Triketide ketolactone formation by module 6-TE without NADPH. b) Triketide lactone and ketolactone formation by module 6-TE with NADPH. c) Comparison of total amount of polyketide products formed by module 6-TE with and without NADPH. d) Total amount of polyketide products formed by module 6-TE with and without NADPH over seven hours.

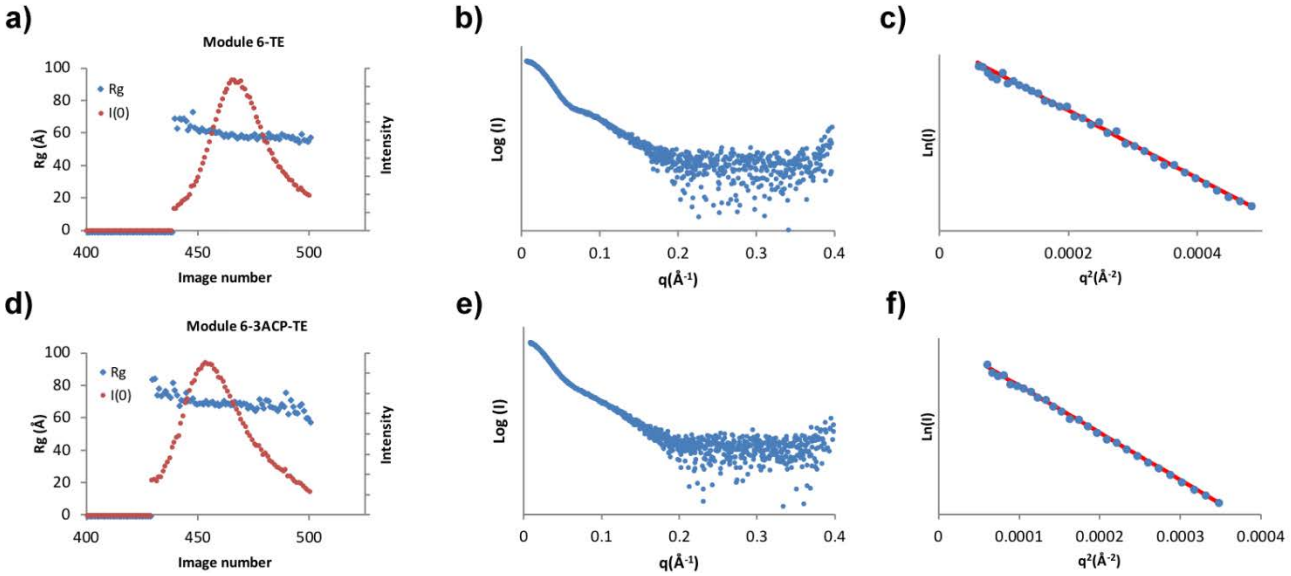

Figure S11. Small-angle X-ray scattering data analysis for module 6-TE (panel a, b, c) and for module 6-3ACP-TE (panel d, e, f). a) and d) Rg and I(0) plots. b) and e) Scattering profile. c) and f) Guinier plot.

References for Supporting Information

1. Mofid, M.R., et al., *Crystallization and preliminary crystallographic studies of Sfp: a phosphopantetheinyl transferase of modular peptide synthetases.* Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 5): p. 1098-100.
2. Wu, N., et al., *Analysis of the Molecular Recognition Features of Individual Modules Derived from the Erythromycin Polyketide Synthase.* Journal of the American Chemical Society, 2000. **122**(20): p. 4847-4852.
3. Martel, A., et al., *An integrated high-throughput data acquisition system for biological solution X-ray scattering studies.* J Synchrotron Radiat, 2012. **19**(Pt 3): p. 431-4.
4. Huang, T.C., et al., *X-ray powder diffraction analysis of silver behenate, a possible low-angle diffraction standard.* Journal of Applied Crystallography, 1993. **26**(2): p. 180-184.
5. McPhillips, T.M., et al., *Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines.* J Synchrotron Radiat, 2002. **9**(Pt 6): p. 401-6.
6. Petoukhov, M.V., et al., *ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis.* Journal of Applied Crystallography, 2007. **40**(s1): p. s223-s228.