JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

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[‡]

[†]Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, Missouri 63132, United States

[‡]Department of Biology, Washington University, One Brookings Drive, Campus Box 1137, St. Louis, Missouri 63130, United States

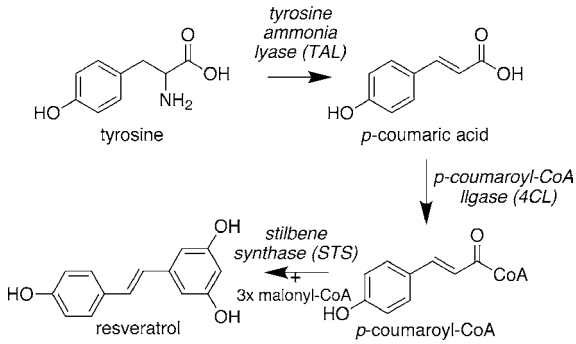
Supporting Information

ABSTRACT: To increase the biochemical efficiency of biosynthetic systems, metabolic engineers have explored different approaches for organizing enzymes, including the generation of unnatural fusion proteins. Previous work aimed at improving the biosynthesis of resveratrol, a stilbene associated a range of health-promoting activities, in yeast used an unnatural engineered fusion protein of *Arabidopsis thaliana* (thale cress) 4-coumaroyl-CoA ligase (At4CL1) and *Vitis vinifera* (grape) stilbene synthase (VvSTS) to increase resveratrol levels 15-fold relative to yeast expressing the individual enzymes. Here we present the crystallographic and biochemical analysis of the 4CL::STS fusion protein. Determination of the X-ray crystal structure of 4CL::STS provides the first molecular view of an artificial didomain adenylation/ketosynthase fusion protein. Comparison of the steady-state kinetic properties of At4CL1, VvSTS, and 4CL::STS demonstrates that the fusion protein improves catalytic efficiency of either reaction less than 3-fold. Structural and kinetic analysis suggests that colocalization of the two enzyme active sites within 70 Å of each other provides the basis for enhanced in vivo synthesis of resveratrol.

A major aim in the metabolic engineering of biochemical pathways in host organisms is to increase system efficiency and the yields of compound production. One strategy for achieving this goal is to organize pathway enzymes either as macromolecular complexes or as artificially scaffolded proteins in close proximity.[1] Both approaches mimic the natural colocalization of proteins in cells and/or the physical channeling of reactive intermediates between enzyme active sites to improve metabolic efficiency.[2] In microbes and plants, chimeric genes encoding different biosynthetic proteins are common and provide a means of grouping sequential metabolic steps or coordinating the regulation of enzyme activities.[3] For engineering of in vivo biochemical pathways, linking genes to generate unnatural fusion proteins is a promising approach for improving metabolic yields and minimizing the number of heterologous expression vectors.

Multiple studies have demonstrated the ability to produce flavonoids and related compounds, such as the stilbene resveratrol, in *Sacchromyces cerevisiae* and *Escherichia coli*.[4] In response to stress, some plants, such as grape and peanut, produce resveratrol as a phytoalexin,[5] but it has also been implicated as a health-promoting compound in red wine and in the extension of life span in various organisms.[6] Resveratrol is synthesized from *p*-coumaric acid in two enzymatic reactions[7] (Scheme 1). In plants, *p*-coumaric acid is formed from phenylalanine by phenylalanine ammonia lyase and cinnamate-4-hydroxylase; however, for metabolic engineering purposes, bacterial tyrosine ammonia lyase can replace these enzymes and generate *p*-coumaric acid directly from tyrosine instead of phenylalanine.[4b,7,8] Next, *p*-coumaroyl-CoA ligase (4CL) activates *p*-coumaric acid to provide a critical coenzyme A (CoA)-linked intermediate. Stilbene synthase (STS) uses *p*-coumaroyl-CoA and three malonyl-CoA molecules to form a tetraketide intermediate that cyclizes via an aldol condensation to yield resveratrol.

Scheme 1. Resveratrol Biosynthesis Pathway



To evaluate the potential of using unnatural protein fusions as a metabolic engineering tool, we previously used the resveratrol biosynthesis pathway as a test case.[1b] Two yeast strains were generated. The first expressed *Arabidopsis thaliana* (thale cress) 4CL1[9a] (At4CL1) and *Vitis vinifera* (grape) STS[9b] (VvSTS) as individual proteins, and the second expressed an unnatural 4CL::STS fusion protein. Both strains were provided *p*-coumaric acid for uptake, and the synthesis of resveratrol was monitored. Comparison of the product yield in each strain showed that *S. cerevisiae* expressing the fusion protein produced 15-fold more resveratrol than yeast expressing the individual proteins.[1b] Here we examine the structural and kinetic properties of the 4CL::STS fusion protein in comparison with the At4CL1 and VvSTS proteins.

Received: September 12, 2011
Published: November 30, 2011

The 979 amino acid 4CL::STS fusion protein, in which At4CL1 is connected to VvSTS by a linker containing three amino acids (Gly-Ser-Gly), was expressed in *E. coli* with an N-terminal hexahistidine tag and purified to homogeneity by nickel-affinity and size-exclusion chromatographies [Figure S1 in the Supporting Information (SI)]. A similar approach was used for both At4CL1 and VvSTS. Each protein eluted from the size-exclusion column as a ~70 kDa species corresponding to monomeric At4CL1 and dimeric VvSTS, respectively, as reported previously.[9] The 4CL::STS fusion protein eluted as a 200 kDa species, indicating a dimeric form and suggesting that the N-terminal addition of 4CL does not interfere with STS oligomerization.

To establish the molecular organization of the 4CL and STS active sites in the engineered fusion protein, the 3.1 Å resolution crystal structure of 4CL::STS was determined by molecular replacement with R_{cryst} = 17.7% and R_{free} = 20.8% (Figure 1 and Table S1 in the SI). The molecular replacement

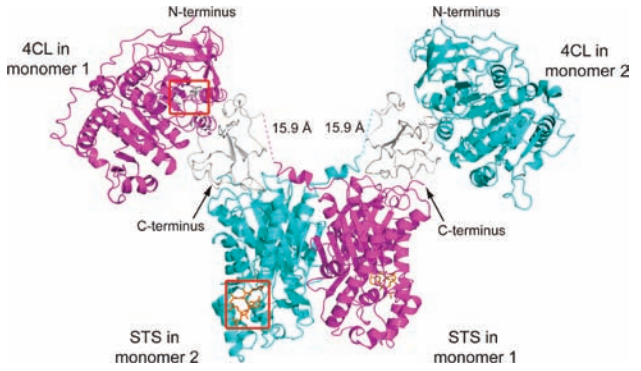


Figure 1. Overall structure of the 4CL::STS fusion protein. The ribbon diagram shows the dimer with crystallographically related monomers colored magenta and cyan. A model of the disordered C-terminal region of At4CL1 based on the poplar 4CL structure is shown in white. Stick representations of AMP (gray) in At4CL1 and CoA (gold) in STS were modeled into each structure to indicate the positions of the active sites, which are highlighted with red boxes.

was performed using two independent search ensembles, poplar 4CL (PDB entry 3A9U)[9a] and peanut STS (PDB entry 1Z1E).[11d] The asymmetric unit contains one copy of the fusion protein, with the physiologic dimer formed by crystallographic symmetry. The refined model of 4CL::STS includes residues corresponding to residues 22−389 of At4CL1 and 2−331 of VvSTS. The C-terminal domain of At4CL1 (102 residues), the GSG linker, and the first residue of VvSTS were disordered; however, the complete structure of poplar 4CL[9a] helped define the connection between the two domains in the fusion protein (Figure 1). The disordered C-terminal domain of At4CL1 corresponds to the same region of the poplar 4CL structure that undergoes a rotational movement upon binding of either AMP or the reaction intermediate analogue adenosine 5′-(3-(4-hydroxyphenyl)propyl)phosphate (APP).[9a] In enzymes that catalyze biosynthetic adenylation reactions, including non-ribosomal peptide synthetases and acyl-CoA synthetases, the C-terminal region is dynamic and can rotate up to 140° during the course of the two half-reactions (i.e., formation of the adenylated intermediate and nucleophilic attack during the transferase step).[10] The position of the disordered C-terminal region of 4CL in the fusion protein was modeled on the basis of the "capped" active site form observed in poplar 4CL[9a] (white ribbon in Figure 1).

The 4CL::STS fusion protein dimer shows twofold rotational symmetry with a dyad axis between the STS enzymes from the two monomers (Figure 1). The "Y"-shaped structure is formed with each 4CL half of the protein as the arms and the STS dimer as the base. The 4CL and STS portions of each monomer are located on opposite sides of the dyad axis, with the active site of At4CL in one monomer facing the active site of VvSTS in the other monomer. The two active sites in each side of the 4CL::STS fusion protein are separated by a distance of ~70 Å.

The three-dimensional structure of VvSTS in the fusion protein shares the canonical thiolase-like fold observed in other plant stilbene, chalcone, and pyrone synthases.[11] The overall structure of the grape STS in the fusion protein showed root-mean-square deviations (rmsd's) of 0.44 Å^2 for 350 aligned C_α atoms relative to the pine STS (PDB entry 1XES) and peanut STS (PDB entry 1Z1E) and an rmsd of 0.39 Å^2 for 323 aligned C_α atoms compared to alfalfa chalcone synthase (PDB entry 1BI5).[12] As observed in the previously solved plant polyketide synthase structures, the positions of the catalytically essential residues Cys751, His890, and Asn923 are also maintained in the VvSTS active site of the fusion protein (Figure S2A).[13] Weak electron density corresponding to the pantothenyl arm of a CoA molecule was observed in the STS active site. Although CoA was not included in the final refined model, the ligand is shown in Figure 1 to highlight the location of the catalytic site in the STS portion of the fusion protein.

The structure of the ordered N-terminal domain of At4CL1 in the fusion protein is similar to that of poplar 4CL (Figure 1 and Figure S3) with rmsd's of 0.50−0.63 Å^2 for 335 aligned C_α atoms.[12] Likewise, key residues of the catalytic (Lys441, Gln446, and Lys570) and substrate (Tyr283, Gly378, Pro384, and Val385) binding site are conserved in both the *Arabidopsis* and poplar enzymes (Figure S2B).[9a] However, Ala353 in At4CL1 was found at a position corresponding to the Gly306 residue in poplar, which was suggested to play a role in substrate binding.[9a] The 102 C-terminal residues of At4CL1 are largely disordered in the fusion protein and were not included in the final structural model, although the weak electron density for this region suggests that its conformation resembles that observed in the poplar 4CL structure with bound AMP (Figure S3). Modeling of the At4CL1 C-terminal domain based on the poplar 4CL structure with bound AMP suggested that its placement in the fusion protein would not impede rotational movement of this domain during the catalytic cycle of At4CL1.[9a] In the model, the distance between Lys579 (the last 4CL residue) and Ala589 (the first STS residue) is 15.9 Å (Figure 1), which suggests that there is sufficient room for the intervening GSG-linker residues.

To compare the functional properties of 4CL::STS, At4CL1, and VvSTS, steady-state kinetic assays were performed (Table 1). For each substrate in the 4CL reaction (*p*-coumaric acid, ATP, and CoA), 4CL::STS exhibited 1.4- to 2.7-fold higher k_{cat}/K_m values than At4CL1. Similar modest improvements in catalytic efficiency (1.3-fold for malonyl-CoA and 2.4-fold for *p*-coumaroyl-CoA) were also observed for STS activity in the fusion protein. These modest differences may result from changes in the mobility of each domain of the fusion protein relative to the separated proteins.

Comparison of resveratrol formation in assays with At4CL and VvSTS [specific activity: 1.55 ± 0.02 nmol min^{-1} (μg of protein)$^{-1}$] versus 4CL1::STS [specific activity: 1.62 ± 0.05 nmol min^{-1} (μg of protein)$^{-1}$] showed similar rates.

dx.doi.org/10.1021/ja2085993 | J. Am. Chem.Soc. 2011, 133, 20684−20687

Table 1. Steady-State Kinetic Parameters[a]

substrate	parameter	At4CL1	4CL1::STS
p-coumaric acid	k_{cat}	14.9	22.5
	K_m	66.7	69.9
	V/K	3723	5365
ATP	k_{cat}	60.3	121
	K_m	39.4	30.9
	V/K	25510	65260
CoA	k_{cat}	53.6	91.1
	K_m	92.0	58.5
	V/K	9710	25950
substrate	parameter	VvSTS	4CL1::STS
p-coumaroyl-CoA	k_{cat}	1.74	1.08
	K_m	18.4	4.89
	V/K	1576	3680
malonyl-CoA	k_{cat}	0.58	0.48
	K_m	55.2	35.4
	V/K	175	226

[a]Units for k_{cat}, K_m, and V/K are min^{-1}, μM, and $M^{-1}\ s^{-1}$, respectively. Mean values ($n = 3$) are shown; the errors are less than 10%.

This result is not unexpected, as the concentrations of protein required for measurement of the activity cannot distinguish diffusion-limited steps in the pathway. Successful demonstrations of improved biosynthesis achieved by organizing metabolic proteins through either fusions or scaffolding strategies are largely supported solely by in vivo experiments.[1b−d] The difference between in vitro protein assay results and in vivo titers likely reflects lower protein concentrations, metabolite levels, competing reactions, and the cellular environment (i.e., macromolecular crowding, viscosity, and protein localization) within a production organism.

Crystallographic and kinetic analysis of the 4CL::STS fusion protein has demonstrated that generation of the engineered system does not drastically alter the structural and functional properties of either At4CL1 or VvSTS and suggests that improved in vivo production of resveratrol in yeast likely results from localization of the active sites in close proximity and not from improvements in catalytic efficiency of either enzyme. Therefore, a reduced physical distance between active centers of biosynthetic enzymes is important and sufficient for improving an engineered pathway in a heterologous host. The architecture of the 4CL::STS fusion protein also provides a scaffold for the addition and/or modification of domain organization in engineered proteins. In conclusion, structure−function analysis of At4CL1, VvSTS, and 4CL::STS indicates that improving the localization and/or organization of metabolic enzymes in close proximity within a host organism is an important design consideration for synthetic biology applications.

■ ASSOCIATED CONTENT

Supporting Information

Details of experimental procedures; X-ray data collection and refinement statistics (Table S1); purification of At4CL1, VvSTS, and 4CL::STS (Figure S1); representative electron density (Figure S2); comparison of 4CL structures (Figure S3); and identification of *p*-coumaroyl-CoA (Figure S4). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

oyu@danforthcenter.org

Author Contributions

[§]These authors contributed equally.

■ ACKNOWLEDGMENTS

This work was funded by DOE-DE-SC0001295 and NSF-MCB-0923779 to O.Y. M.W., a student at Ladue Horton Watkins High School, was sponsored by the Pfizer-Solutia Students/Teachers as Research Scientists (STARS) Program. Portions of this research were carried out at the Argonne National Laboratory Structural Biology Center at the Advanced Photon Source, a National User Facility operated by the University of Chicago for the Department of Energy Office of Biological and Environmental Research (DE-AC02-06CH11357).

■ REFERENCES

(1) (a) Tian, L.; Dixon, R. A. *Planta* **2006**, *224*, 496. (b) 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. (c) Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L.; Keasling, J. D. *Nat. Biotechnol.* **2009**, *27*, 753. (d) Moon, T. S.; Dueber, J. E.; Shiue, E.; Prather, K. L. *Metab. Eng.* **2010**, *12*, 298.

(2) (a) Smith, S.; Tsai, S. C. *Nat. Prod. Rep.* **2007**, *24*, 1041. (b) Kittendorf, J. D.; Sherman, D. H. *Curr. Opin. Biotechnol.* **2006**, *17*, 597. (c) Winkel, B. S. *Annu. Rev. Plant Biol.* **2004**, *55*, 85.

(3) (a) Rousseau, G. G.; Hue, L. *Prog. Nucleic Acid Res. Mol. Biol.* **1993**, *45*, 99. (b) Osumi, T. *Biochimie* **1993**, *75*, 243. (c) Ivanetich, K. M.; Santi, D. V. *FASEB J.* **1990**, 1591. (d) Kurland, J. J.; Pilkis, S. J. *Protein Sci.* **1995**, *4*, 1023. (e) Nagradova, N. *IUBMB Life* **2003**, *55*, 459. (f) Bretonnet, A. S.; Jordheim, L. P.; Dumontet, C.; Lancelin, J. M. *FEBS Lett.* **2005**, *579*, 3363. (g) Liberles, J. S.; Thorolfsson, M.; Martinez, A. *Amino Acids* **2005**, *28*, 1. (h) Schroeder, A. C.; Zhu, C.; Yanamadala, S. R.; Cahoon, R. E.; Arkus, K. A. J.; Wachsstock, L.; Bleeke, J.; Krishnan, H. B.; Jez, J. M. *J. Biol. Chem.* **2010**, *285*, 827.

(4) (a) Becker, J. V.; Armstrong, G. O.; van der Merwe, M. J.; Lambrechts, M. G.; Vivier, M. A. *FEMS Yeast Res.* **2003**, *4*, 79. (b) Watts, K. T.; Lee, P. C.; Schmidt-Dannert, C. *ChemBioChem* **2004**, *5*, 500. (c) Ralston, L.; Subramanian, S.; Matsuno, M.; Yu, O. *Plant Physiol.* **2005**, *137*, 1375.

(5) Schröder, J.; Schröder, G. *Z. Naturforsch., C: J. Biosci.* **1990**, *45*, 1.

(6) (a) Jang, M.; Cai, L.; Udeani, G. O.; Slowing, K. V.; Thomas, C. F.; Beecher, C. W.; Fong, H. H.; Farnsworth, N. R.; Kinghorn, A. D.; Mehta, R. G.; Moon, R. C.; Pezzuto, J. M. *Science* **1997**, *275*, 218. (b) Howitz, K. T.; Bitterman, K. J.; Cohen, H. Y.; Lamming, D. W.; Lavu, S.; Wood, J. G.; Zipkin, R. E.; Chung, P.; Kisielewski, A.; Zhang, L. L.; Scherer, B.; Sinclair, D. A. *Nature* **2003**, *425*, 191. (c) Valenzano, D. R.; Terzibasi, E.; Genade, T.; Cattaneo, A.; Domenici, L.; Cellerino, A. *Curr. Biol.* **2006**, *16*, 296. (d) Kaeberlein, M.; McDonagh, T.; Heltweg, B.; Hixon, J.; Westman, E. A.; Caldwell, S. D.; Napper, A.; Curtis, R.; DiStefano, P. S.; Fields, S.; Bedalov, A.; Kennedy, B. K. *J. Biol. Chem.* **2005**, *280*, 17038.

(7) Yu, O.; Jez, J. M. *Plant J.* **2008**, *54*, 750.

(8) (a) Kyndt, J. A.; Meyer, T. E.; Cusanovich, M. A.; Van Beeumen, J. J. *FEBS Lett.* **2002**, *512*, 240. (b) Schroeder, A. C.; Kumaran, S.; Hicks, L. M.; Cahoon, R. E.; Halls, C.; Yu, O.; Jez, J. M. *Phytochemistry* **2008**, *69*, 1496.

(9) (a) 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. (b) Melchior, F.; Kindl, H. *FEBS Lett.* **1990**, *268*, 17.

(10) Gulick, A. M. *ACS Chem. Biol.* **2009**, *4*, 811.

(11) (a) Ferrer, J. L.; Jez, J. M.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. *Nat. Struct. Biol.* **1999**, *6*, 775. (b) Jez, J. M.; Austin, M. B.;

Ferrer, J. L.; Bowman, M. E.; Schröder, J.; Noel, J. P. *Chem. Biol.* **2000**, *7*, 919. (c) Austin, M. B.; Bowman, M. E.; Ferrer, J. L.; Schröder, J.; Noel, J. P. *Chem. Biol.* **2004**, *11*, 1179. (d) Shomura, Y.; Torayama, I.; Suh, D. Y.; Xiang, T.; Kita, A.; Sankawa, U.; Miki, K. *Proteins* **2005**, *60*, 803.

(12) Holm, L.; Rosenstrom, P. *Nucleic Acids Res.* **2010**, *38*, W545.

(13) (a) Jez, J. M.; Ferrer, J. L.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. *Biochemistry* **2000**, *39*, 890. (b) Jez, J. M.; Noel, J. P. *J. Biol. Chem.* **2000**, *275*, 39640.