Supporting Information

Unusual Activities of the Thioesterase Domain for the Biosynthesis of the Polycyclic Tetramate Macrolactam HSAF in *Lysobacter enzymogenes* C3

Lili Lou,[†] Haotong Chen,[†] Ronald L Cerny,[†] Yaoyao Li,[‡] Yuemao Shen,[‡] Liangcheng Du*[,†]

[†]Departments of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588, USA, and

[‡]School of Life Science, Shandong University, Jinan 250100, China

General Materials and Methods. Chemicals used in this study were purchased from Fisher Scientific or Sigma. All oligonucleotide primers for PCR were synthesized by Integrated DNA Technologies (IDT, Coralville, IA). Plasmid preparation and DNA extraction were carried out with Qiagen kits (Valencia, CA), and all other DNA manipulations were carried out according to standard methods. The *Escherichia coli* strain XL Blue was used as the host for general plasmid DNA propagation, and cloning vectors were pANT841.

Expression of HSAF-TE, Srf TE and EntF TE. The 732-bp thioesterase (TE) domain of HSAF hybrid PKS/NRPS was amplified by PCR using Pfu DNA polymerase and cosmid COS 8-1 of *Lysobacter enzymogenes* C3 as template. The forward primer was P1, and the reverse primer was P2 (Table S1). The PCR product was digested with *Nco*I and *Bam*HI, and then ligated into pANT841 that has been linearized with the same enzymes to generate pANT841-TE. The plasmid pANT841-TE was sequenced to confirm the fidelity of the TE domain. To express the TE domain, the *Nco*I/*Bam*HI fragment was released from pANT841-TE and ligated into pQE60 at the same sites, to generate pQE60-TE. This expression construct was then introduced

into *E. coli* SG13009 (pREP4), which is expected to produce a protein of 256 amino acid residues. Single colonies were inoculated in LB at 37 ºC supplemented with 50 µg/ml ampicillin and 25 µg/ml kanamycin. When OD600nm reached 0.7, the cultures were cooled for 20 min at 4 ºC. IPTG (0.5 mM) was added to the cultures, which were allowed to grow for additional 14 h at 25 ºC. The cells were harvested and resuspended in 20 mM PBS, pH 7.8, containing 5 mM imidazole, 500 mM NaCl, 1 mg/mL lysozyme, and 0.5% Triton X-100. The mixture was incubated on ice for 30 min, followed by sonication, and centrifugation. The soluble fraction was loaded to a Ni-NTA column (Qiagen), and the C-His_6-tagged TE (28.3 kDa) was purified by using an imidazole step-gradient as instructed by the manufacturer's protocol. The purity of the protein was analyzed by SDS-PAGE, and the fractions containing purified protein was pooled, concentrated, and dialyzed against 50 mM PBS, pH 7.8, containing 250 mM NaCl and 15% glycerol. Finally the protein solution was frozen in liquid nitrogen and stored at -80 ºC until use. To express the two non-PTM TE proteins, Srf TE and EntF TE, we obtained two constructs, Srf TE/pET30a and EntF TE/pET30a, which were generous gifts from Prof. Walsh's group at Harvard Medical School. Srf TE and EntF TE were expressed and purified as described.[1,2]

Site-directed mutagenesis. The active site Ser91 in the conserved motif "GxSxG" and a nearby Ser119 were changed to Ala using the overlap extension PCR method. The primers used in the experiments are summarized in Table S1. Primer pairs P1/P4 and P3/P2 were used in the TE-S91A generation, and primer pairs P1/P6 and P5/P2 were used in the TE-S119A generation. Cosmid COS 8-1 was used as template. For the double mutant TE-S91A/S119A, primer pairs P1/P6 and P5/P2 were used, with the plasmid TE-S91A-pANT841 as the template. All mutated sites were confirmed by DNA sequencing. Additionally, the sequencing data showed that one of the colonies that were supposed to generate the TE-S119A mutant contained a random PCR

mutation that changed Arg71 to Ser. This generated an unexpected double mutant TE-R71S/S119A, which was also expressed and purified and used as a control in the assays.

LC-MS analysis. The peptide fragments resulting from the protease activity of TE were analyzed by LC-MS on a Waters (Micromass) Q-TOF Ultima (Waters; Micromass UK, Beverly, MA, USA) and a Shimadzu HPLC system consisting of a SCL-10A controller with two pumps (LC-10AT and LC-10AD). A Micro-Tech scientific column (reverse phase, 15 cm × 1.000 mm id, P/N mm-15-C4W) was used for separations. The flow rate was 50 µl/min. Solvent A was H_2O containing 0.1% formic acid; solvent B was acetonitrile containing 0.1% formic acid. The gradient was as follows: 0-5 min, 0% B; 5-10 min, 0% B to 20% B gradient; 10-35 min, 20% B to 70% B gradient; 35-38 min, 70% B to 85% B gradient; 38-43 min, 85% B; 43-46 min, 85% B to 0% B gradient; 46-56 min, 0% B. The mass spectrometer was operated in positive ion mode with electrospray ionization. The software Masslynx V3.5 was used in analysis. To measure the mass of the intact proteins, samples were analyzed with a Qstar XL system (Applied Biosystems Inc., Foster City, CA) using a turbo ion spray source probe source. The analysis was performed by loading 100 µL of the protein sample into a 2 mm × 20 mm preconcentration loop filled with perfusion material POROS 10 R2 (PerSeptive Biosystems). The salt was removed from the protein sample by passing 2 mL of 0.25% formic acid through the preconcentration loop. After being desalted, the intact protein was directed to a Micro-Tech Scientific C18 reverse phase column. A Shimadzu (SCL-10A) HPLC system (Shimadzu Scientific Instruments, Inc., Columbia, MD) was used for gradient elution with a flow rate of 50 µL/min at ambient temperature. Analytes of interest were eluted from the column by using gradient elution of 0.3% formic acid in H_2O (solvent A) and 0.3% formic acid in acetonitrile (solvent B). The percentage of solvent B was gradually increased from 10 to 90% with a linear gradient over a time period of

10 min followed by washing and an equilibration step. The total run time for each sample was 20 min. The data were acquired in TOF (time-of-flight) positive ion mode, and the mass range was 800–1000 amu (atomic mass units). Analyst QS version 1.1 was used to process the data. Molecular masses of proteins were generated from several multiply charged peaks using the Bayesian Protein Reconstruct option in Bioanalyst Extensions version 1.1.5.

Circular dichroism Measurements. CD measurements were carried out on a Jasco J-815 (Japan) equipped with a Jasco PTC-423S/15 Peltier temperature controller under a nitrogen flow rate of 10 L/min. All spectra reported in this work had high tension (HT) values on the photomultiplier tube below 600 V. Protein concentration was determined by ultraviolet-visible absorbance at 280 nm using a molar extinction coefficient of 20130 M^{-1} cm^{-1}. The protein concentration was 7 μM in 5 mM PBS buffer at pH 7.8. Prior to each scan, the background was set using the appropriately pH-adjusted solvent (5 mM K_2HPO_4/KH_2PO_4, pH7.8). Sample solution was placed in a 1-cm path-length strain-free quartz cell, and the spectrum was recorded every 0.1 nm from 200 to 260 nm at varying temperatures: 20, 37, 50, 65, 80 and 100 ºC. Each spectrum represents the average of three scans. The scanning speed was 100 nm/min, data pitch was 0.1-nm increments, the bandwidth was 1 nm, Digital Integration Time was 1 s, the accumulation was 3, the scanning resolution was 0.1 nm, and the standard sensitivity was used. For the thermal denaturation curves, samples were heated from 20 ºC to 100 ºC at 1.5 ºC/min. Deconvolution of the spectra was performed using the program of CDpro.

In vitro activity assays. In the assays of the peptide ligase-like activity, a 3 μl purified enzyme (0.2 mg/ml of C-A-PCP-TE, 1.69 mg/ml of TE, 1.68 mg/ml of TE-S91A, 2.2 mg/ml of TE-S119A, 2.14 mg/ml of TE-S91A/S119A, 2.12 mg/ml of TE-R71S/S119A) was mixed with 3 μl BSA (2.5 mg/ml). Each sample was boiled for 0-30 min as specified in the individual figures.

The samples were added with an equal volume (6 μl) of SDS-PAGE loading buffer and boiled for 8 min. After centrifugation at 13200 rpm for 5 min, the supernatant of each of the samples was collected and analyzed by SDS-PAGE (12%). In the same way, fumonisin ACP [3] and lysosome were also tested with HSAF TE. In experiments testing TE's protease-like activity, four different protease inhibitors (from Sigma), PMSF (1.0 mM, 2.5 mM, 25 mM), leupeptin (175 μm, 315 μm,1.75 mM), TPCK (100 μm) , and TLCK (100 μm), were added to the samples prior to the 0-30 min boiling. A control without inhibitor was tested simultaneously.

Sequence alignment and homology modeling. A template search was first performed through the BLAST programs. Two NRPS TE (fengycin TE and surfactin TE) and two PKS TE (DEBS TE and picromycin TE) were chosen to do multiple sequence alignment as they showed the highest sequence similarity with HSAF TE. Sequence alignments were generated using CLUSTALW 2,[4] and secondary structure annotations for HSAF TE were generated using Espript 2.2.[5] Homology modeling for 3D structure of HSAF TE was performed using MODELLER Version 9v7.[6] Alignment file for MODELLER was prepared by CLUSTALW. Molecular visualization program VMD 1.8.6 was used for displaying, animating, and analyzing the structures. Protein 3D structure alignments were done using the DaliLite Pairwise comparison.

Table S1: Primers used in this study

P1 (TE-NcoI-Fr)	5'-GTC ACC ATG GGA AAG ACG GTG GAG GCG ATC AGC-3'
P2 (TE-BamHI-Rv)	5'-TTA GGA TCC GGC GAC ATG GCC CGT CTC CCC-3'
P3 (TE-S91A-Fr)	5'-CTG TTC GGC TAC GCG CTC GGC GGC-3'
P4 (TE-S91A-Rv)	5'-GTT GCC GCC GAG CGC GTA GCC GAA CAG -3'
P5 (TE-S119A-Fr)	5'-GTG GTC ATC ATG GAT GCC TAC CGC-3'
P6 (TE-S119A-Rv)	5'-TTC CGG GAT GCG GTA GGC ATC CAT GAT-3'

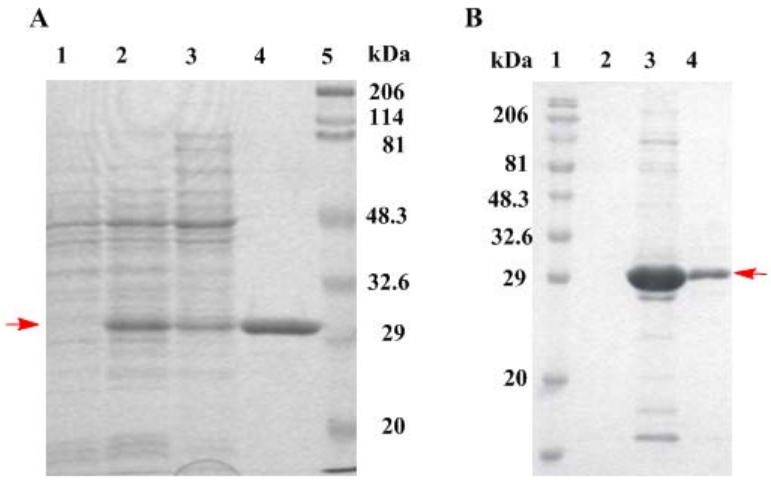


Figure S1. Expression and purification of the TE domain. (A) The wild type TE. Lane-1, total proteins before IPTG induction; lane-2, total proteins after IPTG induction; lane-3, soluble proteins; lane-4, purified TE-domain; lane-5, markers. (B) The TE-S91A mutant. Lane-1, markers; lane-2-4, eluted fractions from a Ni-NTA column.

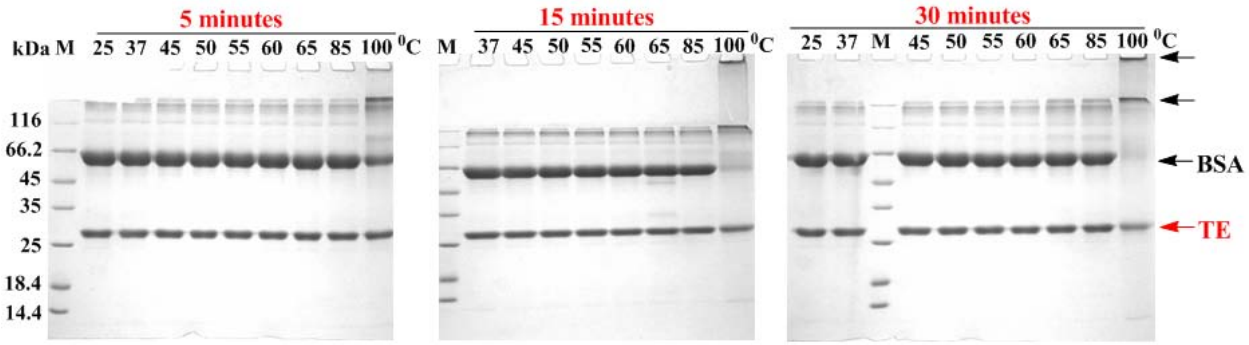
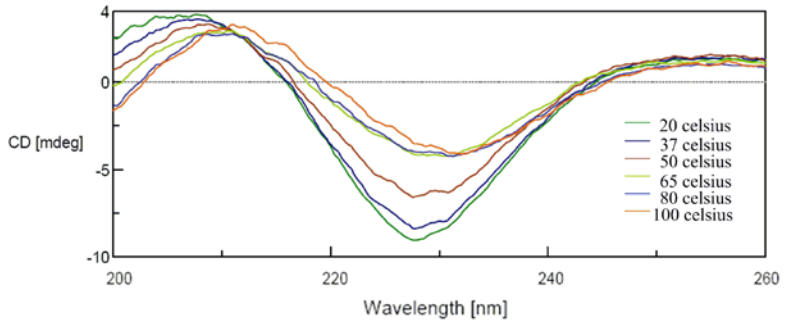


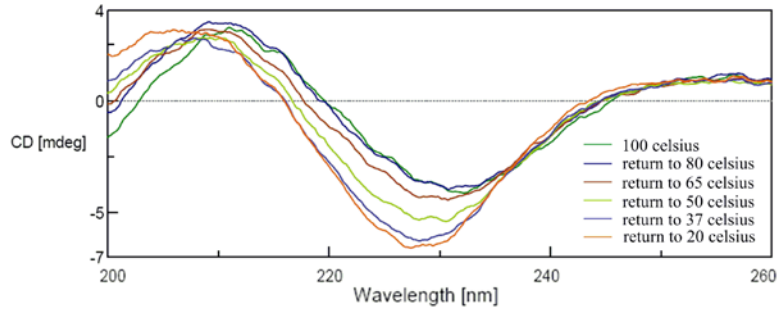
Figure S2. SDS-PAGE of the TE domain incubated with BSA and treated at different temperatures for different periods of time.

A



Temp (°C)	helix	sheet	Turn	Unrd
20	0.183	0.302	0.207	0.309
37	0.156	0.317	0.216	0.311
50	0.091	0.363	0.212	0.334
65	0.064	0.434	0.214	0.288
80	0.064	0.433	0.239	0.264
100	0.07	0.432	0.233	0.265

B



Temp (°C)	helix	sheet	Turn	Unrd
100	0.07	0.432	0.233	0.265
80	0.069	0.423	0.227	0.282
65	0.068	0.429	0.217	0.285
50	0.077	0.377	0.227	0.318
37	0.097	0.357	0.216	0.33
20	0.098	0.349	0.213	0.338

Figure S3. Far-UV circular dichroism spectra of HSAF TE and secondary structure analysis. CD measurements were carried out on a Jasco J-815 equipped with a Jasco PTC-423S/15 Peltier temperature controller. The secondary structure analysis was done by software CDPro.

A. CD spectra of TE (in pH 7.8 PBS buffer) with increasing temperatures (from 20 ºC to 100 ºC) and the predicted secondary structure elements. Note: Unrd, unordered structure.

B. CD spectra of TE (in pH 7.8 PBS buffer) with decreasing temperatures (from 100 ºC to 20 ºC) and the predicted secondary structure elements.

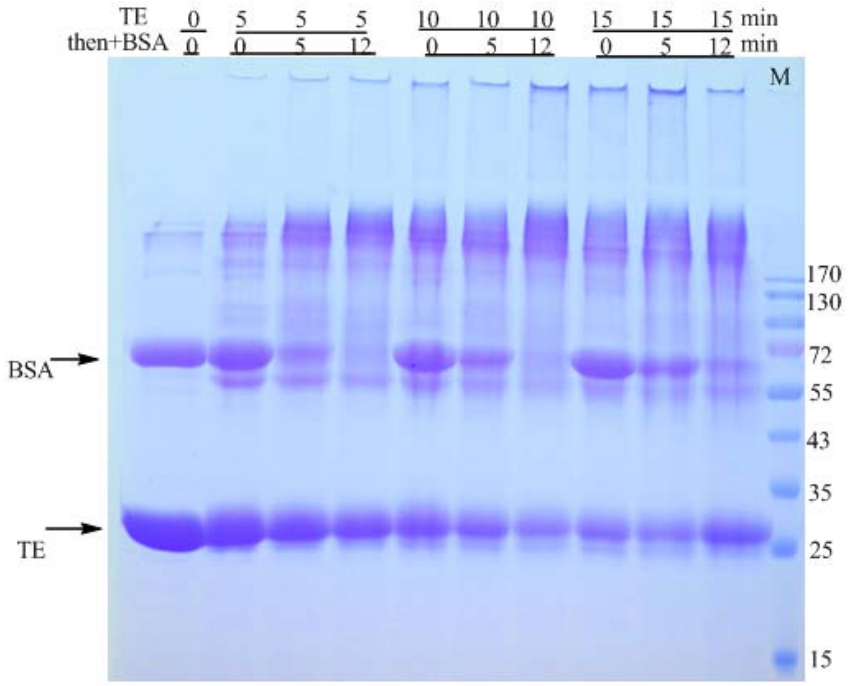
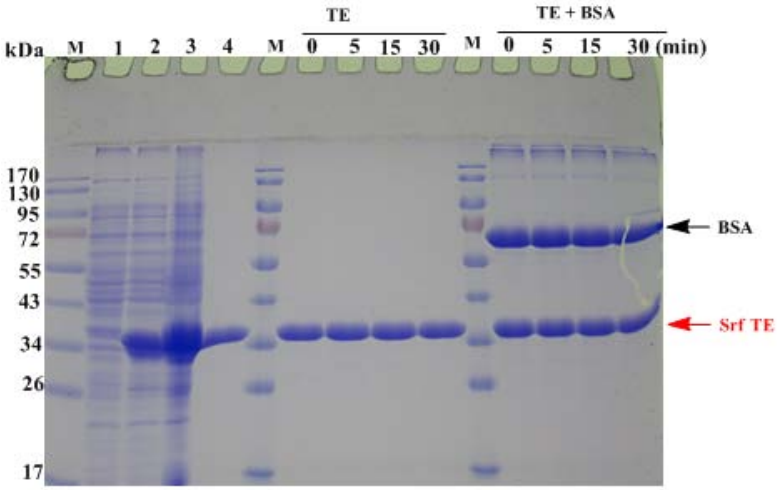


Figure S4. SDS-PAGE of the TE domain after it was pre-boiled for 5-15 min and then co-heated with BSA for 0, 5, or 12 min.

A



B

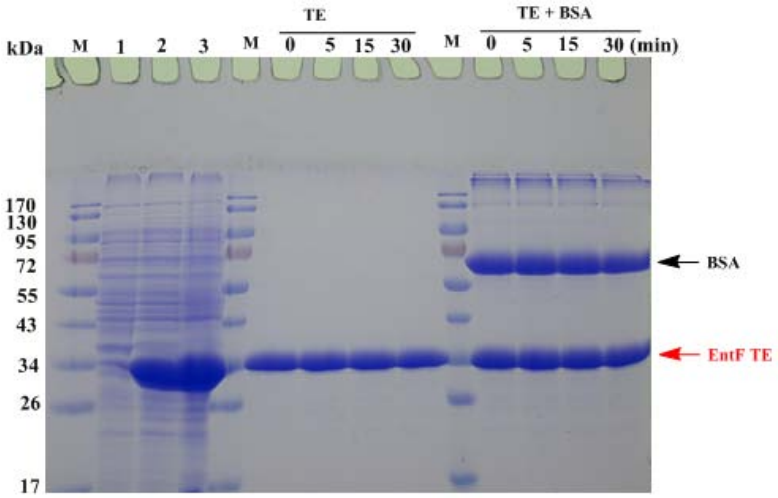


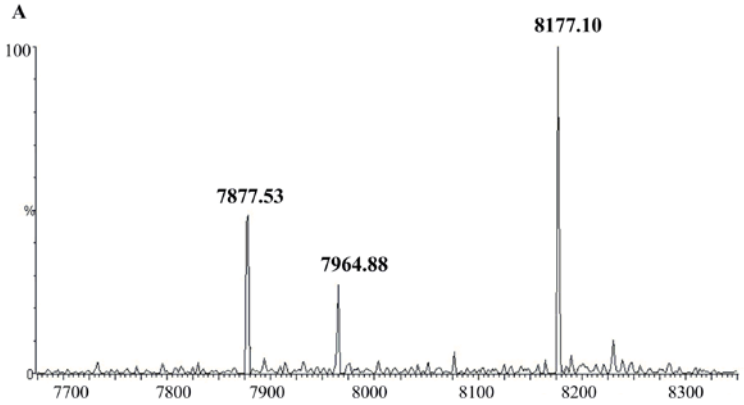
Figure S5. Expression, purification and activity test of two non-PTM TE domains.

A. Srf TE. Lane-1, total proteins before IPTG induction; lane-2, total proteins after IPTG induction; lane-3, soluble proteins; lane-4, purified Srf TE; M, size markers; the next four lanes, Srf TE alone boiled for 0, 5, 15, 30 min, respectively; the last four lanes, Srf TE with BSA boiled for 0, 5, 15, 30 min, respectively.

B. EntF TE. Lane-1, total proteins before IPTG induction; lane-2, total proteins after IPTG induction; lane-3, soluble proteins; M, size markers; the next four lanes, EntF TE alone boiled

for 0, 5, 15, 30 min, respectively; the last four lanes, EntF TE with BSA boiled for 0, 5, 15, 30 min, respectively.

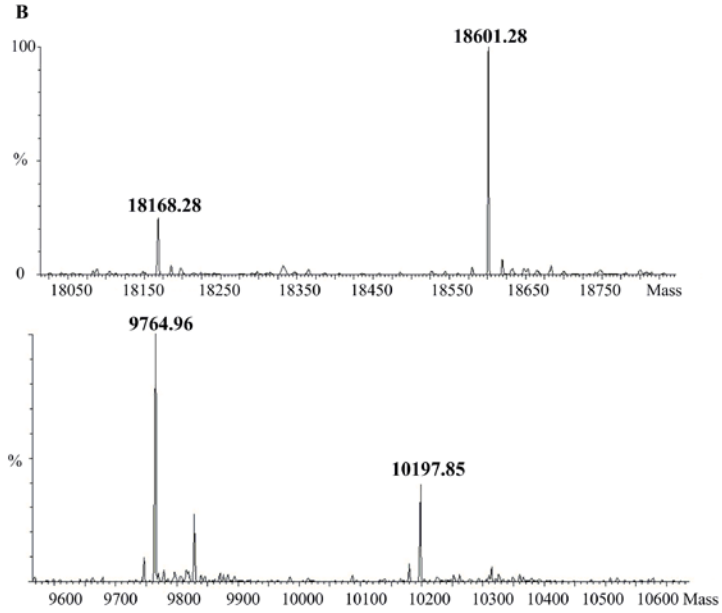
A




```
  1  GKTVEAISL GRVAGAQPYL RFNAGRGHEQ TIFCFPPAGG HGLVYRQLAV
 51 HLPEYEFVSF NYLLGDDKVA RYADLIEGIH AEGHCTLFGY SLGGNLAFEV
101 AKELERRGRE VPNVVIMDSY RIPESFELGN EHFEAFEHEL TEHLRKHTGS
151 DLLAHDTLEQ ARDYIRFCSQ TPNLGTIAAP VS↑VIS↑D↑EDKL VFYGTGQRGT
                                    8177 7965 7878

201 WHGSSLTRSQ VFRGYGKHAD MLDADYIEKN AALARGILVG ETGHVAGSRS
251 HHHHHH
```

B




```
  1  GKTVEAISL GRVAGAQPYL RFNAGRGHEQ TIFCFPPAGG HGLVYRQLAV
 51 HLPEYEFVSF NYLLGDDKVA RYADLIEGIH AEGHCTLFGY SLGGNLAFEV
101 AKELERRGRE VPNVVIMDSY RIPESFELGN EHFEAFEHEL TEHLRKHTGS
                  18168    18601
151 DLLAHDTLEQ ARD↓YIR↓↑FCSQ TPNLGTIAAP VSVISDEDKL VFYGTGQRGT
                  10198    9765

201 WHGSSLTRSQ VFRGYGKHAD MLDADYIEKN AALARGILVG ETGHVAGSRS
251 HHHHHH
```

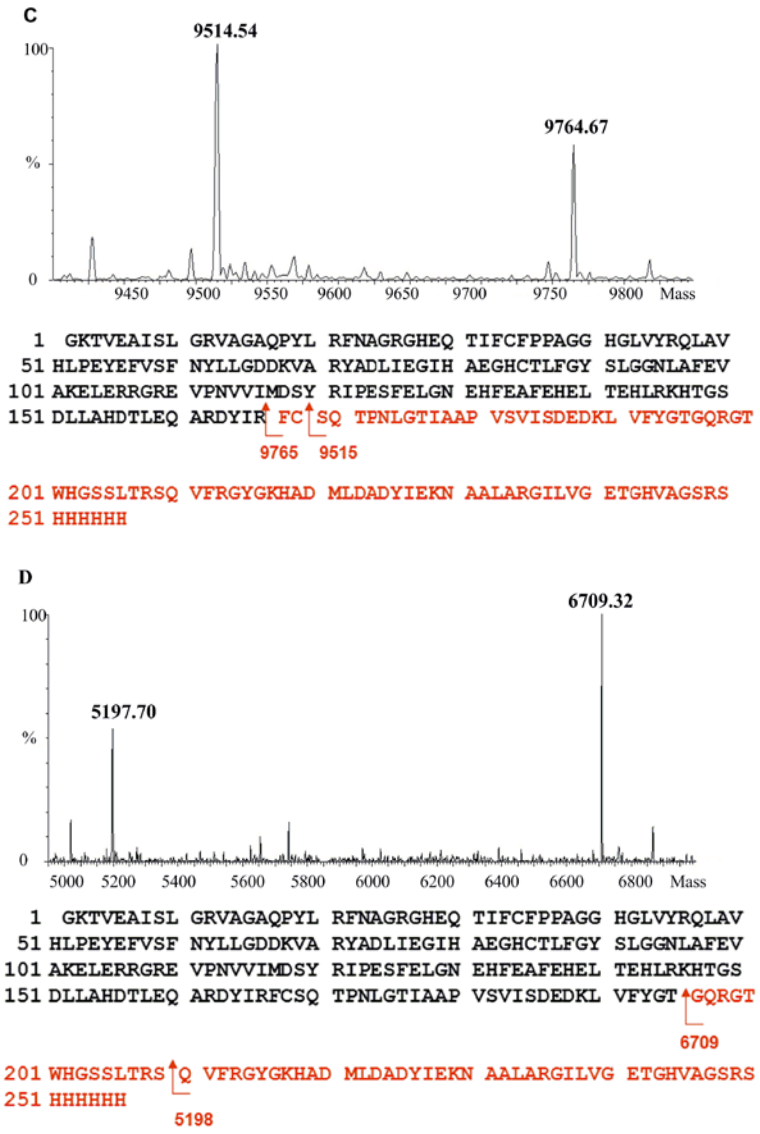


Figure S6. Peptide fragments identified by LC-MS resulted from self-cleavage of HSAF TE. The cleavage sites and corresponding mass are indicated on the amino acid sequence under each of the spectra.

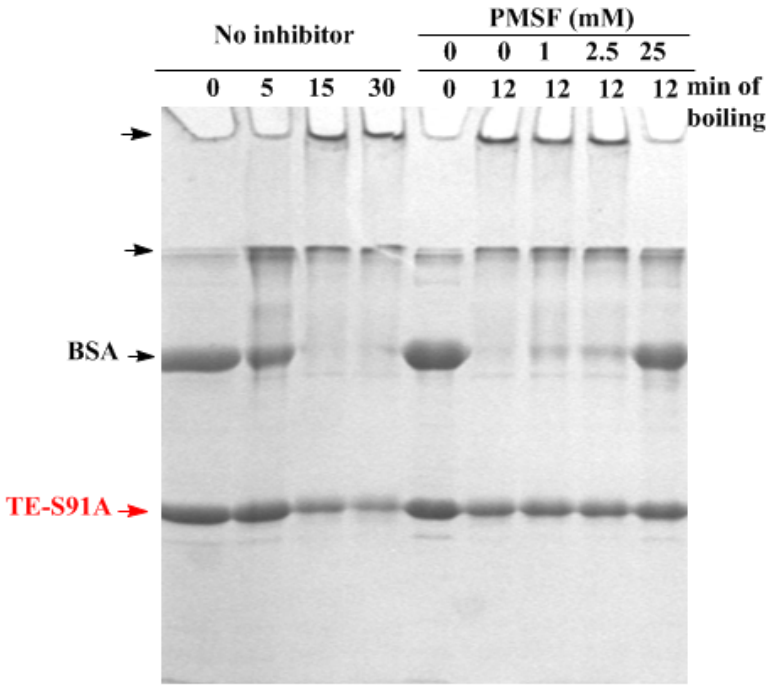


Figure S7. SDS-PAGE of TE-S91A mutant incubated with BSA with or without the protease inhibitor PMSF. Lane 1-4, TE-S91A + BSA, no inhibitor, boiled for 0, 5, 15, 30 min, respectively; lane 5, TE-S91A + BSA, no boiling; lane 6-9, TE-S91A + BSA + PMSF boiled for 12 min, with 0, 1, 2.5, 25 mM PMSF, respectively.

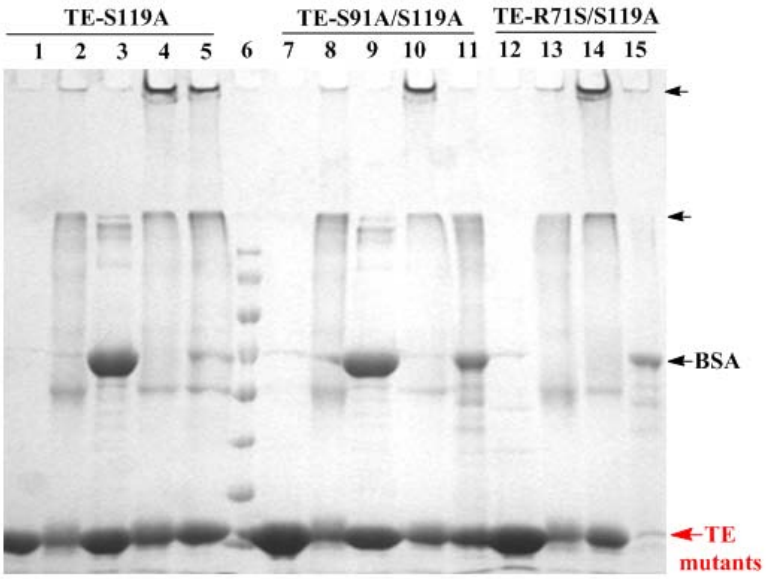


Figure S8. SDS-PAGE of TE-S119A mutant, TE-S91A/S119A double mutant, and TE-R71S/S119A double mutant incubated with BSA with or without the protease inhibitor PMSF. Lane 1, TE-S119A, no boiling; lane 2, TE-S119A, boiled 15 min; lane 3, TE-S119A + BSA, no boiling; lane 4, TE-S119A + BSA, boiled 15 min; lane 5, TE-S119A + BSA + 100 mM PMSF, boiled 15 min; lane 6, markers; lane 7, TE-S91A/S119A, no boiling; lane 8, TE-S91A/S119A, boiled 15 min; lane 9, TE-S91A/S119A + BSA, no boiling; lane 10, TE-S91A/S119A + BSA, boiled 15 min; lane 11, TE-S91A/S119A + BSA + 100 mM PMSF, boiled 15 min; lane 12, TE-R71S/S119A, no boiling; lane 13, TE-R71S/S119A, boiled 15 min; lane 14, TE-R71S/S119A + BSA, boiled 15 min; lane 15, TE-R71S/S119A + BSA + 100 mM PMSF, boiled 15 min.

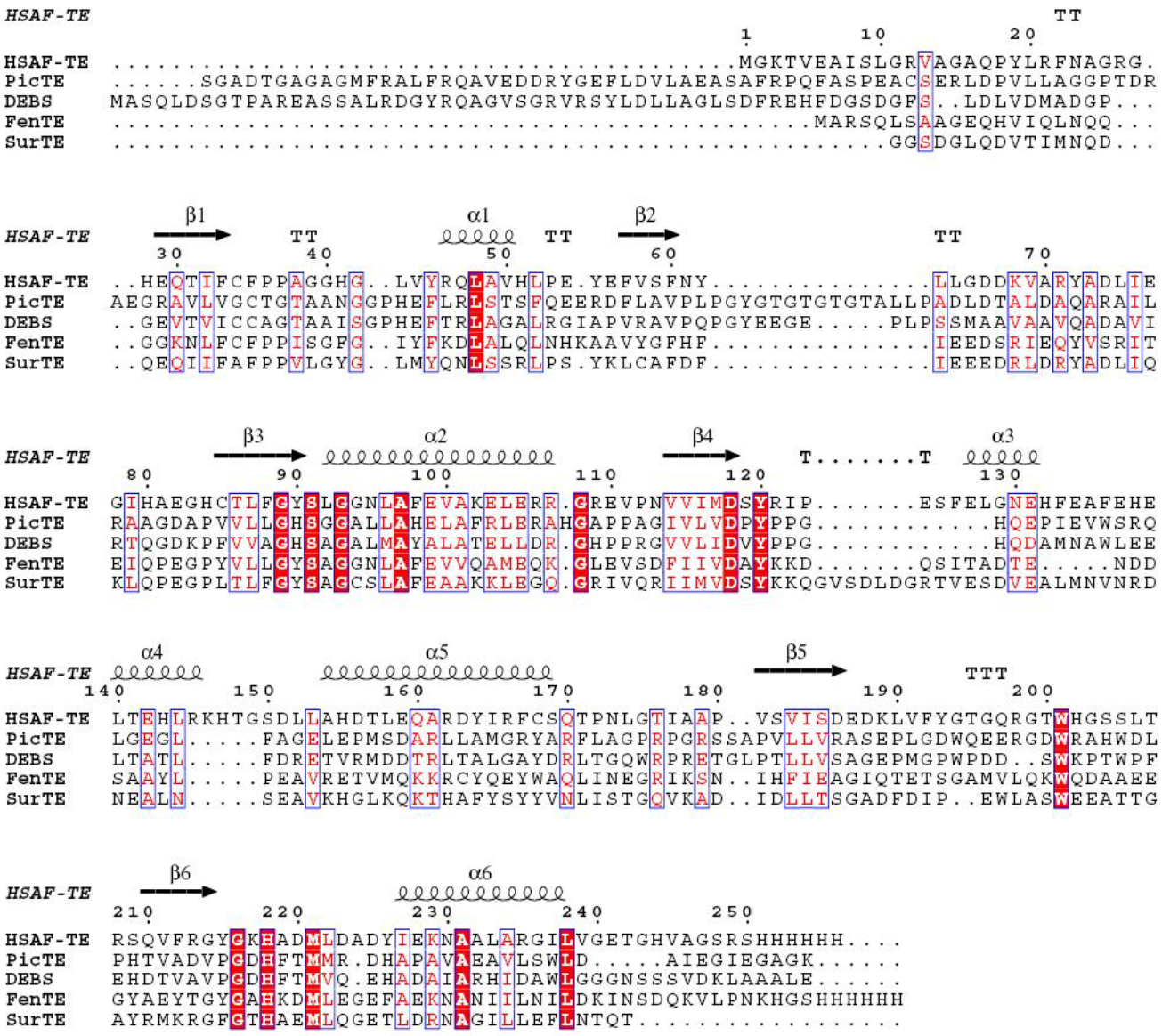


Figure S9. Primary sequence alignment of several TEs and secondary structure prediction of HSAF TE. PicTE, picromycin PKS TE;[7] DEBS, 6-deoxyerythronolide B synthase TE;[8] FenTE, fengycin NRPS TE;[9] SurTE, surfactin NRPS TE.[10] Invariant residues are in white with a red background, other conserved sites are in red. Espript 2.2 was used to generate secondary structure annotations for HSAF TE.

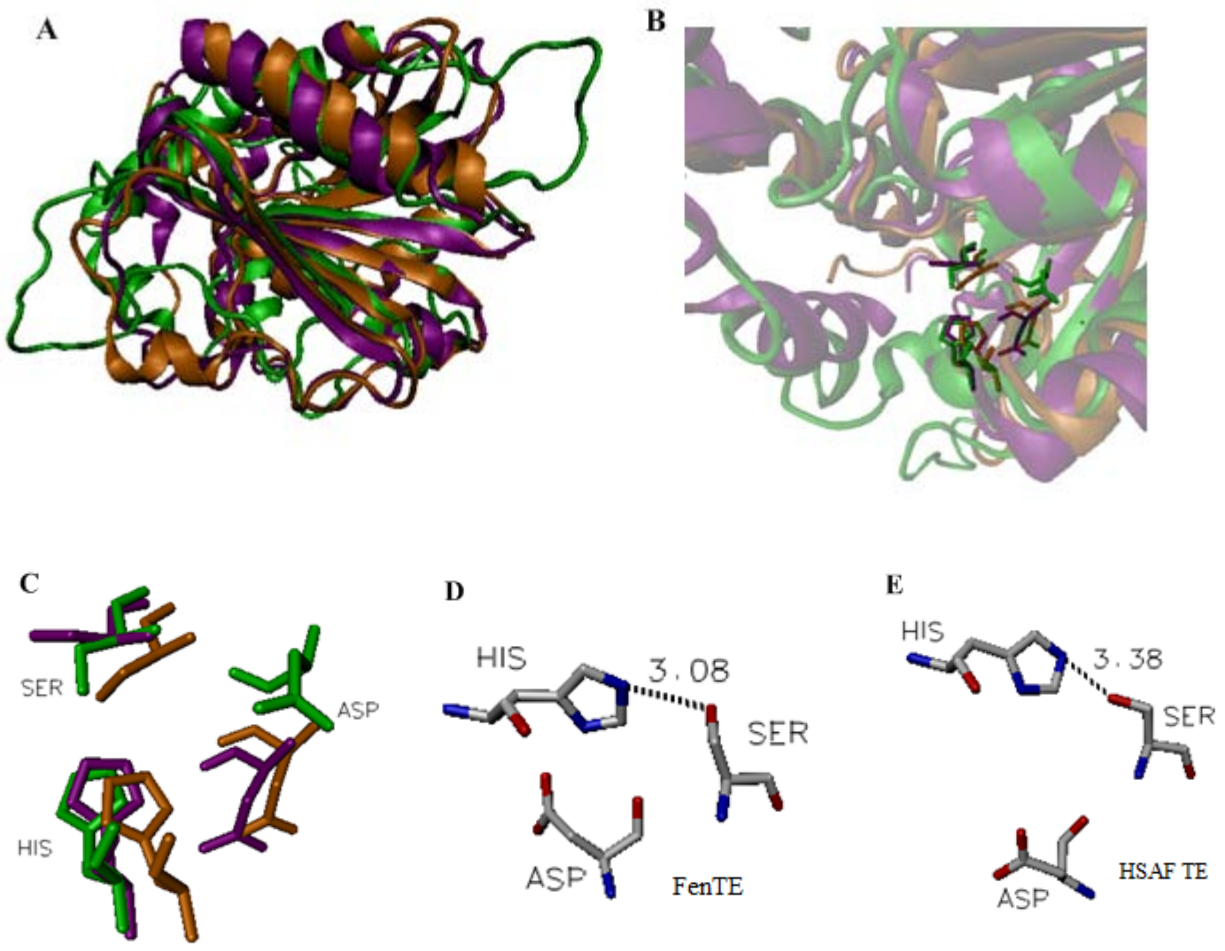


Figure S10. The structure of HSAF TE derived from homology modeling on known TE structures. A. The three-dimensional structure superimposition of TEs. Fengycin TE (orange, PDB entry 2CB9), surfactin TE (purple, PDB entry 1jmkc), and HSAF TE (green). B. The well conserved catalytic triad, shown in ball-and-stick, within the substrate binding pocket of the TE structures. C. Superimposition of the active site to show the deviation of HSAF TE's Asp from the known structures. FenTE (orange), SrfTE (purple), HSAF TE (green). D-E. Active site comparison between fengycin TE and HSAF TE. Note that the similar H-bond distance between the hydroxyl of serine and the NE2 atom of histidine (3.08Å for FenTE and 3.38Å for HSAF TE).

References

(1) Tseng, C. C., Bruner, S. D., Kohli, R. M., Marahiel, M. A., Walsh, C. T., and Sieber, S. A. (2002) *Biochem. 41*, 13350-13359.

(2) Frueh, D. P., Arthanari, H., Koglin, A., Vosburg, D. A., Bennett, A. E., Walsh, C. T., Wagner, G. (2008) *Nature 454*, 903-906.

(3) Gerber, R., Lou, L., and Du, L. (2009) *J. Am. Chem. Soc. 131*, 3148-3149.

(4) Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., Higgins, D. G. (2007) *Bioinformatics 23*, 2947-2948.

(5) Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) *Bioinformatics 15*, 305-308.

(6) Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., Karplus, M. (1995) *Proteins 23*, 318-326.

(7) Tsai, S. C., Lu, H., Cane, D. E., Khosla, C., and Stroud, R. M. (2002) *Biochem. 41*, 12598-12606.

(8) Tsai, S. C., Miercke, L. J., Krucinski, J., Gokhale, R., Chen, J. C., Foster, P. G., Cane, D. E., Khosla, C., and Stroud, R. M. (2001) *Proc. Natl. Acad. Sci. U S A 98*, 14808-14813.

(9) Samel, S. A., Wagner, B., Marahiel, M. A., and Essen, L. O. (2006) *J. Mol. Biol. 359*, 876-889.

(10) Bruner, S. D., Weber, T., Kohli, R. M., Schwarzer, D., Marahiel, M. A., Walsh, C. T., and Stubbs, M. T. (2002) *Structure 10*, 301-310.