SUPPORTING INFORMATION

Structural and Kinetic Analysis of the Unnatural Fusion Protein 4-Coumaroyl-CoA Ligase::Stilbene Synthase

Yechun Wang, Hankuil Yi, Melissa Wang, Oliver Yu,* and Joseph M. Jez

1. EXPERIMENTAL PROCEDURES

Chemicals. Standard chemicals, including 4-coumaric acid, malonyl-CoA, CoA, and resveratrol were purchased from Sigma-Aldrich. Acetonitrile was from MP Biomedicals and all other solvents for HPLC and Orbitrap ESI-MS/MS were from Fisher Scientific. All enzymes were from New England Biolabs, unless specified otherwise.

Gene isolation and expression vector construction. Oligonucleotides were designed for PCR amplification of At4CL1[1], as follows: 4CL1F, 5'-dGCTCTAGA**ATG**GCGCCACAAGAACAAGCAGTTTCTCAG-3' (XbaI site underlined) and 4CL1R: 5'-dCCCTCGAG**TCA**CAATCCATTTGCTAGTTTTGCCCTCAGA-3' (XhoI site underlined). For amplification of VvSTS[2], oligonucleotide primers were designed as follows: STSF: 5'-dTAAGGATCC**ATG**GCTTCAGTCGAGGAATTTAGAAACGCTC-3' (XbaI site underlined) and STSR: 5'-dCCCTCGAGTTAATTTGTAACCATAGGAATGCTATGTAGC-3' (XhoI site underlined). To generate bacterial expression constructs, the vector pESC-TRP-4CL::STS[3] was used as a template to PCR amplify At4CL1, VvSTS, and 4CL::STS for sub-cloning into pET-28a(+).

Expression and purification of At4CL1, VvSTS and 4CL::STS. Each expression construct was transformed into *Escherichia coli* Rosetta(DE3) grown in LB media containing 50 μg mL^{-1} kanamycin and 35 μg mL^{-1} chloramphenicol at 37 °C until A_{600nm} ~0.8-1.0. Protein expression was induced by addition of 1 mM isopropyl-®-D-thiogalactopyranoside and the culture grown at 16 °C for 20 hr. Cells were harvested by centrifugation (3,000 x g; 10 min; 4 °C) with the cell pellets either used immediately or stored at -80 °C. Pellets were re-suspended in lysis buffer (50 mM Tris, pH 8.0, 20 mM imidazole, 500 mM NaCl, 10% glycerol, and 1% Tween-20). Cells were disrupted by sonication (10 x 30 s; 4 °C). Cell debris was clarified by centrifugation (40,000 x g; 30 min). Supernatant was added to a pre-equilibrated (lysis buffer) Ni^{2+}-nitrilotriacetic acid (Qiagen) affinity column. After loading of protein sample, the column was washed (50 mM Tris, pH 8.0, 20 mM imidazole, 500 mM NaCl, and 10% glycerol) to remove unbound contaminant proteins. His-tagged proteins were eluted (50 mM Tris, pH 8.0, 250 mM imidazole, 500 mM NaCl, and 10% glycerol). The eluted 4CL and 4CL::STS were then dialyzed in 20 mM Tris, pH 8.0, 100 mM NaCl buffer; eluted STS was dialyzed in 20 mM HEPES, pH 8.0, 5 mM EDTA and 100 mM NaCl buffer. All proteins were purified by size-exclusion chromatography on a HiLoad 16/60 Superdex-200 FPLC column in their corresponding buffers. For crystallization, 4CL::STS was further purified using a second round of size-exclusion chromatography (20 mM Tris, pH8, 50 mM NaCl). Protein concentrations were determined using the BioRad protein assay with BSA as a standard.

Enzymatic synthesis of *p*-coumaroyl-CoA by At4CL1. At4CL1 was used to enzymatically prepare *p*-coumaroyl-CoA[4]. Reactions (10 mL; 25 °C) were performed in 100 mM Tris, pH 8, 2 mM dithiothreitol (DTT), 10 mM $MgCl_2$, 5 mM ATP, 200 μM *p*-coumaric acid, and 1 mM CoA with 0.5 mg At4CL1. After 30 min, the reaction was terminated by addition of acetic acid, mixed, and centrifuged (20,000 x g; 30 min). Purification of *p*-coumaroyl-CoA was performed using an Agilent 1100 HPLC (Agilent) with a Gemini-NX C18 column (Phenomenex, 5 μm, 150 mm x3 mm). The mobile phase flow was set to 0.6 mL min^{-1} with a

binary gradient elution using solvent A (100% acetonitrile) and solvent B (0.5% TFA in water). The gradient was as follows: 0-5 min, 10% A; 5-25 min, 10-30% A; 25-40 min, 30% A; 40-40.1 min, 30-100% A; 40.1-43 min, 100% A; 43-43.1 min, 100-10% A; 43.1-46 min, 10% A. A diode array detector was used to record A_{333nm} and the *p*-coumaroyl-CoA peak collected in 2 mL fractions, which were then pooled and lyophilized. To confirm the product, a sample was dissolved in 1% formic acid in methanol to 1 mM and analyzed on an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) by direct infusion. The LTQ-Orbitrap Velos was operated in positive ion mode. The scan mode is FTMS + pNSI full ms from 500 to 1200. The resolution was 30k and the ion injection times were set to 9.094 ms. The theoretical $[M-H]^+$ of *p*-coumaroyl-CoA is 914.15 with single charge (H^+) and the m/z of the main peak on the spectrum resulting from enzymatic synthesis is 914.16 (Figure S4). The extinction coefficient used for *p*-coumaroyl-CoA quantification was ε_{333}=21 mM cm^{-1} [4].

Enzyme assays and steady-state kinetic analysis. Assays for 4CL and STS activity of At4CL1, VvSTS, and the 4CL::STS fusion protein were performed using a standard buffer condition: 100 mM Tris, pH 8, 2 mM DTT, and 10 mM $MgCl_2$. For 4CL activity, 5 mM ATP, 200 μM *p*-coumaric acid, and 2 mM CoA were used in the assay described by Knobloch & Hahlbrock[5]. For STS activity, 400 μM *p*-coumaroyl-CoA and 500 μM *p*-malonyl-CoA were used as previously described[6]. The reaction product was analyzed using a 4000 QTRAP LC-MS/MS system (Applied Biosystems) with multiple reaction monitoring (MRM) to obtain specific MS/MS ions of resveratrol[7]. The electrospray ion source of the mass spectrometer was operated at 550°C in negative ion mode. MRM-selected ions for resveratrol was 227/143[8]. The quantification was performed based on standard curves constructed using pure resveratrol. Kinetic parameters were determined for At4CL1, VvSTS, and 4CL::STS using the above assays with varied substrate concentrations and the resulting data fit to $v = (k_{cat}[S])/(K_m + [S])$ in GraphPad Prism.

Crystallization and structure determination of the 4CL::STS fusion protein. Crystals of 4CL::STS were produced by the vapor diffusion method at 4 °C. Protein (10 mg mL^{-1}) was mixed with crystallization buffer (18.5% PEG-3350, 0.24 M potassium citrate, and 10 mM sarcosine) in a 1:1 ratio of a 4 μL drop over 0.7 mL of crystallization buffer. Before data collection at 100 K, crystals were soaked in mother liquor containing 30% glycerol. Collection, integration, and reduction of diffraction data were performed at beamline 19-ID of the Advanced Photon Source Argonne National Laboratory using SBC COLLECT and HKL3000[9]. The structure of 4CL::STS was solved by molecular replacement performed with PHASER[10]. For molecular replacement, two independent search ensembles were used, as follows: poplar 4CL (PDB: 3A9U)[11] and peanut STS (PDB: 1Z1E)[12]. Each ensemble was searched for in sequence, which lead to placement of each molecule with a rotation Z-score of 26.4 and a translation Z-score of 68.5 with an initial FOM = 0.70. Model building was performed in COOT[13] and all refinements (resolution range: 49.21 - 3.1 Å with a 5% of reflections used for R_{free}) were performed with PHENIX[14]. For refinement, each domain (i.e., At4CL1 and VvSTS) of the fusion protein was used as a TLS group. Data collection and refinement statistics are reported in Table S1. The atomic coordinates and structure factors for the 4CL::STS fusion protein have been deposited in the Protein Data Bank (PDB: 3TSY).

2. EXPERIMENTAL RESULTS

Table S1: Data Collection and Refinement Statistics

Crystal	
Space group	P $4_1$22
Cell dimensions	*a* = *b* = 117.7 Å, *c* = 244.0 Å
Data Collection	
Wavelength	0.979 Å
Resolution (highest shell)	49.21 - 3.1 Å (3.15-3.10 Å)
Reflections (total/unique)	200,672 /
Completeness (highest shell)	100% (100%)
$\langle I/\sigma \rangle$ (highest shell)	17.7 (1.9)
R_{sym} (highest shell)	8.2% (87.6%)
Refinement	
R_{cryst} / R_{free}	17.7% / 20.8%
No. of protein atoms	6,323
R.m.s. deviation, bond lengths	0.016 Å
R.m.s. deviation, bond angles	1.33°
Avg. B-factor - protein	87.0 $Å^2$
Stereochemistry: favored, allowed, generously allowed.	96.7, 3.2, 0.1%

Figure S1. Purification of At4CL1, VvSTS and 4CL::STS. (i) SDS-PAGE gel of purified proteins. Lane 1: molecular weight standard, Lane 2: At4CL1 (61 kDa), Lane 3: VvSTS (47 kDa), Lane 4: 4CL::STS (114 kDa). (ii) Elution profiles of At4CL1 (A), VvSTS (B), and 4CL::STS (C), from a HiLoad 16/60 Superdex-200 FPLC column are shown. At4CL1, VvSTS, and 4CL::STS elute from the column as 70 kDa, 70 kDa, and 200 kDa species corresponding to monomeric, dimeric, and dimeric forms, respectively.

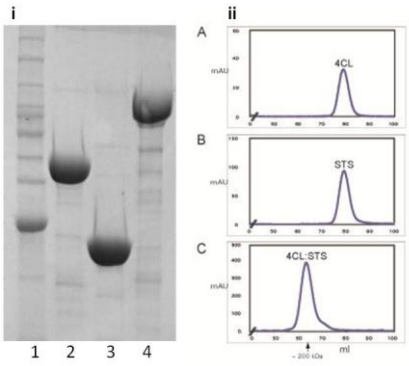


Figure S2. Representative electron density in the 4CL::STS fusion protein. (A) The $2F_o$-F_c omit map (1.5 σ) for the STS active site showing the catalytic residues. (B) The $2F_o$-F_c omit map (1.5 σ) for the 4CL active site showing ATP binding residues.

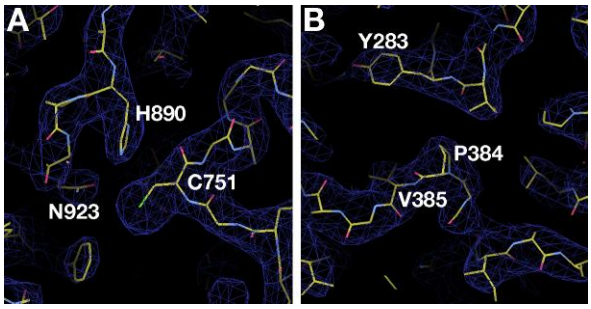


Figure S3. Comparison of plant 4CL structures highlighting movement of the C-terminal domain. (A) Structure of At4CL1 in the 4CL::STS fusion protein. The disordered C-terminal domain is shown in gold based on the poplar 4CL apoenzyme structure. (B) Structure of poplar 4CL with AMP bound in the active site (PDB: 3A9V). (C) Structure of the poplar 4CL apoenzyme (PDB: 3A9U). Note the rotational movement of the C-terminal domain, corresponding to the region colored gold in (A), when compared to that in (C).

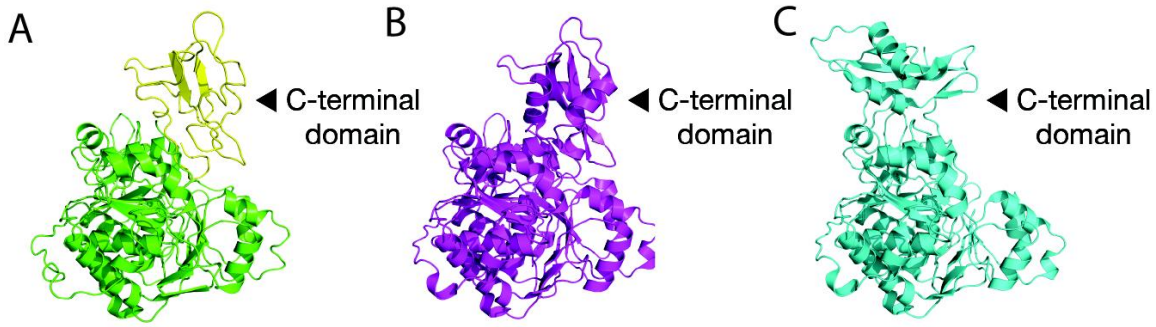
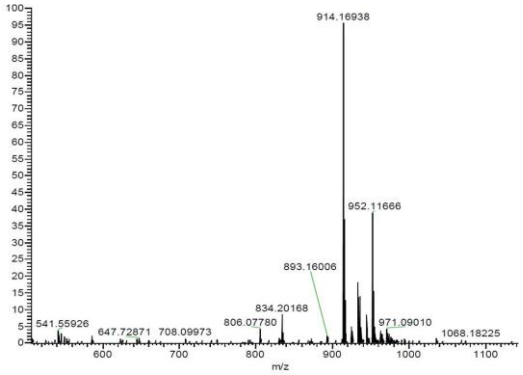


Figure S4. Identification of *p*-coumaroyl-CoA. The putative *p*-coumaroyl-CoA peak was confirmed by mass spectrometry using an LTQ Orbitrap Velos instrument. The theoretical $[M\text{-}H]^+$ is 914.15 with single charge (H^+) and the *m/z* of the main peak on the spectrum is 914.16.



3. REFERENCES FOR SUPPORTING INFORMATION

(1) Lee, D.; Ellard, M.; Wanner, L.A.; Davis, K.R.; Douglas, C.J. *Plant Mol. Biol.* **1995**, *28*, 871-884.

(2) Sparvoli, F.; Martin, C.; Scienza, A.; Gavazzi, G.; Tonelli, C. *Plant Mol. Biol.* **1994**, *24*, 743-755.

(3) Zhang, Y.; Li, S.Z.; Pan, X.; Cahoon, R.E.; Jaworski, J.G.; Wang, X.; Jez, J.M.; Chen, F.; Yu, O. *J. Am. Chem. Soc.* **2006**, *128*, 13030-13031.

(4) Obel, N.; Scheller, H.V. *Anal. Biochem.* **2000**, *286*, 38-44.

(5) Lim, C.G.; Fowler, Z.L.; Hueller, T.; Scaffer, S.; Koffas, M.A. *Appl. Environ. Microbiol.* **2011**, *77*, 3451-3460.

(6) Knobloch, K.H.; Hahlbrock, K. *Arch. Biochem. Biophys.* **1977**, *184*, 237-248.

(7) Wang ,Y.; Halls, C.; Zhang, J.; Matsuno, M.; Zhang, Y.; Yu, O. *Metab. Eng.* **2011**, *13*, 455-463.

(8)Careri, M.; Corradini, C.; Elviri, L.; Nicoletti, I.; Zagnoni, I. *J Agric Food Chem* **2004**, *52*, 6868-74.

(9)Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307-326.

(10)McCoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. *J. Appl. Cryst.* **2007**, *40*, 658-674.

(11)Hu, Y.; Gai, Y.; Yin, L.; Wang, X.; Feng, C.; Feng, L.; Li, D.; Jiang, X.N.; Wang, D.C. *Plant Cell* **2010**, *22*, 3093-3104.

(12) Austin, M.B.; Bowman, M.E.; Ferrer, J.L.; Schröder, J.; Noel, J.P. *Chem. Biol.* **2004**, *11*, 1179-1194.

(13)Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. *Acta Crystallogr. D* **2010**, *66*, 486-501.

(14)Adams, P.D. et al. *Acta Crystallogr. D* **2010**, *66*, 213-221.