

## Chapter 4

### The Hydrocarbon-producing Thioesterase From Olefin Synthase

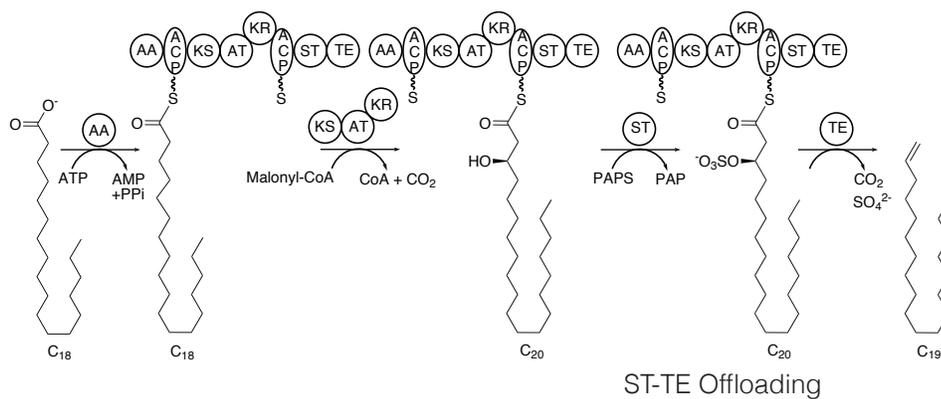
#### Summary

Olefin Synthase (OLS) is a cyanobacterial system that converts fatty acids to terminal double bond hydrocarbons, which is a promising target for biofuel engineering. The OLS tandem C-terminal domains, a sulfotransferase (ST) and a thioesterase (TE), resemble the ST-TE system in the curacin A polyketide synthase that also produces a terminal alkene. In the curacin system, CurM ST activates a  $\beta$ -hydroxy group by sulfonation and CurM TE catalyzes a decarboxylation and sulfate elimination as well as thioester hydrolysis of the acyl carrier protein (ACP)-tethered substrate. OLS ST has recently been characterized as a sulfotransferase, but little is known about OLS TE. Here we report the structure and functional investigation of the OLS TE and the OLS ST-TE didomain. OLS TE showed  $\beta$ -sulfate specific hydrolysis for  $C_{12}$  and  $C_{14}$  acyl-ACP substrates. The 1.7-Å crystal structure of OLS TE closely resembles CurM TE and retains the lid structure, dimer interface, and open active-site cleft distinctive to this family of decarboxylating TEs. A region on the lid containing positively charged amino acids was predicted to be a site for ACP interaction and was tested through charge-altering amino acid substitutions. The 3.1-Å structure of the OLS ST-TE didomain shows no interaction between the ST and TE on the same polypeptide except for the disordered linker that connects them. Assays varying the ST-TE linker length confirm that there is no evidence for direct substrate channeling between OLS ST and TE.

#### Introduction

Biofuel provides an attractive alternative to petroleum-derived fuel, potentially addressing energy security and environmental concerns (64). Terminal-alkene hydrocarbons are appealing biofuels compatible with existing engine, fuel storage, and fuel distribution systems (61, 62). Olefin synthase (OLS), which produces terminal olefin hydrocarbons (n+1) from fatty acids (n) (52, 10), is a promising target for biofuel

development (Figure 4.1). Olefin synthase activity was demonstrated by gene deletion and feeding studies for OLS in *Synechococcus* PCC 7002 (52) where C<sub>18</sub> fatty acids were converted into C<sub>19</sub> terminal olefins. The *ols* gene has so far been found in a total of five cyanobacterial genome sequences, including *Synechococcus* PCC 7002 (10).



#### Figure 4.1 Olefin synthase (OLS) catalysis

The acyl activating (AA) domain loads a fatty acid onto the adjacent acyl carrier protein (ACP) in an ATP dependent manner. The ketosynthase (KS) and acyltransferase (AT) catalyze malonate extension of the acyl-ACP substrate and transfer to the second ACP in the module. Subsequently, the ketoreductase (KR) domain processes the  $\beta$ -carbonyl to a  $\beta$ -hydroxy. The sulfotransferase (ST) sulfonates the  $\beta$ -hydroxy group from the sulfonate donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS). The thioesterase (TE) domain then works in a  $\beta$ -sulfate dependent manner to hydrolyze, decarboxylate, and eliminate sulfate to produce a terminal olefin.

The OLS protein contains eight domains, which are predicted to activate, extend, and decarboxylate fatty acids to produce terminal-olefin hydrocarbons. The N-terminal acyl-activating (AA) domain is similar to other proteins known to load a fatty acid onto the phosphopantetheine prosthetic arm of an acyl carrier protein (ACP) (53) (Figure 4.1). The AA and subsequent ACP are followed by ketosynthase (KS), acyl transferase (AT), ketoreductase (KR), and ACP domains (Figure 4.1). The KS-AT-KR reactions are chemically identical to reactions of fatty acid synthesis (two-carbon extension and reduction to a  $\beta$ -hydroxy), but the domains are more closely related to polyketide synthase (PKS) enzymatic domains in sequence and assembly. The C-terminus of the OLS polypeptide contains a sulfotransferase (ST) domain followed by a decarboxylating thioesterase (TE) domain (Figure 4.1). A similar ST-TE didomain in the curacin A

biosynthetic pathway (2) was shown to activate the  $\beta$ -hydroxy as a leaving group through sulfonation (ST) from the donor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) followed by a  $\beta$ -sulfate specific hydrolysis and decarboxylative elimination (TE) (27). The ST-TE acts upon the ACP-linked intermediates, tethered via a thioester, which are released as terminal olefins instead of the carboxylic acids or macrolactones expected from canonical TEs of PKS and FAS systems.

The closely related PKS CurM, although not an olefin synthase, contains the same canonical domains (KS-AT-KR-ACP) and offloading domains (ST-TE) as OLS and functions to create the terminal olefin in the small-molecule natural product curacin A. The high sequence identity (47%) and the identical chemical group in the final products indicate that the PKS and OLS ST-TE domains perform the same chemistry. In earlier work (Chapter 3) we showed that OLS ST has sulfotransferase activity with  $\beta$ -hydroxyacyl-ACP linked substrates (121). OLS ST does not tolerate the C5-methoxy chemical group that mimics the natural CurM PKS ST substrate while CurM ST is indiscriminate with regards to the C5-methoxy (121). OLS ST and CurM ST have similar 3-D structures and contain the same PAPS binding region and active site residues. A dynamic flap covering the active site is important for substrate selectivity and product formation based on the activities of chimeras of the PKS and OLS STs with exchanged active-site flaps (121).

The structure of CurM TE (1.7 Å) shows how a typical  $\alpha/\beta$  hydrolase fold is adapted to select for  $\beta$ -sulfate substrates and catalyze decarboxylative elimination (99). A Ser-His-Glu/Asp catalytic triad sits on top of a conserved  $\alpha/\beta$  hydrolase core. The lid, which differs from lids of other  $\alpha/\beta$  hydrolases, is anchored open and away from the active site on one side by a buried Arg salt bridge with the core and on the other by a dimer interface, which appears unique to the decarboxylating TEs (99). An Arg on the lid-to-core linker was shown to be important for recognizing  $\beta$ -sulfate substrates and may assist in the efficient positioning of substrates for catalysis (99). It appears that selectivity for  $\beta$ -sulfate substrates is achieved by the open nature of the CurM TE, which is unable to efficiently position a substrate for catalysis except for those bearing a  $\beta$ -sulfate.

Since the ST-TE domains are integral to the creation of the terminal olefin in OLS, we were motivated to characterize the OLS TE and the OLS ST-TE didomain for a complete view of the OLS decarboxylative offloading system. OLS TE catalyzes  $\beta$ -sulfate specific hydrolysis of acyl-ACP linked substrates. The OLS TE structure retains the structural features identified as important for  $\beta$ -sulfate specificity in CurM TE. From amino acid variants, Arg201 is identified as important for activity with  $\beta$ -sulfate substrates. The upper lid is identified as a putative ACP interaction site. Studies of the OLS ST-TE didomain structure and activity show no evidence for interaction of direct substrate channeling between OLS TE and OLS ST.

## **Experimental**

### *Cloning and site directed mutagenesis*

Inserts encoding OLS TE (OLS residues 2435-2720 here referred to as 1-286) and OLS ST-TE (OLS residues 2121-2720 here referred to as 1-600) were amplified from synthetic DNA (GeneArt) and inserted in the pMCSG7 vector encoding an N-terminal 6xHis tag. Site directed mutagenesis and deletions were performed using the QuikChange protocol (Agilent). All constructs were verified by sequencing. CurM ACP, CurM ST, and OLS ST were expressed as previously described (27, 121).

### *Protein expression and purification*

*E. coli* strain BL21(DE3) was transformed with expression plasmid, grown at 37°C in 500 mL TB with 4% glycerol to an OD<sub>600</sub> of 1.0, cooled to 20°C, induced with IPTG (final concentration 0.2 mM), and grown for an additional 18 h. All purification steps were performed at 4°C. The cell pellet from 500 mL of cell culture was resuspended in 40 mL Buffer A (20 mM Tris pH 7.9, 500 mM NaCl, 10% glycerol) plus 20 mM imidazole and lysed by sonication, and the soluble fraction loaded onto a 5-mL HisTrap Ni NTA column (GE Healthcare). The proteins were eluted in Buffer A with a linear gradient of 20 to 650 mM imidazole. The 6xHis tag was removed by 2-h incubation with 1 mM DTT and tobacco etch virus (TEV) protease (1 mg protease for 50 mg protein) at room temperature. After overnight dialysis at 4°C in Buffer A with 1 mM DTT, the remaining His-tagged proteins were removed by Ni-affinity chromatography, followed by size exclusion chromatography with a HiLoad 16/60 Superdex 200 column

(GE Healthcare) pre-equilibrated with Buffer A. The purified proteins were concentrated to 10 mg/mL, flash frozen in liquid N<sub>2</sub>, and stored at -80°C. OLS TE variants used for assay were purified as above but using a single HisTrap NiNTA purification step followed by overnight dialysis in Buffer A with yields similar to wild type. Yields per 500 mL culture were 120 mg for OLS TE and 60 mg for OLS ST-TE.

### *Crystallization*

Crystals of OLS TE grew at 4°C within 4 days by vapor diffusion from a 1:1 mix of protein stock (8.5 mg/mL OLS TE, 100 mM NaCl, 20 mM Tris pH 7.9) and well solution (20% PEG3350, 200 mM ammonium acetate, 100 mM Bis Tris pH 5.9). OLS TE crystals were cryo protected in well solution with 10% glycerol. OLS ST-TE crystals grew at 20°C within 6 days from a 1:1 mixture of protein stock (13.5 mg/mL OLS ST-TE, 500 mM NaCl, 10% glycerol, 20 mM Tris pH 7.9, 2 mM PAP) and well solution (21% PEG3350, 50 mM urea, 5% glycerol, 100 mM Bis Tris pH 5.9). Crystals were harvested in loops and flash cooled in liquid N<sub>2</sub>.

### *Data collection and structure determination*

Data were collected at GM/CA beamline 23ID-D at the Advanced Photon Source (APS) at Argonne National Lab (Argonne, IL). A 1.70-Å dataset was collected for a rod-shaped OLS TE crystal in 4x36° wedges, 25 μm apart with a 20-μm mini-beam (Table 4.1). A 3.10-Å dataset was collected for OLS ST-TE (Table 4.1). All data were processed using the HKL2000 suite (80) (Table 4.2, Table 4.3). The OLS TE structure was solved by molecular replacement with Phaser (81) using CurM TE (3QIT) (99) (47% sequence identity) as a search model. Refinement was done in REFMAC5 (86) from the CCP4 (87) suite using 7 TLS groups per monomer (88) (Figure 4.2). One dimer in the asymmetric unit provides two independent views of OLS TE. OLS ST-TE was solved by molecular replacement in Phaser (81) using OLS ST (4GBM) (121) and OLS TE as a search model. Refinement was done in BUSTER (106) using restraints to the high-resolution structures of the individual domains and 1 TLS group per chain (Figure 4.3). Model quality was evaluated with MolProbity (107). PyMOL was used to align structures and prepare figures (109).

**Table 4.1 Crystallographic summary**

	<b>OLS TE</b>	<b>OLS ST-TE</b>
<b>Diffraction Data</b>		
Space group	$P2_1$	$P2_12_12_1$
X-ray source	APS 23-ID-D	APS 23-ID-B
a, b, c (Å)	44.9, 54.2, 127.6	98.9, 132.2, 247.1
$\alpha, \beta, \gamma$ (°)	90, 91.3, 90	90, 90, 90
Wavelength (Å)	1.03320	1.03322
$d_{\min}$ (Å)	1.70 (1.76-1.70) <sup>a</sup>	3.1 (3.21-3.10)
Avg I/ $\sigma_I$	11.7 (2.1)	15.8 (2.7)
$R_{\text{sym}}$	0.080 (0.426)	0.108 (0.751)
Completeness	99.3 (96.6)	100.0 (100.0)
Avg. redundancy	2.9 (2.4)	7.3 (7.3)
Unique reflections	67,640	59,786
<b>Refinement</b>		
Data range (Å)	127.62-1.70	45.09 -3.10
No. reflections	67,522	59,697
$R_{\text{work}}/R_{\text{free}}$ <sup>b</sup>	0.158/0.190	0.325/0.348
RMS deviations		
Bonds (Å)	.008	.010
Angles (°)	1.11	1.11
Avg <i>B</i> -factors (Å <sup>2</sup> )		
Protein	17.5	124.4
Water	25.4	
Other	29.9	103.3
Ramachandran		
Allowed	100.0%	99.0%
Outliers	0.0%	1.0%
Protein Atoms	4628	17601
Water Molecules	761	0
Other Atoms	20	108

<sup>a</sup>Outermost shell in parentheses.

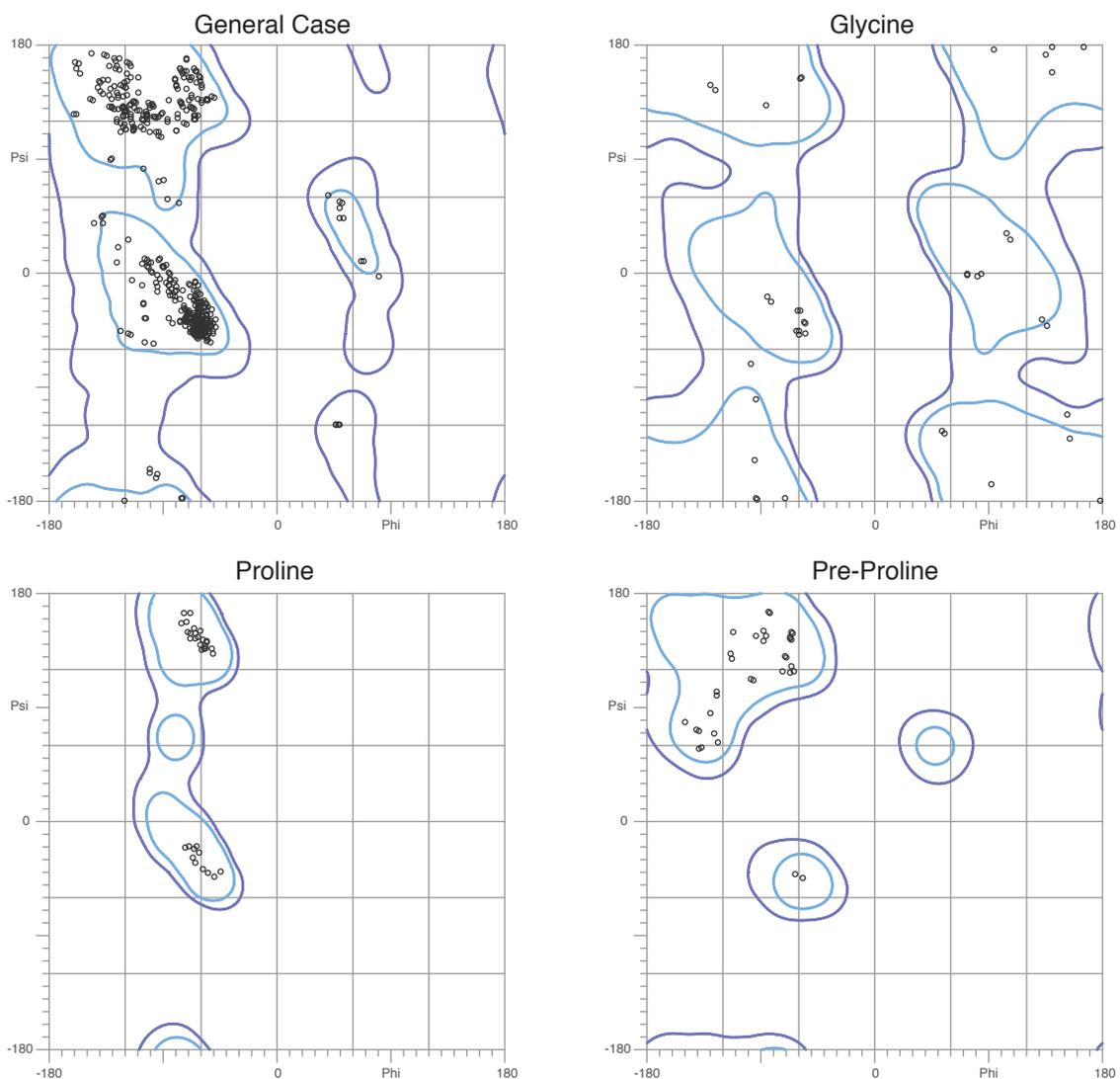
<sup>b</sup>The  $R_{\text{free}}$  data set included a random 5% of reflections.

**Table 4.2 Scaling statistics for OLS TE**

Shell Limit (Å)		Average I	Avg. error	Chi <sup>2</sup>	Linear R-factor	Square R-factor
Lower	Upper					
50.00	3.66	354.1	22.6	0.856	0.048	0.056
3.66	2.91	198.5	13.3	1.317	0.071	0.084
2.91	2.54	90.9	6.9	0.995	0.072	0.074
2.54	2.31	65.3	5.5	1.200	0.094	0.096
2.31	2.14	50.7	5.0	1.137	0.110	0.111
2.14	2.02	37.6	4.5	1.019	0.127	0.120
2.02	1.91	26.8	4.0	1.013	0.163	0.151
1.91	1.83	17.5	3.7	1.000	0.235	0.214
1.83	1.76	11.9	3.5	0.986	0.328	0.300
1.76	1.70	8.6	4.1	0.964	0.426	0.400
All	Reflections:	87.1	7.4	1.052	0.08	0.068

**Table 4.3 Scaling statistics of OLS ST-TE**

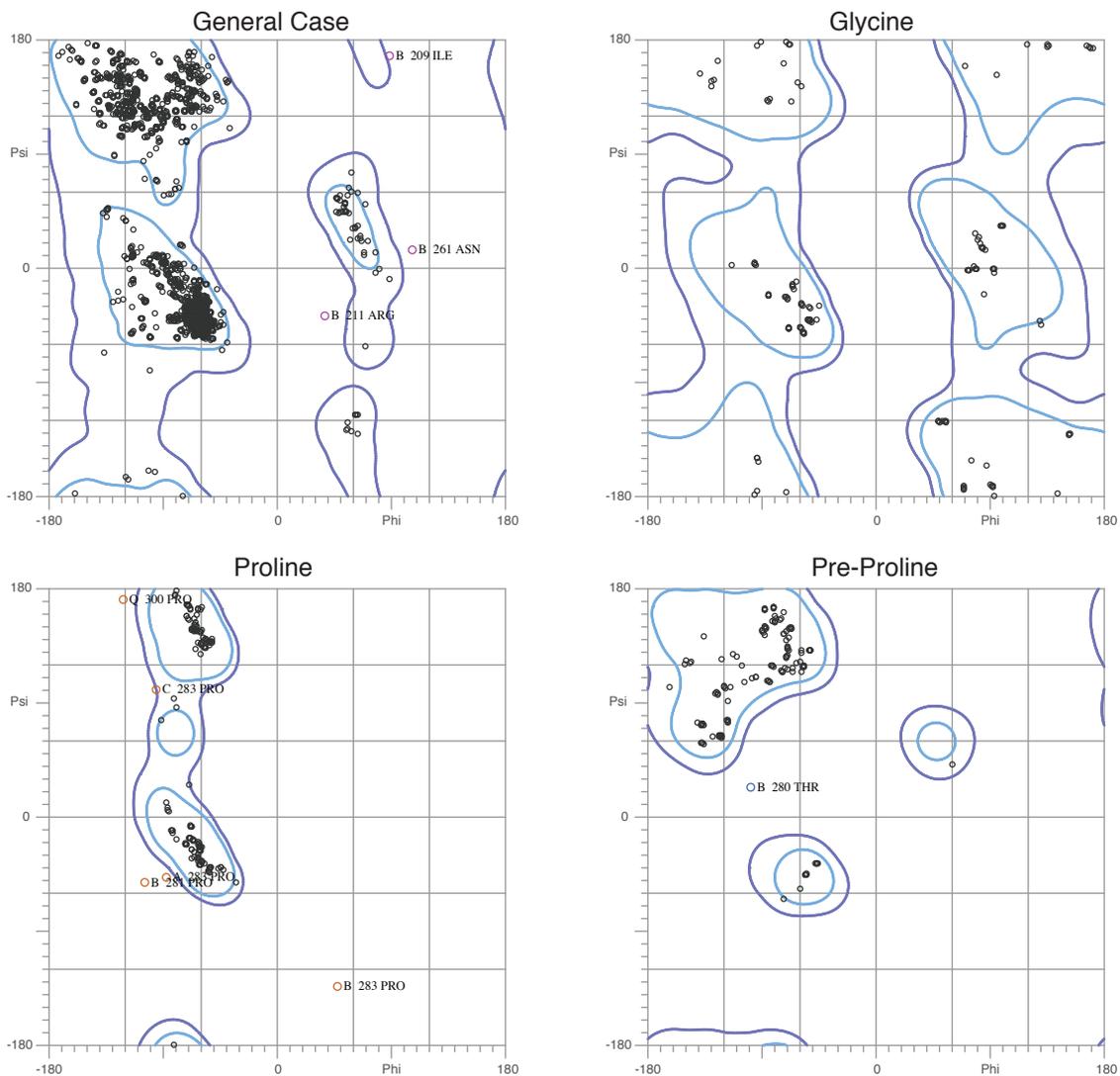
Shell Limit (Å)		Average I	Avg. error	Chi <sup>2</sup>	Linear R-factor	Square R-factor
Lower	Upper					
50.00	6.67	290.3	14.9	0.785	0.067	0.078
6.67	5.30	64.2	3.5	0.998	0.097	0.105
5.30	4.63	63.8	3.5	0.951	0.095	0.101
4.63	4.21	52.5	3.1	1.098	0.111	0.109
4.21	3.91	37.4	2.5	1.234	0.141	0.124
3.91	3.68	24.7	2.1	1.168	0.179	0.160
3.68	3.49	17.1	1.8	1.125	0.233	0.204
3.49	3.34	10.7	1.6	1.092	0.335	0.279
3.34	3.21	6.4	1.5	1.033	0.511	0.379
3.21	3.10	4.0	1.5	0.950	0.751	0.531
All	Reflections:	58.6	3.7	1.043	0.108	0.084



97.3% (581/597) of all residues were in favored (98%) regions.  
 100.0% (597/597) of all residues were in allowed (>99.8%) regions.

**Figure 4.2 Ramachandran analysis of OLS TE**

Ramachandran plots of the refined model of OLS TE. Plots were generated using MolProbity (<http://kinemage.biochem.duke.edu>) (89).



97.4% (2143/2201) of all residues were in favored (98%) regions.  
 99.6% (2192/2201) of all residues were in allowed (>99.8%) regions.

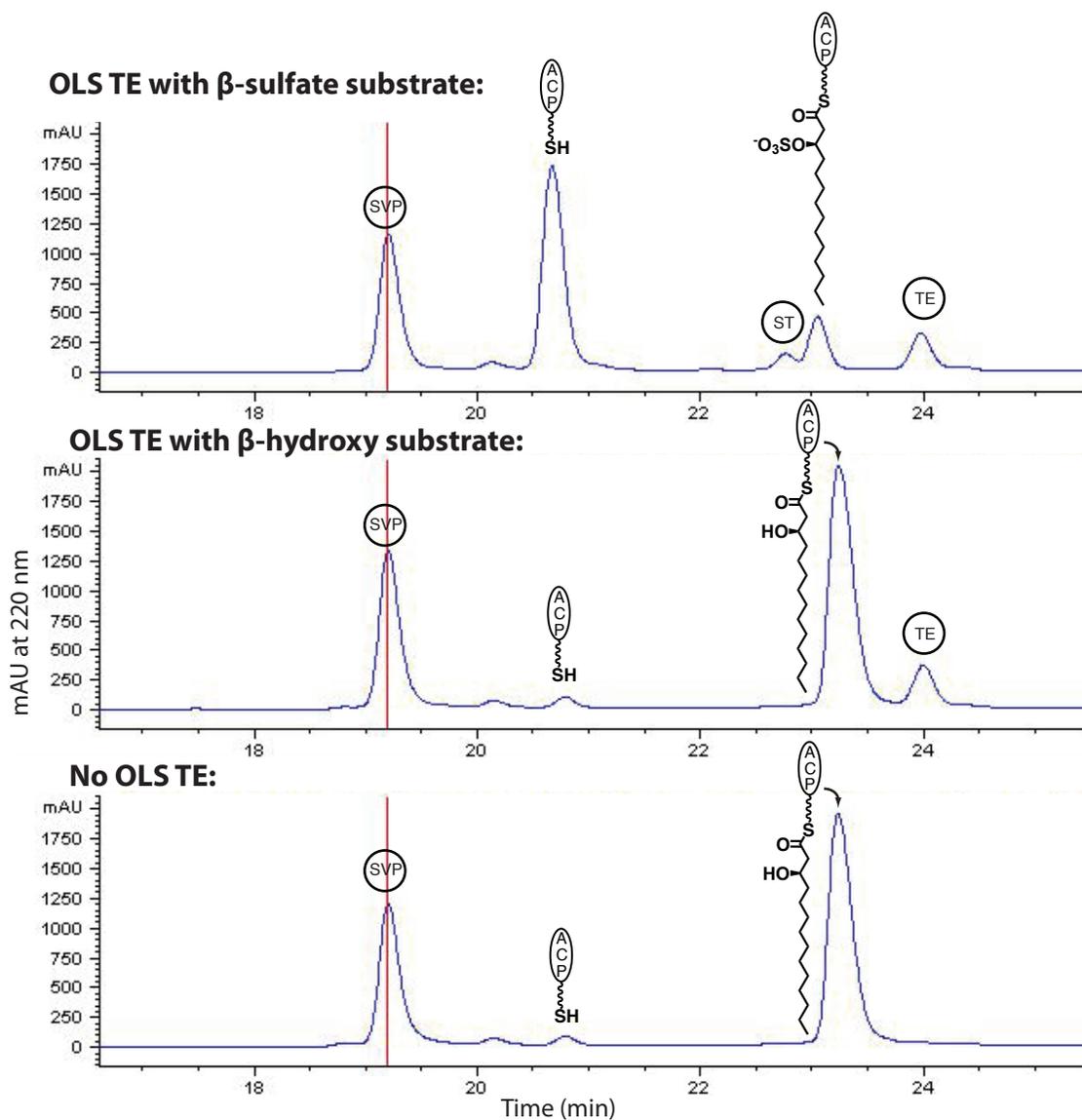
**Figure 4.3 Ramachandran analysis of OLS ST-TE**  
 Ramachandran plots of the refined model of OLS ST-TE. Plots were generated using MolProbity (<http://kinemage.biochem.duke.edu>) (89).

### Enzyme assays

OLS TE and OLS ST-TE were assayed using a modification of our previous protocols (27, 99, 121). Apo CurM ACP was loaded with substrate analog by 2-h incubation of 100  $\mu$ M (*R*)-3-hydroxytetradecanoyl-CoA (121), (*R*)-3-hydroxydodecanoyl-CoA (121), (*3R*)-3-hydroxy-5-methoxytetradecanoyl-CoA (27), or (*3S*)-3-hydroxy-5-methoxytetradecanoyl-CoA (27), with 50  $\mu$ M apo-ACP, 10  $\mu$ M *Streptomyces verticillus* Svp (94), 10 mM MgCl<sub>2</sub>, and 100 mM Tris pH 7.9 at 30°C. Complete loading was confirmed by reverse phase HPLC using a Jupiter C4 column (200 x 2.0 mm, 5  $\mu$ m, 300 Å, Phenomenex) and a linear elution gradient from 30% to 90% CH<sub>3</sub>CN (0.1% CF<sub>3</sub>CO<sub>2</sub>H)/H<sub>2</sub>O (0.1% CF<sub>3</sub>CO<sub>2</sub>H) over 45 min. After exchange into Buffer A and concentration (Amicon Ultra 10 kDa concentrators, Millipore), substrate-loaded ACP was flash frozen and stored at -80°C. To generate the sulfonated substrate for the TE assay, 100  $\mu$ M loaded ACP was incubated with 5  $\mu$ M CurM ST, 1.75 mM PAPS (Sigma), and 50 mM Tris pH 7.9 at room temperature for 30 min. Complete sulfonation was confirmed by HPLC. The TE reaction was initiated by addition of OLS TE (4  $\mu$ M). After 5 min the reaction was quenched with 5% formic acid. Conversion of loaded to holo ACP was quantitated as described above (Figure 4.4). Control reactions adding Buffer A in place of the ST or both the ST and TE were performed in parallel.

Assays of OLS ST-TE didomain were performed by incubating 100  $\mu$ M (*R*)-3-hydroxytetradecanoyl-ACP, 1.75 mM PAPS (Sigma), 50 mM Tris pH 7.9, and 2  $\mu$ M ST-TE (either OLS ST-TE, OLS ST <sub>$\Delta$ 5</sub>TE, or OLS ST and OLS TE *in trans*) for 3 hrs. The reaction was quenched with 10% formic acid and the conversion of loaded to holo ACP was quantitated by HPLC as described above.

Assays with the non-sulfonated ( $\beta$ -hydroxy) substrate were performed by incubating 1 mM (*3R*)-hydroxytetradecanoyl-CoA, 50 mM Tris pH7.9, and 40  $\mu$ M TE for 5 h. The reaction was quenched with an equal volume of 1M CH<sub>3</sub>CO<sub>2</sub>H, neutralized with 1M NaOH, and crotonyl-CoA was added as an internal standard. Hydrolysis was analyzed using a Luna C18 column (250 x 4.60 mm, 5  $\mu$ m, 100 Å, Phenomenex) with a linear gradient from 10% to 90% CH<sub>3</sub>OH/H<sub>2</sub>O (10 mM CH<sub>3</sub>CO<sub>2</sub>NH<sub>4</sub>) over 20 min.



**Figure 4.4 HPLC detection of OLS TE substrates and products**

The one-pot assay mixture was separated by reverse-phase HPLC resulting in peaks for SVP, holo ACP, CurM ST, *(R)*-3-sulfotetradecanoyl-ACP, *(R)*-3-hydroxytetradecanoyl-ACP, and OLS TE.

## Results

### *OLS TE activity*

To obtain a complete view of the process of terminal olefin production by the olefin synthase (OLS) we investigated the C-terminal OLS TE domain. The recombinant, excised OLS TE is a dimeric protein like the related CurM TE. OLS TE is also highly stable, in contrast to the poor stability of the excised CurM TE, which precluded extensive mutagenesis to probe function. Thioester hydrolysis was assayed with  $\beta$ -sulfo-acyl-ACP and  $\beta$ -hydroxy-acyl-ACP substrates (Table 4.4), using the CurM ACP (27). Following reaction with OLS TE, ACP products were analyzed by HPLC (Table 4.4, Figure 4.4). We probed acyl chain length ( $C_{12}$ ,  $C_{14}$ ),  $\beta$  substituent ( $SO_4^-$ , OH),  $\beta$ -carbon chirality, and the substituent at C5 (methoxy, H). OLS TE, similar to CurM TE, was specific for  $\beta$ -sulfate substrates, an uncommon specificity among TEs (Table 4.4, (27)). OLS TE was also highly specific for the *R* configuration of the  $\beta$ -sulfate group, consistent with the configuration predicted for the natural substrate (118). In contrast with OLS ST, which does not tolerate substrates with a C5-methoxy group (121), OLS TE accepts substrates both with and without a methoxy at the C5 position. The results of OLS TE activity are consistent with our model of substrate binding in the CurM TE active site.

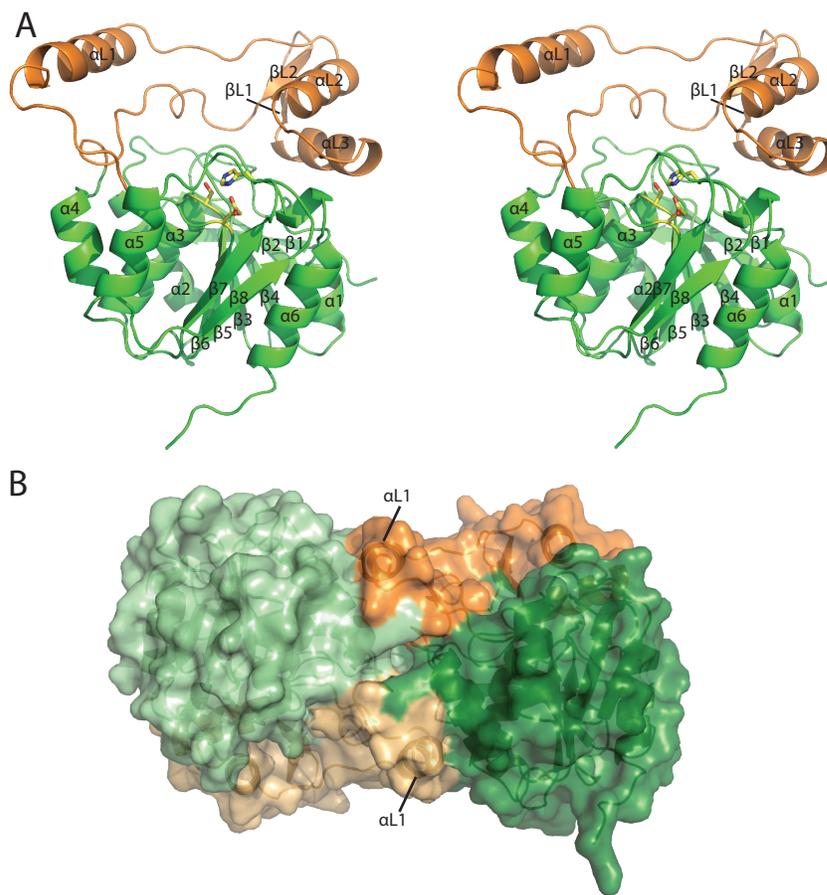
**Table 4.4 Substrate preference of OLS TE**

	Percent Hydrolyzed <sup>a</sup>		
	$\beta$ -sulfate substrates	$\beta$ -hydroxy substrates	No Enzyme
( <i>R</i> )-3-hydroxydodecanoyl-ACP	78 $\pm$ 4	4.1 $\pm$ 0.1	3.9 $\pm$ 0.2
( <i>R</i> )-3-hydroxytetradecanoyl-ACP	94.2 $\pm$ 0.1	3.45 $\pm$ 0.04	4.5 $\pm$ 0.1
(3 <i>R</i> )-3-hydroxy-5-methoxytetradecanoyl-ACP	36.1 $\pm$ 0.4	7 $\pm$ 1	6.6 $\pm$ 1.6
(3 <i>S</i> )-3-hydroxy-5-methoxytetradecanoyl-ACP	6.1 $\pm$ 0.2	5.0 $\pm$ 0.2	4.73 $\pm$ 0.02

<sup>a</sup>Raw HPLC chromatogram peak area for the ACP-loaded substrates and holo ACP products were used to calculate fraction of substrate hydrolyzed.  $\beta$ -sulfo-acyl-ACP substrates were generated by reacting  $\beta$ -hydroxy-acyl-ACP substrates with CurM ST. TE reactions were quenched after 5 min. Data are mean  $\pm$  standard deviation from duplicate experiments.

### Structure of OLS TE

The 1.7-Å crystal structure of OLS TE was solved by molecular replacement using CurM TE as a search model (Table 4.1). Continuous electron density is seen from the N-terminus to three residues before the C-terminus of the polypeptide (residues 1-250 of OLS TE correspond to 2435-2720 of the full length OLS polypeptide). The asymmetric unit of the crystal contains one dimer, providing two independent views of OLS TE (Figure 4.5).



**Figure 4.5 Structure of OLS TE**

**A.** OLS TE polypeptide. The stereo ribbon diagram is colored as green for the  $\alpha/\beta$  hydrolase core, as orange for the lid, and with the catalytic triad active site in stick (yellow C). Secondary structure is labeled. **B.** OLS TE dimer ( $1435 \text{ \AA}^2$  buried surface area per subunit). Backbone trace and surface representation of OLS TE dimer viewed along the molecular dyad. Monomers are colored dark green (core) and dark orange (lid) (right monomer) and light green (core) and light orange (lid) (left monomer). In the main dimer contact,  $\alpha L1$  interacts with the core of the partner subunit.

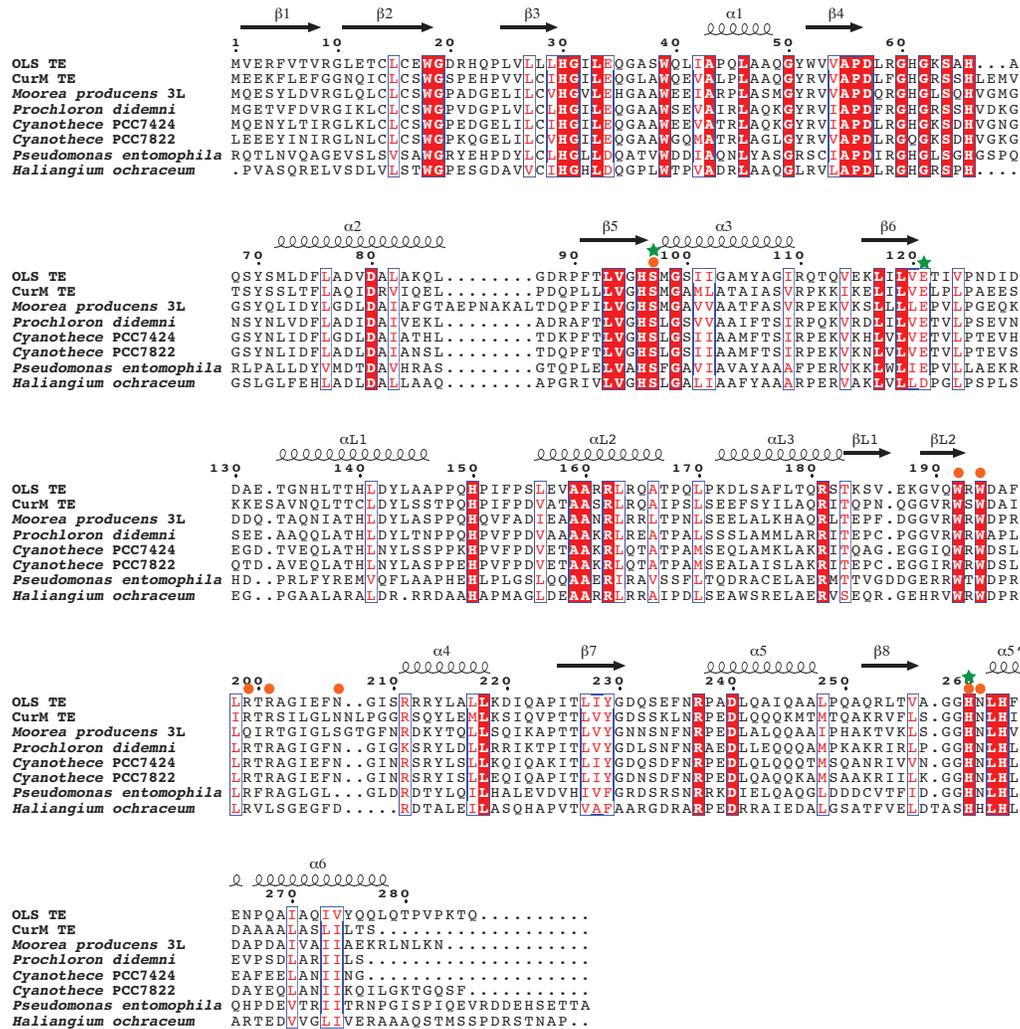
As expected, OLS TE and CurM TE have the same fold and position of the catalytic triad (0.82 Å RMSD for 197 C $\alpha$  atoms, 47% sequence identity, Figure 4.6A). OLS TE retains the same fixed-open lid as CurM TE (Figure 4.7, Figure 4.8A), which greatly differs from other  $\alpha/\beta$  hydrolase lids (99). The same lid conformation in decarboxylating TEs from two separate pathways (PKS and OLS) supports the previous view that the lid is anchored open and does not undergo a conformational change to close during catalysis (99). In both OLS and CurM TE, the lid is fixed open by a similar dimer interface unique to the decarboxylating TEs (Figure 4.5B, Figure 4.8B). Comparison of the dimer interfaces shows that they are able to undergo a slight (9°) flexure (Figure 4.8B). The dimer and lid architecture create a large and accessible active-site cleft lined with conserved residues (Figure 4.7B, (99)).

#### *OLS TE active site*

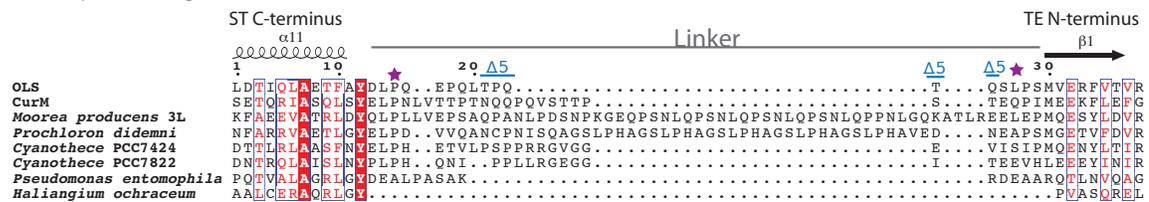
The highly stable OLS TE domain presents an opportunity to probe conserved features of the unusual large and open active site cavity (Table 4.5). Amino acid substitutions of the catalytic triad (S97A/C, H260R) support the identification of the active site residues. The Arg substitution at His260 and the Ala substitution at Ser97 decreased activity 70-fold and 40-fold respectively. Remarkably, a Cys nucleophile is nearly as effective as the wild type Ser. Alanine substitutions at conserved asparagines (Asn207 and Asn261) near the active site resulted in activity almost indistinguishable from WT.

Arg205 was identified as the sulfate recognition element in CurM TE (99); the analogous amino acid in OLS TE is Arg201 (Figure 4.7B). To test the role of OLS TE Arg 201 in  $\beta$ -sulfate recognition, Ala and Lys substitutions were made by site-direct mutagenesis. A significant reduction in activity was observed for R201A (15% of WT) (Table 4.5). The activity was partially rescued with the Lys substitution (32% of WT) (Table 4.5), which bears the same positive charge as Arg and should be able to interact with the sulfate. To demonstrate that Arg 201 is involved in the activity of only  $\beta$ -sulfate substrates, the variants were tested in the much slower reaction with the  $\beta$ -hydroxy substrate, where they performed identically to WT (Table 4.6). Arg201 is important for activity only for  $\beta$ -sulfate substrates and not for  $\beta$ -hydroxy substrates and thus is assumed to play a role in  $\beta$ -sulfate specificity.

## A Sequence alignment of decarboxylating TEs



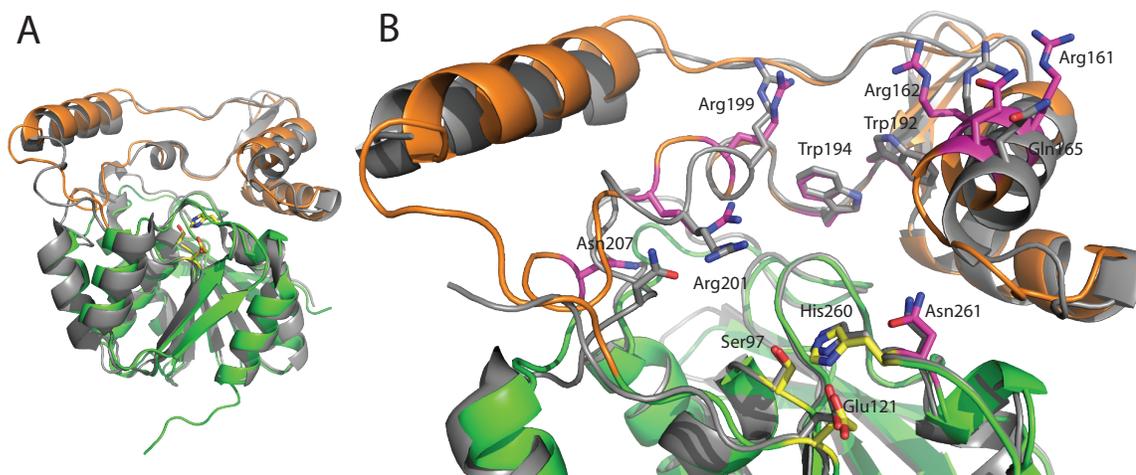
## B Sequence alignment of ST to TE linker



**Figure 4.6 Sequence alignment of decarboxylating TEs**

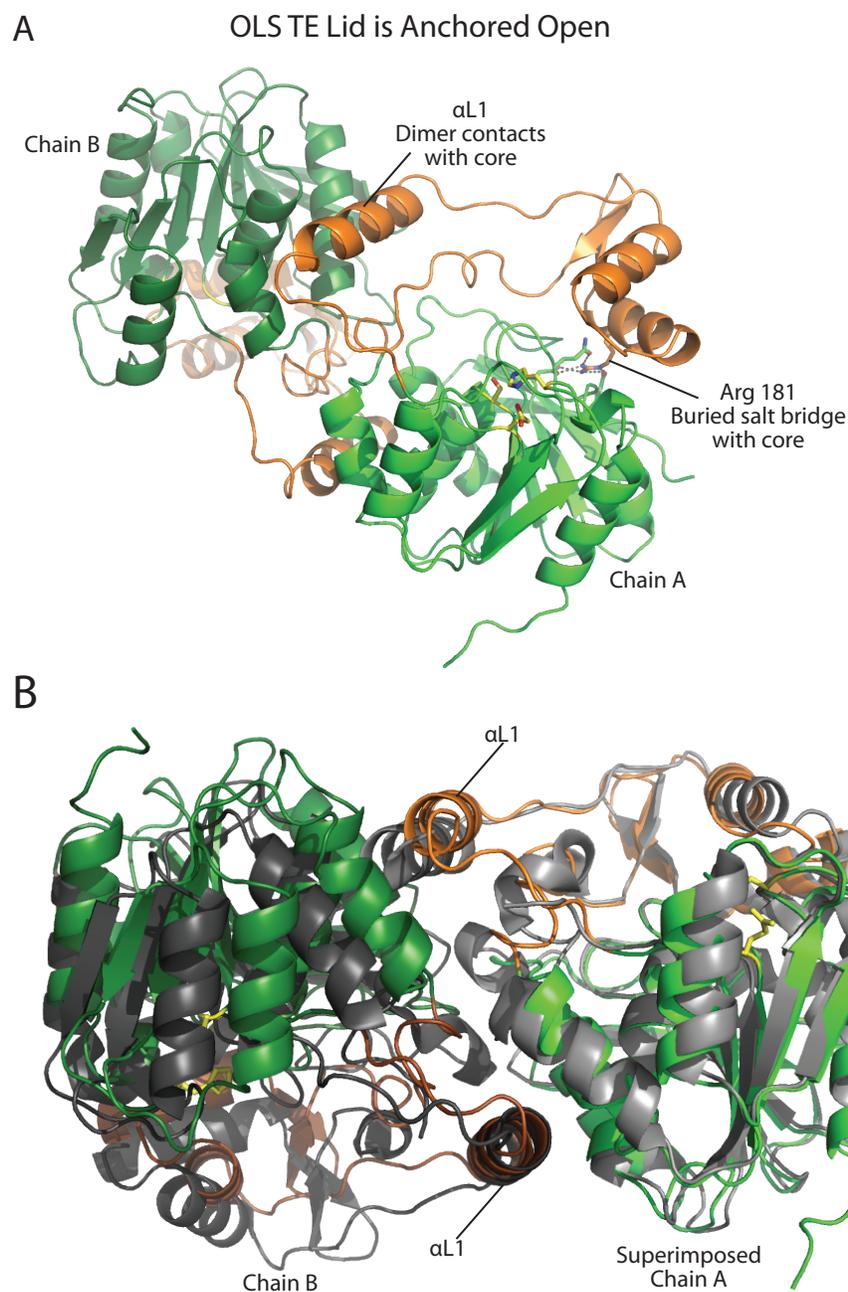
A. OLS TE and CurM TE are aligned with putative decarboxylating TEs from open reading frames encoding tandem ACP-ST-TE tridomains. Invariant amino acids are highlighted in red, conserved residues are printed in red. Green stars indicate amino acids in the active site and orange circles indicate those amino acids probed by site-directed mutagenesis. Secondary structures are indicated above the alignment. Sequence alignment was performed by MUSCLE (91) and the figure was prepared using ESPript

(95). GenBank entries are: *Moorea producens* (CurM\_ST) (GenBank ACV42478), *Synechococcus* PCC 7002 (OLS\_ST) (YP\_001734428), *Cyanothece* PCC 7424 (YP\_002377174), *Cyanothece* PCC 7822 (ZP\_03153601), *Moorea producens* 3L (ZP\_08425908), *Prochloron didemni* (AEH57210), *Pseudomonas entomophila* L48 (YP\_610919), *Haliangium ochraceum* DSM 14365 (YP\_003265308). **B.** Alignment of the linker between the ST and TE from open reading frames containing ACP-ST-TE tridomains. The last secondary structure element of the ST ( $\alpha_{11}$ ) and the first from the TE ( $\beta_1$ ) are depicted above the sequence. The linker varies in length between 0 and 60 amino acids and is not conserved. Purple stars indicate the last ordered amino acid of the excised OLS ST structure (121) and the first ordered amino acid in the excised OLS TE structure. Blue bars labeled  $\Delta 5$  indicate the residues deleted in OLS ST $\Delta 5$ TE.



**Figure 4.7 Comparison of OLS TE and CurM TE**

**A.** Structural alignment of OLS TE and CurM TE. OLS TE is colored as in Figure 4.5A and CurM TE is colored in gray. OLS and CurM TE show the same  $\alpha/\beta$  hydrolase core and lid structure (0.82 Å RMSD for 197 C $\alpha$  atoms). **B.** Zoom view of OLS TE and CurM TE superposition. OLS TE is colored as in A with important residues near the active site and lid represented in stick (magenta C). CurM TE is represented in gray. Many conserved amino acids line the active site cleft of the decarboxylating TEs.



**Figure 4.8 OLS TE lid**

**A.** OLS TE lid is fixed open. OLS TE core is colored green and lid is colored orange. Active site is shown in stick form (yellow). One side of the lid is fixed open by a buried Arg salt bridge and the other by dimer interface contacts. **B.**  $\alpha$ L1 associates with the core of the partner subunit. “Chain A” of OLS TE and CurM TE dimers were superimposed (0.82 Å RMSD for 197 Ca atoms). The dimer interface is conserved between OLS TE and CurM TE and exhibits flexure of  $9^\circ$ . OLS TE is colored as in Figure 4.5B, CurM TE is colored in light and dark gray.

**Table 4.5 Activity of OLS TE variants**

OLS TE variant	% Activity <sup>a</sup>	Location
Wild Type	100	
S97A	2.9 ± 0.5	Catalytic Triad
S97C	85 ± 4	Catalytic Triad
W192A	83 ± 1	Lid
W194A	101 ± 2	Lid
R201A	15 ± 2	Lid-Core loop
R201K	32 ± 3	Lid-Core loop
N207A	95 ± 3	Lid-Core loop
N261A	93 ± 11	Core
H260R	1.4 ± 0.6	Catalytic Triad
R161A	98 ± 1	Lid
R162A	24 ± 1	Lid
R199A	69 ± 2	Lid
R161A/R162A	22 ± 2	Lid
R161A/R199A	64 ± 3	Lid
R162A/R199A	5 ± 4	Lid
R161A/R162A/R199A	<0.1	Lid
R161D/R162D/R199A	<0.1	Lid
NO TE	0	

<sup>a</sup>Raw HPLC chromatogram peak areas for the substrate ((*R*)-3-sulfotetradecanoyl-ACP) and product (holo ACP) were used to calculate the fraction of substrate hydrolyzed. The activity of each mutant was background subtracted from the no TE control and normalized to the wild type within each replicate. Mean ± standard deviation from triplicate experiments are shown

**Table 4.6 OLS TE activity with  $\beta$ -hydroxy substrates**

OLS TE variant	% Activity <sup>a</sup>
WT	100
R201A	103 ± 6
R201K	104 ± 7

<sup>a</sup>The non-sulfonated CoA substrate, (*3R*)-hydroxytetradecanoyl-CoA, was reacted with OLS TE for 5 h. The reaction was monitored by HPLC analysis of the CoA species. The fraction of substrate hydrolyzed was calculated from the ratio of peak areas of the substrate ((*3R*)-hydroxytetradecanoyl-CoA) and product (CoA). The negative-control catalytic-triad substitution (H260R) was used for background subtraction, and results were normalized to the wild type within each replicate. Means ± standard deviations of triplicate experiments are shown.

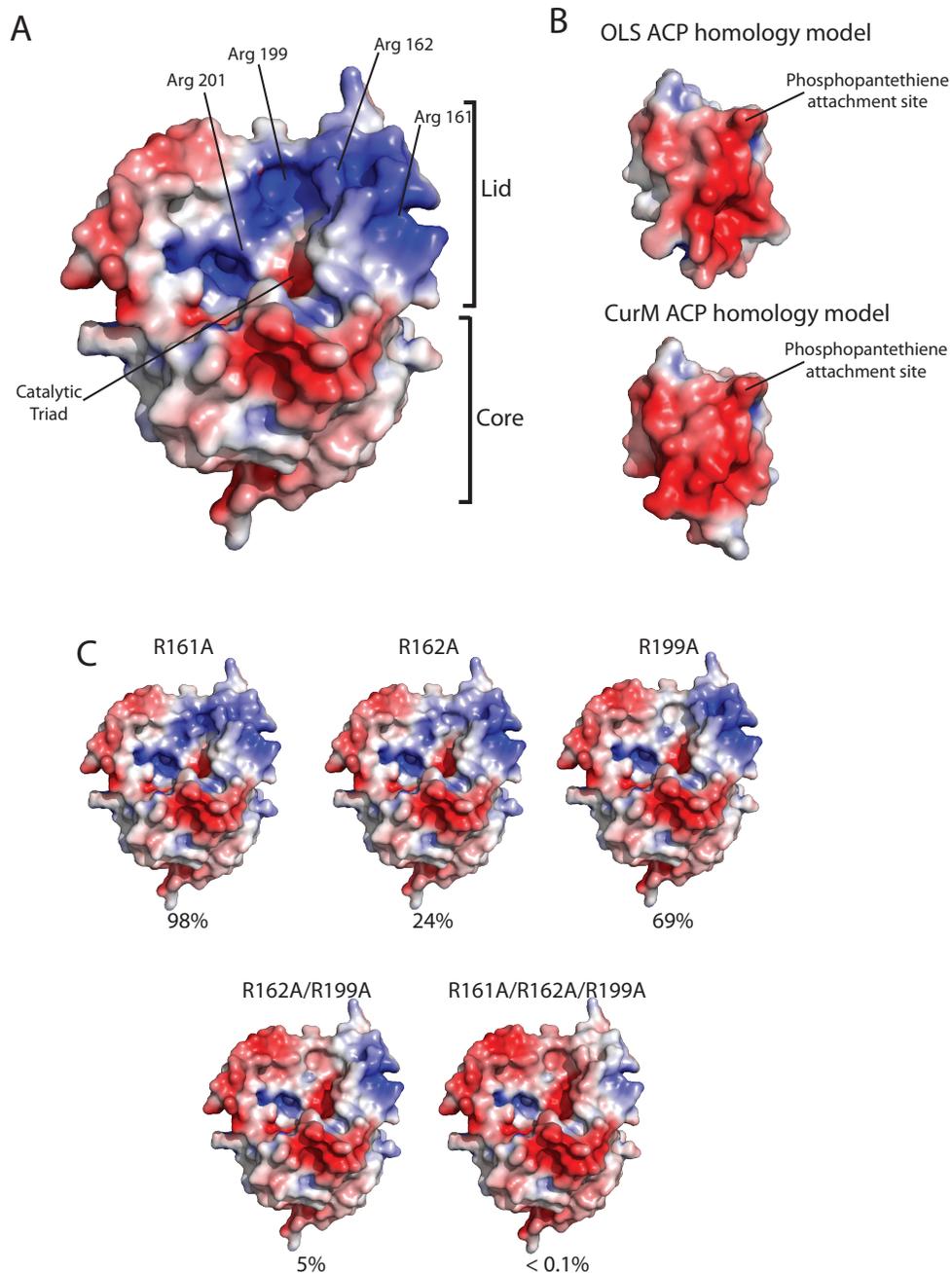
### *OLS TE lid*

Additional conserved amino acids in the unique open lid were probed by site-directed mutagenesis of OLS TE. Due to problems with protein solubility, the lid substitutions could not be investigated in CurM TE. Alanine substitutions of two conserved tryptophans, Trp192 and Trp194 (Figure 4.7B), resulted in 83% and 101% of WT activity respectively (Table 4.5). These drastic substitutions would presumably deepen the active-site cleft formed by the lid but would not alter the electrostatic environment or the contacts that fix the lid in an open position. Activity similar to WT for the Trp variants suggests that the openness of the lid is more important than the structure of the cleft that it forms. We obtained crystals of the Trp variants in the same condition as WT, but they did not diffract beyond 10Å, suggesting that the Trp substitutions may alter the lid structure.

In contrast to the Trp variants, substitutions that altered the electrostatic surface of the lid greatly affected activity. Three basic amino acids in the lid (Arg161, Arg162, and Arg199) create a pronounced electrostatic gradient and positive surface patch (Figure 4.9A), which is a potential ACP docking site. ACP interaction sites on other PKS or FAS domains often include a positive patch with potential interactions facilitated by arginines (122, 123). Homology models of OLS and CurM ACPs show a significant negative surface around the phosphopantetheine attachment site (Figure 4.9B). The positively charged surface at the top of the lid is approximately 18 Å away from the active site, similar to the length of the phosphopantetheine arm, which would extend from the ACP to the catalytic triad. Additionally, the surrounding negative surface on OLS TE creates an electrostatic gradient, which may guide the ACP to the correct docking location.

To probe the role of the positive patch, we made substitutions at three OLS TE amino acids that appear important for patch electrostatics. Initially, single alanine substitutions were made for each of the three arginines in the positive patch (R161A, R162A, and R199A, Table 4.5). The Ala substitution at Arg161 resulted in no change in activity and little change in the positive surface patch or electrostatic gradient (Figure 4.9C). On the other hand, activity was reduced 4-fold by Ala substitution at Arg162 (Table 4.5). Arg162 is conserved in decarboxylating TEs (Figure 4.6A) and is located on the inside of the cleft, closer to the active site than is Arg161 (Figure 4.7B, Figure 4.9A).

The Ala substitution at Arg199 had a modest effect, with activity reduced 30% from WT levels. These results are somewhat at odds with electrostatic calculations for the single amino acid variants (Figure 4.9C), which predicted that no single substitution would significantly alter the positive character of the TE lid. However, two of the double substitutions (R161A/R162A, R162A/R199A) and the triple substitutions (R161A/R162A/R199A, R161D/R162D/R199A) had substantially reduced activity (Table 4.5) and significantly altered the surface electrostatics (Figure 4.9C). The activity reduction is more than additive for the double R162A/R199A substitution and for the triple substitutions (Table 4.5). Loss of positive charge at all three sites reduced activity to below detectable levels. The mutagenesis results support the identification of an ACP recognition site at the top of the lid and suggest that, in addition to creating an open active site, the lid of decarboxylating TEs plays an active role in substrate recognition and binding.



**Figure 4.9 OLS TE electrostatics**

**A.** Electrostatics of OLS TE surface. Positive areas are represented in blue (+5 kT), neutral in white, and negative in red (-5 kT). Orientation is the same as Figure 4.8A and the locations of key amino acids and the catalytic triad are indicated. Electrostatic surface was generated using APBS (124). **B.** Electrostatic surface of a homology model of OLS and CurM ACPs. Homology model was created with SWISS-MODEL (125). **C.** *In silico* electrostatic surface analysis of substitutions at the putative ACP docking site. Activity decreases as the positive patch is removed.

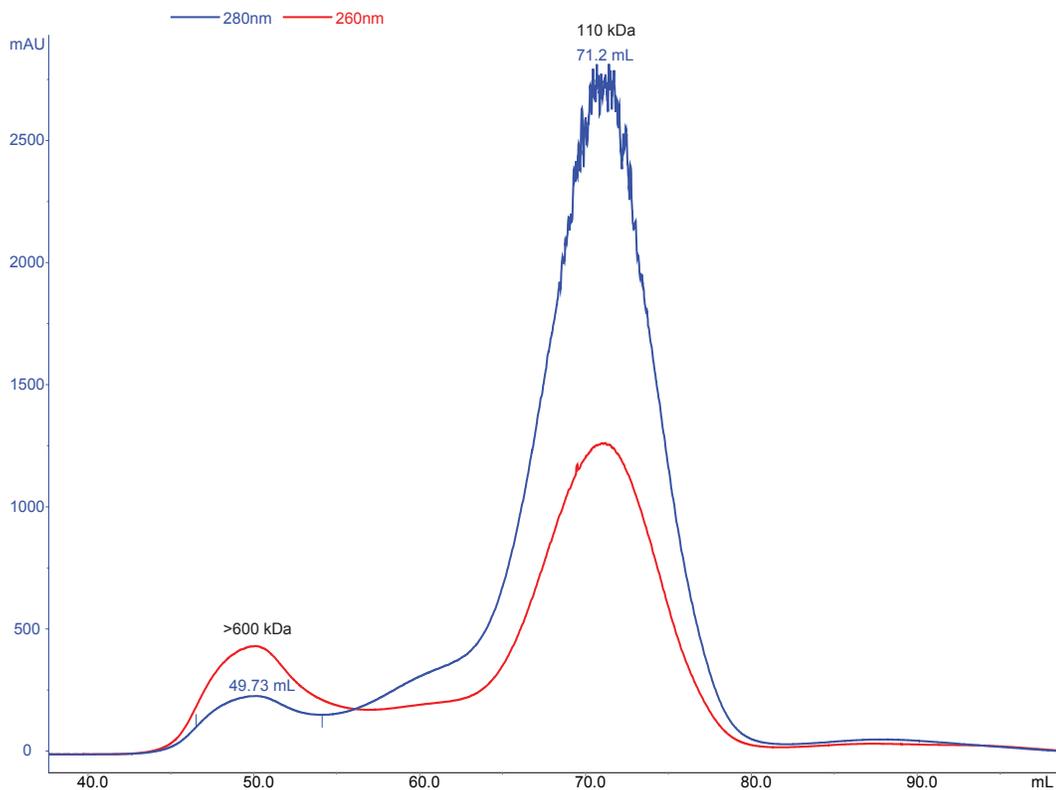
### *OLS TE in the full-length polypeptide*

The crystal structures of the excised OLS ST and OLS TE domains identify the linker between the two domains (Figure 4.6B). The linker includes 17 amino acids (residues 2418 – 2434) in the OLS polypeptide. Three linker residues are ordered in each of the OLS ST and TE structures. The sequence of the linker is not conserved in amino acid content and varies from 0-60 amino acids in length in other ACP-ST-TE sequences (Figure 4.6B).

OLS TE is chemically dependent upon OLS ST; OLS TE requires OLS ST to sulfonate the  $\beta$ -hydroxy, which the TE subsequently acts upon (Figure 4.1). Since OLS ST is also naturally tethered to the TE, the question arises how OLS TE might interact with OLS ST in the context of the full polypeptide. To probe the interaction of the naturally tethered ST and TE, the OLS ST-TE didomain was recombinantly expressed and purified as a dimer (Figure 4.10). The 3.1-Å crystal structure was solved by molecular replacement using the OLS ST monomer (121) and OLS TE dimer as a search model (Table 4.1).

### *Structure of OLS ST-TE didomain*

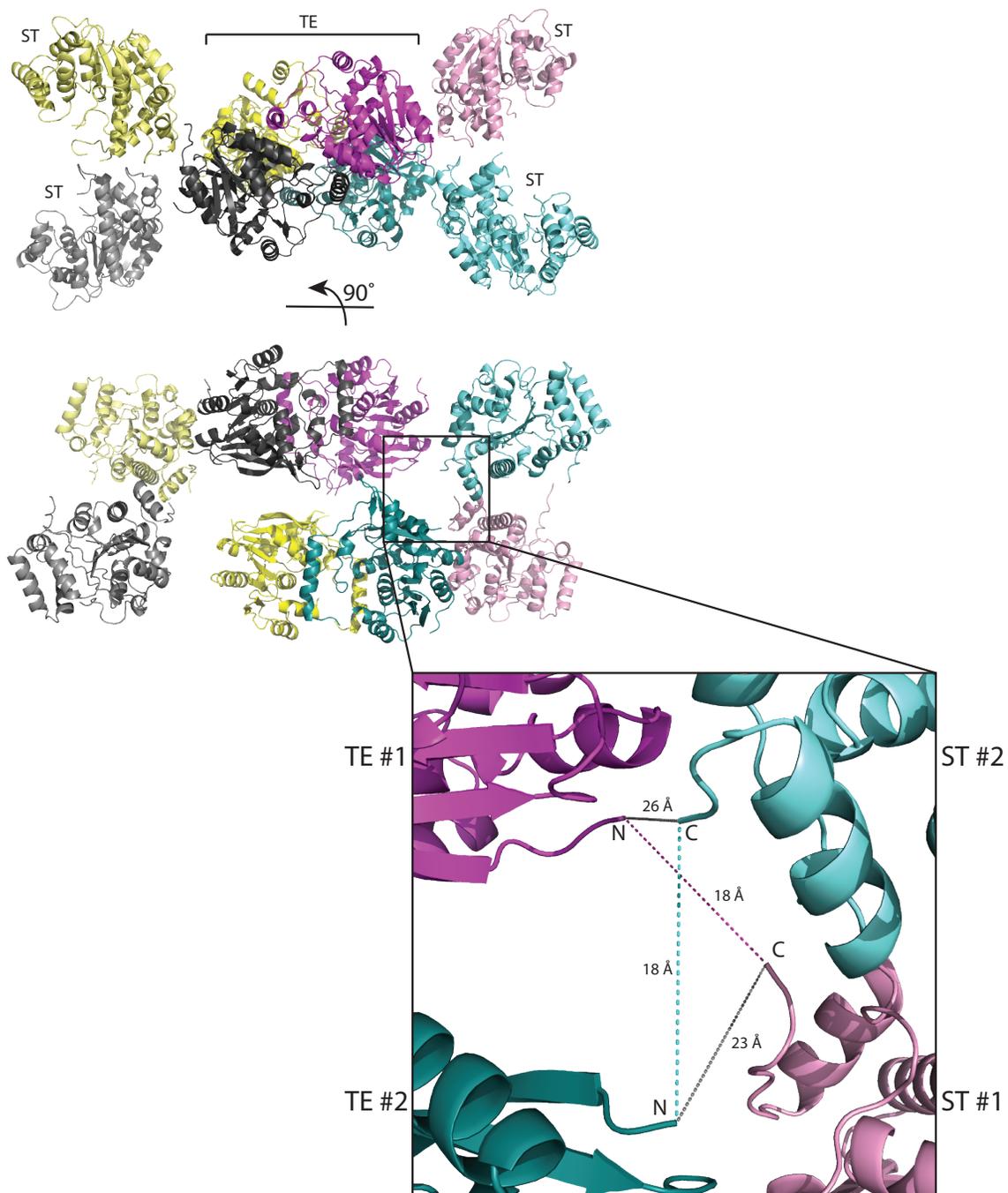
Crystals of OLS ST-TE contain four didomains in the asymmetric unit and four polypeptides of both OLS ST and OLS TE were found in the structure (Figure 4.11). The domain structures within the didomain are identical to the higher-resolution structures of the individual excised domains (ST RMSD = 0.16 Å for 253 C $\alpha$  atoms, TE RMSD = 0.18 Å for 261 C $\alpha$  atoms). It was not possible to trace a continuous chain from ST to TE; 8-10 amino acids of the ST-TE linker were disorder in the crystal. In order to identify which ST and TE domains belong to the same polypeptide we looked for distances from ST C-terminus to TE N-terminus that were less than 27 Å (~3 Å per amino acid). Two ST-to-TE junctions fall within the 27-Å requirement, causing ambiguity as to which ST and TE are connected on the same polypeptide (Figure 4.11). It is possible that the crystal is statistically disordered and both connections exist in the crystal. Chain assignments were selected based on the shortest distance between ST C-termini and TE N-termini. The disordered linker suggests a flexible, dynamic tether between the two domains.



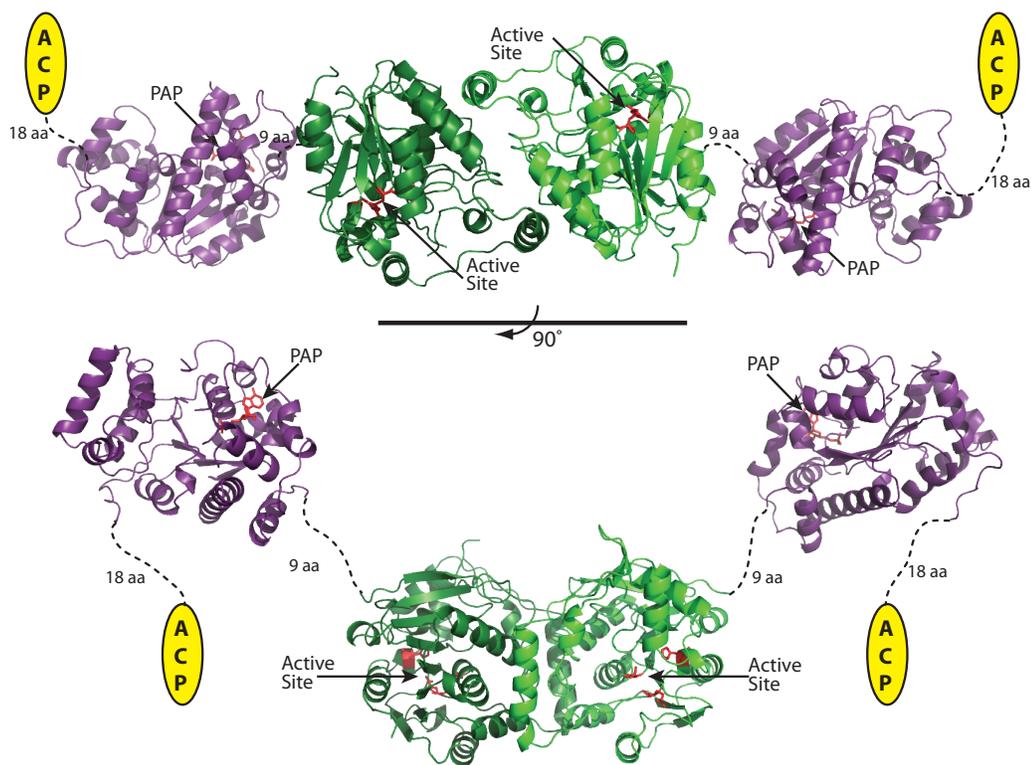
**Figure 4.10 Size exclusion analysis of OLS ST-TE**

In the elution profile from a HiLoad 16/60 Superdex 200 size exclusion column, OLS ST-TE has an apparent molecular mass of 110 kDa, similar to the calculated molecular weight of an OLS ST-TE dimer (135 kDa).

OLS ST-TE contains the same TE dimer interface as observed in the structures of the excised OLS TE and CurM TE domains (Figure 4.12). The ST domains flank the TE dimer and create an extended dimer structure. The active sites of the ST and TE are over 50 Å apart (Figure 4.12). There is no contact of the ST and TE domains on the same polypeptide apart from the linker, consistent with the hypothesis that the ST and TE function as separate tethered domains. To preclude the possibility that a biologically significant interaction is made by a crystal contact instead of by domains on the same polypeptide, the other ST-to-TE interfaces in the crystal structure were examined. The only significant interface (823 Å<sup>2</sup>) (Figure 4.13) does not appear to be biologically relevant since the termini are too far apart (60 Å) for fusion of the domains, and the interaction is made by non-conserved polar amino acids.

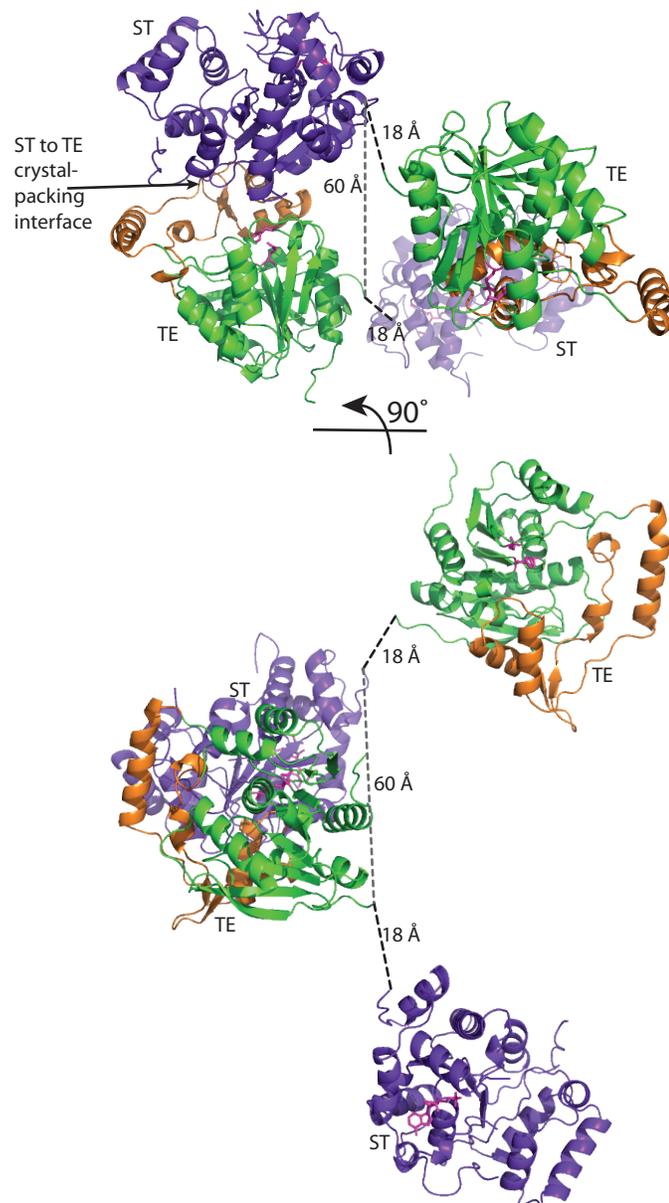


**Figure 4.11 Contents of the asymmetric unit for the OLS ST-TE crystal structure**  
 Two ST-TE dimers in the asymmetric unit are shown. ST (lighter) and TE (darker) domains that belong to the same polypeptide are shown with matching colors. Zoom view is of the region connecting the ST domain to the TE domain. Linkage of the same polypeptide is represented by the colored (magenta or teal) dotted line and the alternative linkage is represented by the grey dotted line.



**Figure 4.12 OLS ST-TE dimer**

OLS ST is represented in purple and the OLS TE in light and dark green. The PAP and TE catalytic triad are shown in stick form (magenta). Dashed lines indicate the disordered amino acid linker between the ST and TE. The place of attachment to ACP and the rest of the OLS polypeptide is cartooned.



**Figure 4.13 Crystal contact in OLS ST-TE structure**

OLS TE domain (green core, orange lid, magenta stick catalytic triad) makes a crystal contact with OLS ST domain of the same chain (purple, magenta stick PAP). The C-terminus of the ST and N-terminus of the TE are over 60 Å apart, too distant for the interaction to occur *in cis*. The distances between ST and TE on the same polypeptide are shown.

### *OLS ST-TE activity in cis and trans*

To probe a potential function of the fusion of OLS TE to OLS ST, the activity of OLS ST-TE was assayed with the domains *in trans*, naturally linked *in cis*, and with an *in cis* variant lacking five amino acids in the ST-TE linker (OLS ST $_{\Delta 5}$ TE, deletion of residues 2426-2430 from the OLS polypeptide (Figure 4.6B)) (Table 4.7). The three reaction mixtures were incubated with the substrate (*R*)-3-hydroxytetradecanoyl-ACP and analyzed by HPLC. Even though the reactions did not go to completion, no sulfonated intermediate was detected, only  $\beta$ -hydroxy-ACP substrate and holo-ACP product. In the *in vitro* assay conditions, the limiting reaction appears to be the ST sulfonation, which is consistent with the reaction times reported for OLS ST (121) and OLS TE (Table 4.4). Compared with the *in trans* reaction, activity was not significantly different for either *in cis* constructs (91% of *in trans*). There was no difference in activity of the ST-TE didomain with the natural linker and the  $\Delta 5$  deletion. The *in cis* and *in trans* comparison supports the conclusion from the didomain structure that OLS ST and TE act independently and are flexibly tethered as seen in the crystal structure and also suggests that the sequential ST and TE reactions do not necessarily occur on domains that reside on the same polypeptide.

**Table 4.7 Comparison of OLS ST-TE activity *in cis* and *in trans***

Construct(s)	% Activity <sup>a</sup>	
OLS ST + OLS TE	100	<i>in trans</i>
OLS ST-TE	91 $\pm$ 9	<i>in cis</i>
OLS ST $_{\Delta 5}$ TE	91 $\pm$ 1	<i>in cis</i>
No Enzyme	0	

<sup>a</sup>Raw HPLC chromatogram peak areas for the substrate ((*R*)-3-hydroxytetradecanoyl-ACP) and product (holo ACP) were used to calculate the fraction of substrate hydrolyzed. The activity was background subtracted from the no enzyme control and normalized to OLS ST + OLS TE *in trans* within each replicate. Mean  $\pm$  standard deviation from triplicate experiments are shown

## Discussion

### *OLS TE substrate specificity*

OLS TE, similar to CurM TE (27), was specific for  $\beta$ -sulfate substrates. Analysis of the CurM TE structure predicted that the specificity comes from the open active site, which is unable to efficiently position substrates for catalysis except for substrates that can form the sulfate-to-Arg205 interaction (99). OLS TE possesses all the structural features predicted or shown to play a role in  $\beta$ -sulfate specificity in CurM TE. The lid is fixed in an open position, aided by the dimer interface, and creates an active site exposed to solvent. OLS TE Arg201 (analogous to Arg205 in CurM TE (99)) is conserved and positioned similarly near the catalytic triad (Figure 4.7B). OLS TE Arg201 was important for catalysis with  $\beta$ -sulfate substrates (Table 4.5), but not for the slower reaction with  $\beta$ -hydroxy substrates (Table 4.6), and thus plays a role in sulfate specificity.

In addition to accepting the native acyl chain substrates, OLS TE accepted substrates bearing a C5-methoxy group. The spacious, solvent-accessible active-site cleft, which lacks specific substrate-protein interaction past the substrate  $\beta$  position, accommodates substrates with non-native functional groups. The ability to accept functionalized substrates contrasts with OLS ST, which does not accept substrates with a C5-methoxy and binds substrates by a narrow channel covered by a flap (121). OLS TE showed a distinct preference for substrates with the *R*  $\beta$ -sulfate configuration (predicted to be the native substrate (118)). The same preference was observed for CurM TE (27). The geometry of CurM Arg205 and the catalytic triad indicate that steric clashes would result when binding the *S*  $\beta$ -sulfate isomer (99). A similar arrangement of Arg201 and active site is present in OLS TE.

### *Decarboxylating TE lid*

A pronounced positive patch created by Arg161, Arg162, and Arg199 at the top of the lid is predicted to be an ACP recognition site. The OLS TE also has an electrostatic potential gradient surrounding the positive patch, which can help guide the ACP to a position for effective substrate delivery. Substitutions that alter patch electrostatics had a more than additive effect on activity ranging from small changes for single amino acid substitutions to dropping below detectable levels for the triple

substitutions. The positive patch and Arg199 and Arg162 are conserved in CurM TE, although the overall patch is less pronounced. The positive area also includes Arg201, and thus serves the double purpose of interacting with the negatively charged ACP and the  $\beta$ -sulfate, subsequently positioning two separate points of the substrate. Although the lid is unique to decarboxylating TEs, the larger TE superfamily exhibits lid structures that are used to modulate the TE function. It is no surprise then to see such an unusual lid on the TEs that catalyze such an unusual decarboxylative elimination reaction.

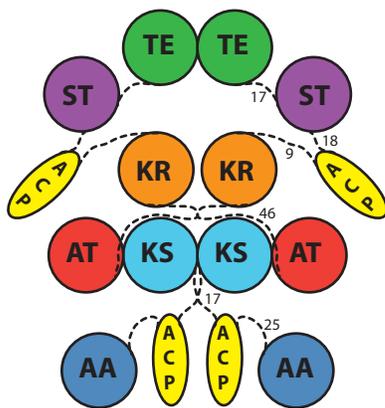
#### *OLS TE flexibly tethered to OLS ST*

OLS ST, immediately preceding OLS TE in the full polypeptide, produces the  $\beta$ -sulfate substrate for the OLS TE reaction. Surprisingly, from all of our experiments, we find no evidence for an interaction between the two domains. There is no contact between the ST and TE domains on the same polypeptide in the crystal structure (Figure 4.12) and no other crystal interfaces were determined to be biologically relevant. The linker between the two domains is disordered and lacks continuous electron density. The reactions *in cis* and *in trans* show no advantage to linked enzymes, and no change in activity was observed for a variant with a shortened linker. *In vitro*, it appears that the ACP reacts with an ST, releases, and diffuses to react with a TE that is not necessarily on the same polypeptide chain and there is no evidence for direct channeling between the two domains.

#### *Architecture of the OLS polypeptide*

Although the OLS domains N-terminal of OLS ST have not been characterized, much can be gathered about OLS architecture from studies of PKS and FAS domains and modules (Figure 4.14). The N-termini of OLS ST are on opposite ends of the extended dimeric ST-TE structure. The ST links to a monomeric ACP, which has N- and C-termini on the same end of the domain (18). The ACP leads to the presumably monomeric KR, which is fused to the KS-AT dimer (18, 126, 127). The monomeric ACP and AA (128, 129) domains are fused to the KS N-termini located at the bottom of the domain (126). The OLS dimer interface is predicted to consist of the KS dimer and the TE dimer (Figure 4.14). Securing the dimer with two domains on opposite ends of the 8-domain polypeptide should result in a compact arrangement. The organization of the

dimeric anchors should allow the two ACPs on each polypeptide to reach the catalytic domains that act upon the substrates they carry. The OLS architecture positions the putative ACP docking sites of OLS TE on the front and back sides of the TE dimer, positions which can be easily accessed by ACP while not being blocked by the KR.



**Figure 4.14 Configuration of full-length OLS polypeptide**

The OLS domains are depicted by circles and the linking regions by dashed lines. The TE and KS dimers and the KS-AT interface are shown. The linker length between domains is indicated.

#### *Olefin offloading chemistry*

The hydrocarbon-producing OLS shows great potential for biofuel production. The use of the common fatty acid metabolite as the initial substrate lends itself to facile engineering in a heterologous host. Indeed, the up-regulation of fatty acid production has already been engineered into *E. coli* (65) as well as in cyanobacteria (130). The ST-TE offloading strategy may also be used to produce the long chain hydrocarbons ( $C_{27}$ ,  $C_{29}$ , and  $C_{31}$ ) in *Botryococcus braunii*, which was shown to produce odd numbered hydrocarbons by  $\beta$ -activation and decarboxylation (56). The decarboxylative elimination route to hydrocarbons from fatty acids is unique from those already explored, such as the acyl-ACP reductase and aldehyde decarbonylase or cytochrome P450 systems (131).

Here we characterized the OLS TE and the interaction to OLS ST through the OLS ST-TE didomain. OLS TE, which performs the final step in hydrocarbon production, shows  $\beta$ -sulfate specific activity and retains the structural features identified as important for specificity in CurM TE. A positive patch on the lid is identified as a potential ACP recognition site. OLS TE is flexibly tethered to OLS ST and the two domains were shown to act independently.

## Chapter 5

### Structure and activity of DmmA, a marine haloalkane dehalogenase

This chapter was published Gehret JJ, Gu L, Geders TW, Brown WC, Gerwick L, Gerwick WH, Sherman DH, Smith JL (2012) Structure and activity of DmmA, a marine haloalkane dehalogenase. *Protein Sci* 21:239-248.

#### Summary

DmmA is a haloalkane dehalogenase (HLD) identified and characterized from the metagenomic DNA of a marine microbial consortium. Dehalogenase activity was detected with 1,3-dibromopropane as substrate, with steady-state kinetic parameters typical of HLDs ( $K_m = 0.24 \pm 0.05$  mM,  $k_{cat} = 2.4 \pm 0.1$  s<sup>-1</sup>). The 2.2-Å crystal structure of DmmA revealed a fold and active site similar to other HLDs, but with a substantially larger active site binding pocket, suggestive of an ability to act on bulky substrates. This enhanced cavity was shown to accept a range of linear and cyclic substrates, suggesting that DmmA will contribute to the expanding industrial applications of HLDs.

#### Introduction

Haloalkane dehalogenases (HLDs) remove halogens from alkanes by hydrolysis, producing an alcohol, a halide ion, and a proton. These dehalogenating enzymes are of long-standing interest for industrial applications such as degradation of environmental pollutants (69, 70) and biocatalysis (71). Newer applications include remediation of chemical weapons (72), biosensing (73, 74), and cellular imaging (75). The applications for HLDs are growing with knowledge of their functional and structural properties.

Each HLD has a unique substrate selectivity, which can include chlorinated, brominated, and iodinated alkanes of varying length. Some HLDs also act on ring-containing substrates, or those with alcohol, ether, nitrile, or alkene functional groups (132). Substrate selectivity is not easily predicted by sequence analysis. For HLDs of

known structure, the size and shape of the active-site cavity are better predictors of substrate preference than is phylogenetic analysis (132). The natural substrates are in general unknown.

HLDs are members of the  $\alpha/\beta$  hydrolase superfamily, with the active site located between a conserved core sub-domain and a more variable lid structure. The HLD active site departs from the typical  $\alpha/\beta$  hydrolase catalytic triad by the replacement of the common Ser nucleophile with an Asp and the addition of two halide-stabilizing residues (Trp and Trp/Asn) to form a “catalytic pentad” (133). The side chains of the catalytic pentad are identically positioned in the active site cleft of all HLDs of known structure, although some pentad residues vary in their locations in the primary sequence. The sequence-location difference of the catalytic pentad residues defines three evolutionary subfamilies of HLDs (134). Two structures are known from evolutionary subfamily I (DhlA (135-137) and DppA (138)), and four from subfamily II (DhaA (139, 140), LinB (141-144), DbjA (145), Rv2578 (146)). Several structures include bound substrates, products, or covalent intermediates, which together with many kinetic studies, led to a detailed reaction mechanism for this enzyme class (133).

DmmA is a putative haloalkane dehalogenase of subfamily II (134). DmmA was originally annotated as CurN, and presumed to be the final gene product of the curacin A biosynthetic gene cluster (2) from the marine cyanobacterium *Lyngbya majuscula* (now designated *Moorea producens* (4)). However, resequencing revealed the authentic 3' end of the gene cluster, which lacks *curN* (27), suggesting that *dmmA* was cloned as part of a chimeric cosmid from the original metagenomic DNA library (2). Herein we present the crystal structure and haloalkane dehalogenase activity of DmmA.

## **Materials and Methods**

### *Cloning, protein expression, and purification*

Gene fragments encoding DmmA were generated from cosmid pLM17 (2), inserted into the pET-24b vector, and verified by sequencing to form pDmmA<sub>long</sub> (encoding residues 1-341) and pDmmA<sub>short</sub> (encoding residues 44-341) (plasmids were made by Dr. Liangcai Gu). *E. coli* strain BL21(DE3) was transformed with the expression plasmid, grown at 37°C in 1 L 2xYT to an OD<sub>600</sub> of 0.5, cooled to 18°C,

induced with IPTG (final concentration 0.4 mM) and grown for an additional 18 hr. Selenomethionyl (SeMet) DmmA<sub>short</sub> was produced in the same strain in SeMet minimal media (147).

All purification steps were performed at 4°C. The cell pellet from 1 L of cell culture was resuspended in 40 mL Buffer A (20 mM Tris pH 7.9, 500 mM NaCl, 10% glycerol) plus 20 mM imidazole, lysed by sonication, and the soluble fraction loaded onto a 5-mL HisTrap Ni NTA column (GE Healthcare). DmmA was eluted with Buffer A with a linear gradient of 20 to 650 mM imidazole. DmmA was further purified by size exclusion chromatography with a HiPrep 16/60 Sephacryl S100 HR column (GE Healthcare) pre-equilibrated with Buffer A. DmmA was concentrated to 14 mg/mL, flash frozen in liquid N<sub>2</sub>, and stored at -80°C. SeMet DmmA<sub>short</sub> was purified as the wild type with addition of 2 mM DDT to Buffer A during the size exclusion step. Yields per 1 L culture were 150 mg for DmmA<sub>short</sub>, 80 mg for DmmA<sub>long</sub>, and 15 mg for SeMet DmmA<sub>short</sub>.

### *Crystallization*

Crystals of DmmA<sub>short</sub> (native and SeMet) grew by vapor diffusion at 20°C within 24 hrs from a 1:1 mix of protein stock (6 mg/mL DmmA<sub>short</sub> in Buffer A) and well solution (2.2 M sodium malonate pH 7.0, 5% glycerol). Crystals of DmmA<sub>long</sub> formed at 20°C within 24-48 hr from a 1:1 mix of protein stock (14 mg/mL DmmA<sub>long</sub> in Buffer A) and well solution (0.6 M lithium sulfate, 27.5% PEG3350, and 0.1 M BisTris pH5.5). Crystals of DmmA<sub>long</sub> were cyroprotected by a 5-10 s transfer to a mixture of well solution also containing 25% PEG400.

### *Data collection and structure determination*

Data were collected at the GM/CA-CAT beamline 23ID-D at the Advanced Photon Source (APS) at Argonne National Lab (Argonne, IL) (Table 5.1). A 2.2 Å single-wavelength anomalous diffraction dataset was collected at the wavelength of peak absorption at the selenium edge from a DmmA<sub>short</sub> SeMet crystal. A 2.9 Å dataset was collected from a DmmA<sub>long</sub> crystal. DmmA<sub>short</sub> data were processed using the HKL2000 suite (80) (Table 5.2, Table 5.3). DmmA<sub>long</sub> data were processed using MOSFLM (148)

**Table 5.1 Crystallographic Summary**

	<b>DmmA<sub>short</sub></b> <b>(SeMet)</b>	<b>DmmA<sub>short</sub> (SeMet)</b> <b>1,5-dibromopentane</b> <b>soaked</b>	<b>DmmA<sub>long</sub></b>
<b>Diffraction Data</b>			
Space group	<i>P6<sub>2</sub></i>	<i>P6<sub>2</sub></i>	<i>P3<sub>1</sub>21</i>
X-ray source	APS 23-ID-D	APS 23-ID-D	APS 23-ID-D
a, b, c (Å)	99.8, 99.8, 122.0	100.3, 100.3, 121.4	144.7, 144.7, 105.0
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Wavelength (Å)	0.97939	0.92003	0.97939
d <sub>min</sub> (Å)	2.20 (2.28-2.2) <sup>a</sup>	2.10 (2.18-2.10)	2.50 (2.50-2.64)
Avg I/σ <sub>1</sub>	13.2 (2.9)	16.7 (2.1)	10.4 (3.5)
R <sub>symm</sub> <sup>b</sup>	0.122 (0.549)	0.118 (0.75)	0.171 (0.607)
Completeness	100.0 (99.9)	99.7 (97.6)	100 (100)
Avg. redundancy	5.8 (5.2)	10.5 (7.6)	9.6 (8.6)
Unique reflections	35,082	40,262	44,165
<b>Refinement</b>			
Data range (Å)	86.39-2.20		
No. reflections	34,919		
R <sub>work</sub> /R <sub>free</sub> <sup>c</sup>	0.139/0.177		
RMS deviations			
Bonds (Å)	0.009		
Angles (°)	1.195		
Avg B-factors (Å <sup>2</sup> )			
Protein	23.4		
Ligand/Ion	53.3		
Water	35.0		
Ramachandran			
Allowed	99.66%		
Outliers	0.34%		
Number of Atoms			
Protein	4649		
Ligand/Ion	18		
Water	497		

<sup>a</sup>Outermost shell in parentheses.

<sup>b</sup>Including anomalous differences.

<sup>c</sup>The R<sub>free</sub> data set included a random 5% of reflections.

**Table 5.2 Scaling statistics for SeMet DmmA<sub>short</sub>**

Shell Limit (Å)		Average I	Avg. error	Chi <sup>2</sup>	Linear R-factor <sup>a</sup>	Square R-factor
Lower	Upper					
50.00	4.74	132.3	4.5	1.010	0.047	0.050
4.74	3.76	161.6	7.6	1.011	0.068	0.076
3.76	3.29	106.8	5.2	1.050	0.076	0.075
3.29	2.99	59.8	4.2	1.011	0.118	0.114
2.99	2.77	39.1	3.9	1.011	0.175	0.168
2.77	2.61	28.3	3.7	1.012	0.235	0.219
2.61	2.48	22.3	4.0	1.001	0.305	0.274
2.48	2.37	17.7	4.1	1.018	0.395	0.348
2.37	2.28	14.6	4.1	1.028	0.479	0.427
2.28	2.20	12.4	4.2	1.072	0.549	0.457
All	Reflections:	59.5	4.5	1.022	0.122	0.093

<sup>a</sup>Including anomalous differences.

**Table 5.3 Scaling statistics for SeMet DmmA<sub>short</sub> soaked with 1,5-dibromopentane**

Shell Limit (Å)		Average I	Avg. error	Chi <sup>2</sup>	Linear R-factor <sup>a</sup>	Square R-factor
Lower	Upper					
50.00	4.52	613.8	20.0	0.993	0.052	0.053
4.52	3.59	559.0	19.3	0.990	0.068	0.070
3.59	3.14	273.4	11.9	0.998	0.094	0.087
3.14	2.85	137.5	9.2	0.999	0.158	0.140
2.85	2.65	84.8	8.3	0.999	0.241	0.207
2.65	2.49	59.8	8.3	1.003	0.335	0.282
2.49	2.37	43.4	8.6	0.992	0.450	0.365
2.37	2.26	33.2	8.7	1.000	0.565	0.455
2.26	2.18	26.4	8.7	1.010	0.642	0.488
2.18	2.10	19.3	9.0	1.006	0.750	0.574
All	Reflections:	186.9	11.2	0.998	0.118	0.079

<sup>a</sup>Including anomalous differences.

**Table 5.4 Scaling statistics for DmmA<sub>long</sub>**

	Overall <sup>a</sup>	Inner Shell	Outer Shell
Low resolution limit (Å)	47.35	47.35	2.64
High resolution limit (Å)	2.50	7.91	2.50
R <sub>merge</sub>	0.171	0.055	0.607
R <sub>meas</sub>	0.180	0.058	0.646
R <sub>pim</sub>	0.058	0.019	0.219
Total number of observations	423211	14374	54726
Total number unique observations	44165	1505	6375
<I/σ <sub>I</sub> >	10.4	21.6	3.5
Completeness	100	99.5	100
Multiplicity	9.6	9.6	8.6

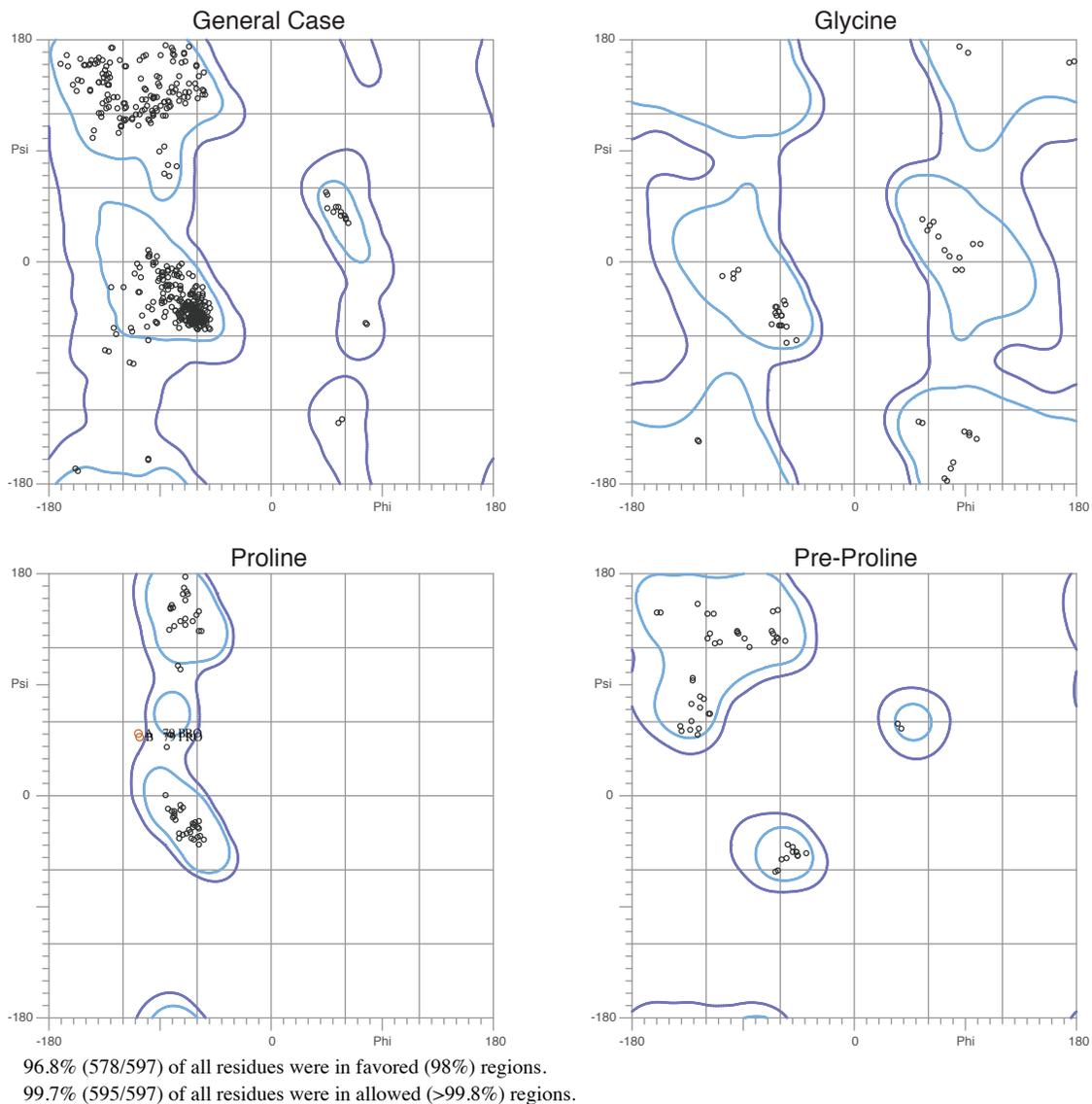
<sup>a</sup>Scaling performed with Scala (149).

**Table 5.5 Phasing statistics for DmmA<sub>short</sub>**

Sites found	22
NCS copies	2
Sites in each NCS group	9
NCS correlations	0.75
FOM <sup>a</sup>	0.47
Solvent content	0.52

<sup>a</sup>Final figure of merit of phasing

and SCALA (149) (Table 5.4). Determination of selenium atomic positions, experimental phasing, density modification phase refinement, and initial model building were performed using the programs SOLVE and RESOLVE (150, 84) (Table 5.5). 22 of the 24 expected selenium positions were identified. The model was finished manually in COOT (85). REFMAC5 (86) from the CCP4 suite (87) was used for refinement (Figure 5.1). The DmmA<sub>long</sub> structure was solved by molecular replacement in Phaser (81) using DmmA<sub>short</sub> as a search model. Model quality was evaluated with MolProbity (107). Figures were made using PyMOL (92) and hollow (151). Electron density was complete throughout the polypeptide chain for both polypeptides for DmmA<sub>short</sub> (residues 44-341) and DmmA<sub>long</sub> (residues 38-341). No density was observed for the C-terminal His tag of either DmmA variant or for the first 37 residues of DmmA<sub>long</sub>.



### Figure 5.1 Ramachandran analysis of DmmA

Ramachandran plots of final refined model of DmmA (PDB code = 3U1T). Plots were generated using MolProbity (<http://kinemage.biochem.duke.edu>) (89).

#### *Enzyme assay*

Activity was measured using a pH indicator dye-based colorimetric method (152, 153). Substrate (0.1-6.0 mM 1,3-dibromopropane (Aldrich)) was prepared in an indicator solution (20  $\mu\text{g}/\text{mL}$  phenol red, 20 mM  $\text{Na}_2\text{SO}_4$ , 1 mM EDTA, 1 mM HEPES pH 8.2). Reactions were initiated by addition of DmmA<sub>short</sub> (1  $\mu\text{M}$  final concentration, 100  $\mu\text{L}$  final volume). Absorbance at 550 nm was monitored at 5-second intervals during the

reaction to detect the decrease in pH generated by the release of H<sup>+</sup>. Parallel negative controls lacking the enzyme or substrate were used to correct for non-enzymatic dehalogenation or pH change. A standard curve relating absorbance at 550 nm to [H<sup>+</sup>] was used to convert absorbance to [H<sup>+</sup>]. The initial rates were fit using KaleidaGraph to the Michaelis-Menten equation ( $v_{\text{initial}} = ([\text{enzyme}] * [\text{substrate}] * k_{\text{cat}}) / (K_m + [\text{substrate}])$ ) to obtain steady state kinetic constants. Comparison with other substrates (1,6-dibromohexane (Acros), 1,6-dichlorohexane (Aldrich), 1,6-diiodohexane (Aldrich), bromocyclohexane (Aldrich)) were performed at 1 mM substrate concentration in duplicate and normalized to the percent of 1,3-dibromopropane activity.

#### *Accession numbers*

Coordinate and structure factors for DmmA<sub>short</sub> have been deposited in the Protein Data Bank with accession number 3UIT.

## **Results**

#### *Biological source of DmmA (formerly CurN)*

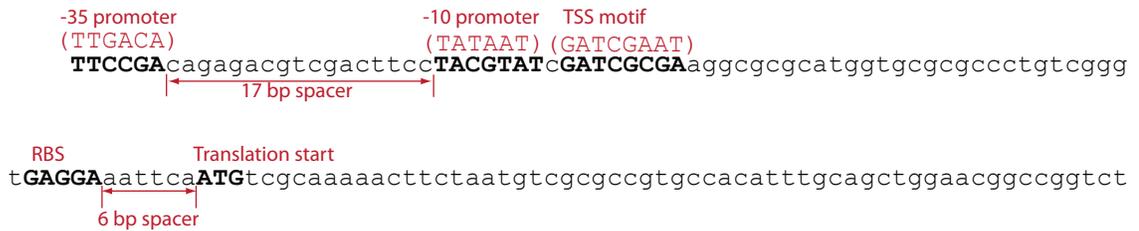
In the original curacin pathway sequencing (2), DNA encoding 64 amino acids at the CurN N-terminus overlapped with DNA encoding the CurM C-terminus in a different reading frame. This unusually long overlap and peculiarities of the CurM protein sequence led us to isolate a new cosmid, and to resequence the *curM-curN* region of the gene cluster, resulting in a corrected 3' terminus that lacked *curN* (27). In addition, no CurN coding sequence was identified by amplification from *M. producens* (formerly *L. majuscula*) genomic DNA, nor was CurN located in the recently sequenced *M. producens* 3L genome (5). Furthermore, the G + C content of *curN* (61%) contrasts with the rest of the curacin biosynthetic gene cluster (45.5%). The scaffolds used to assemble the *M. producens* genome ranged from 37% to 66% G + C content, with the great majority falling between 40% and 45% (5). Additionally, CurN (now referred to as DmmA (dehalogenase A from a marine microbe)) in accord with other dehalogenases of unknown biological function (132)) has high sequence similarity (~50%) to HLDs from other marine bacteria, although none are from the cyanobacterial phyla. From these data, we conclude that *curN* (*dmmA*) is most likely a product of one of the organisms that grew in close association with the *M. producens* field isolate whose metagenomic DNA cosmid

library was used to sequence the *cur* gene cluster. It is evident that the 3' end of the curacin gene cluster concatenated with *curN* (*dmmA*) genetic material from one of the heterotroph bacteria that was growing on the *M. producens* filament at the time of cosmid preparation.

Possible promoter sequences for *dmmA*

Consensus sequence in parenthesis

cagatctgtctgtgtagttggggttccccggagcatcctgtagttctgtgtattcatggaatTTTggaaca  
 aggattagcttggcaagaagtcgcacttctctagcagcacaagggtatcgggtagtggctcctgacttat  
 ttggtcacgggcgctcttctcatttggaaatgggtgacttcttatagttcactgacatttttagctcagata  
 gaccgggtaattcaggaattaccagaccaaccttggttgtagtaggtcattccatgggtgccatgctggc  
 aactgcatcatcggaacgggcctctcacactggatcgtttacgcaccgaggcggcgcaatccgaccg



ggttgccgtggcggccgggtgtgocggttgocgggcccagtcgcagccgcccgttcaactgccgatct  
 cgtcggaaattccattcgcacaaaaggacgggtcgaggtcgagggggcgacaatcgcctatggtgacgagggc  
 agtggtcagccagtgctttttctgcacggcaatccgacatcgtcctacctgtggcgcaacatcatcccata  
 tgtcgtcgtgcccgggtatcgcgcggtggcgccggatctgatcggcatgggcgacagcgcgaagcccgaca  
 tcgaatatcgggtgcaggaccacgtggcatatatggatggttttatagacgcgcttgggtctcgtatgacatg  
 gtccgtggtcatccacgattggggctcgggtgatcggcatgctcagcgcggttgaatcccgatcgggtggc  
 cgctgtcgccttcatggagggcgttgttccaccggcattgccaatgccaagttacgaggccatggggccgc  
 aactgggcccgtgtttcgtgatctgcgcacagcggatgtcggcgagaagatgggttctcgacggcaatttc  
 ttgtcagacgatattgcccagatgggctggtgctgggtccctgagcgaagcagagatggcagcctatcg  
 cgcgcggttcccagcgcgtcagagccggttgcccaccttgacgtggcgcgcaagtcctcggcggcg  
 agcccgcctttgcccgaagccgaggttctgaagaatggcgaatggctgatggcgagcccgataccgaaactg  
 ctgtttcatgcggaaccgggtgctttggctccgaagccggtggctcactacctgagcgaaaactgccc  
 tcttgagggtccgggttcggtggtgctggcacacatttctcgcaggaagatcacccgcacctgatcggtcagg  
 gcattgccgactggctgcccgcgaacaagccgcagcctcctgacaaaaggctgtcaggaggccgatcgtcac  
 gacctttacgggt

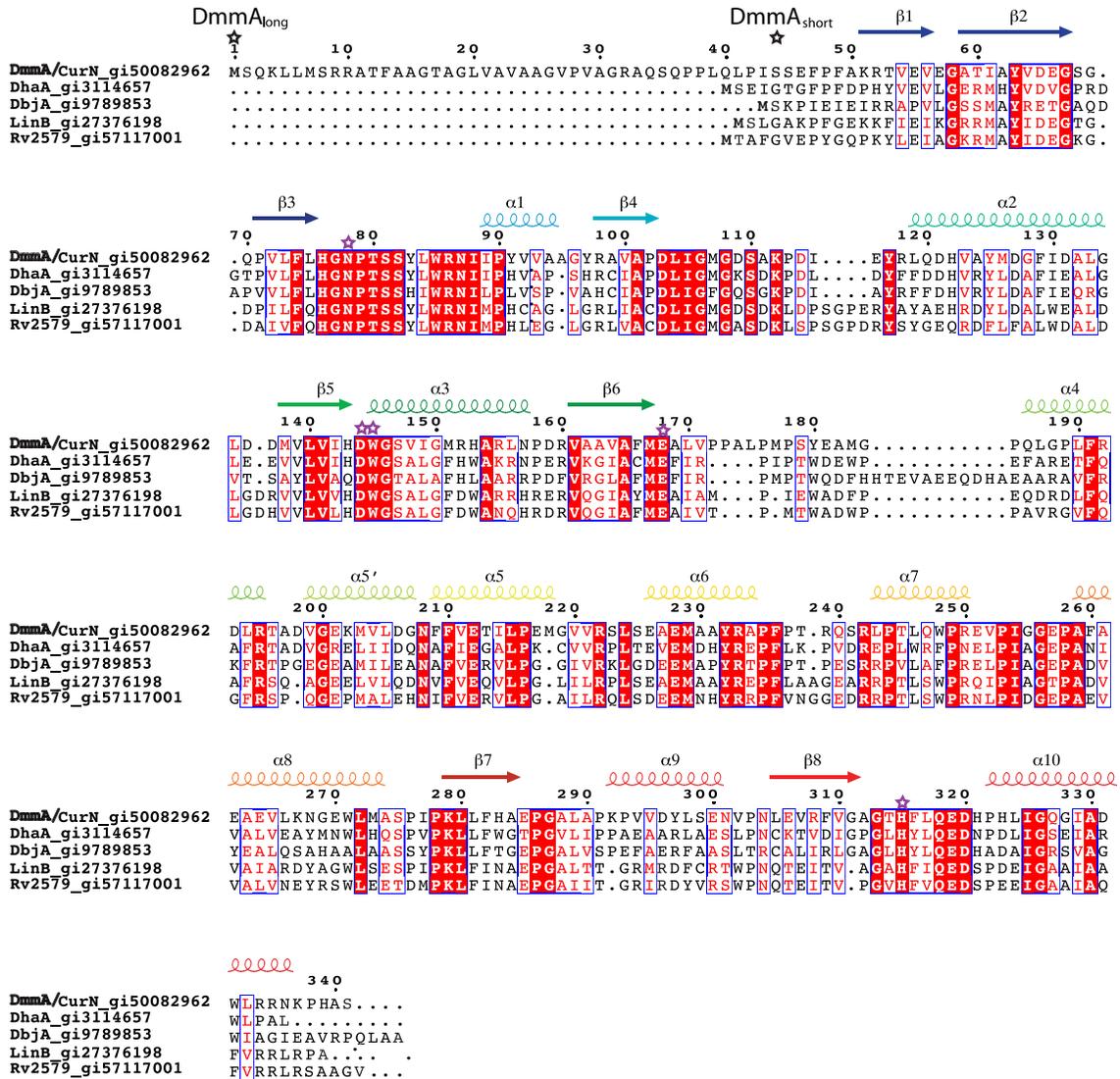
**Figure 5.2 Possible promoter sequences for *dmmA***

Proposed sequence motifs and spacers similar to prokaryotic  $\sigma^{70}$  promoter and ribosome binding sequences upstream of the *DmmA* coding sequence (154). The -35 and -10 promoter regions, transcription start site (TSS), ribosome binding site (RBS) and translation start site are indicated in bold (consensus in parentheses) and labeled in red. Figure was made by Dr. William Clay Brown.

Because the biological source and the natural termini of DmmA are uncertain, we turned to sequence analysis to design *dmmA* constructs for protein expression. HLD homologs have C-termini of very similar length and sequence to the DmmA C-terminus, but the N-termini are poorly conserved and of variable length (134). Additionally, DmmA amino acids 1-40 are predicted to be unstructured. Strong sequence conservation among subfamily-II HLDs begins at DmmA residue 48. This evidence could suggest that the N-terminus of DmmA is not in the original gene, however several considerations led us to consider the full open reading frame as the natural length of DmmA. First, the DNA sequence upstream of the putative translation start site contains sequence motifs and spacers similar to prokaryotic  $\sigma^{70}$  promoter regions and ribosome binding sequences (Figure 5.2). Additionally, several other annotated HLDs have N-termini of similar length to DmmA (134). Thus, we conclude that the natural protein likely includes all the residues encoded in the *curN* (*dmmA*) open reading frame. Two *dmmA* constructs were made, one encoding the full-length protein (DmmA<sub>long</sub>, residues 1-341) and one with a 43-residue N-terminal truncation (DmmA<sub>short</sub>, residues 44-341) to produce a potentially more structured N-terminus (Figure 5.3). Both DmmA<sub>short</sub> and DmmA<sub>long</sub> were produced in high yield, as stable and soluble proteins in an *E. coli* expression system.

#### *Haloalkane dehalogenase activity*

As for many other HLDs, the natural substrate for DmmA is unknown. Small dihalogenated alkanes are typical substrates to assess the activity of HLDs. We chose 1,3-dibromopropane, a standard substrate, for DmmA assays. Haloalkane dehalogenase activity was tested in a colorimetric assay to detect the production of protons (152, 153). Both DmmA<sub>short</sub> and DmmA<sub>long</sub> were active, however DmmA<sub>short</sub> exhibited tenfold higher activity than the longer form of the protein (Table 5.6), suggesting that the N-terminus may interfere with catalysis. Steady-state kinetic constants were determined for DmmA<sub>short</sub> (Table 5.7), which revealed that these parameters are comparable to other HLDs (139, 152, 155). DhIA has 3-fold greater catalytic efficiency than DmmA for 1,3-dibromopropane (4-fold lower  $k_{\text{cat}}$ , 12-fold lower  $K_{\text{m}}$ ) (152), while DhaA has a nearly equal  $k_{\text{cat}}$  and a 50-fold lower  $K_{\text{m}}$  for 1,3-dibromopropane resulting in a 50-fold greater catalytic efficiency (155).



**Figure 5.3 Alignment of DmmA to other subfamily II HLDs of known structure**  
DmmA is aligned with DhaA 1CQW (139), DbjA 3A2M (145), LinB 1IZ8 (143), and Rv2578 2O21 (146). PSI-COFFEE (156) was used to generate the alignment, ESPrpt (95) to prepare the figure, and the STRIDE server (157) to assign secondary structure. Black stars mark the N-termini of DmmA<sub>long</sub> and DmmA<sub>short</sub>, and purple stars mark the catalytic pentad.

**Table 5.6 DmmA activity towards halogenated substrates**

Construct	Substrate	% Activity <sup>a</sup>
DmmA <sub>short</sub>	1,3-dibromopropane	100
	1,6-dibromohexane	3.5 ± 0.7
	bromocyclohexane	14.8 ± 0.6
	1,6-dichlorohexane	3.7 ± 1.4
	1,6-diiodohexane	3.2 ± 0.7
DmmA <sub>long</sub>	1,3-dibromopropane	10 ± 2

<sup>a</sup>Normalized to activity of DmmA<sub>short</sub> with 1 mM 1,3-dibromopropane.

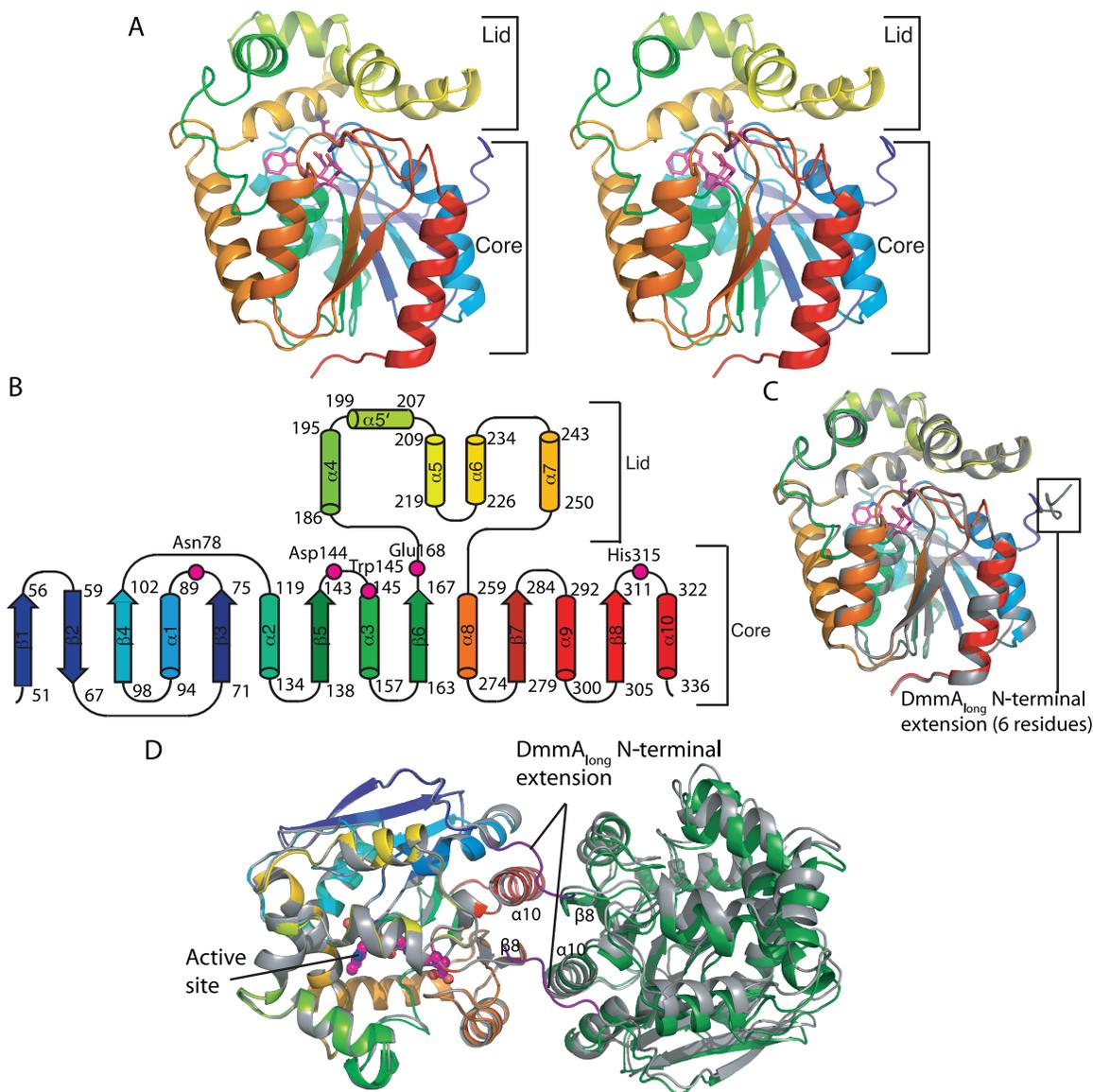
**Table 5.7 Steady state kinetic parameters of DmmA<sub>short</sub>**

$k_{\text{cat}}$ (s <sup>-1</sup> )	$K_{\text{m}}$ (mM)	$k_{\text{cat}}/K_{\text{m}}$ (M <sup>-1</sup> s <sup>-1</sup> )
2.4 ± 0.1	0.24 ± 0.05	1.0 ± 0.3x10 <sup>4</sup>

### *Structure of DmmA*

We solved crystal structures for both DmmA<sub>short</sub> and DmmA<sub>long</sub> (Table 5.1). The first structure was determined from selenomethionyl DmmA<sub>short</sub> (2.2 Å). DmmA<sub>long</sub> (2.5 Å) was solved from the DmmA<sub>short</sub> structure. The short and long variants crystallized under different conditions in different crystal forms, each with two polypeptides per asymmetric unit. The two structures thus provide four independent views of the DmmA dehalogenase.

Due to the higher-resolution of DmmA<sub>short</sub>, our efforts focused on analysis of that form of the protein. DmmA adopts the  $\alpha/\beta$  hydrolase fold as expected (Figure 5.4A), consisting of core and lid sub-domains with the active site at the top of the core, which is covered by the lid (Figure 5.4A,B). DmmA<sub>short</sub> and DmmA<sub>long</sub> are identical within experimental error (RMSD = 0.32 Å for 298 C $\alpha$  atoms). The DmmA<sub>short</sub> structure is complete with no disordered residues. In DmmA<sub>long</sub>, six additional N-terminal residues are visible compared to DmmA<sub>short</sub>, but 37 residues at the N-terminus are disordered (Figure 5.4C).



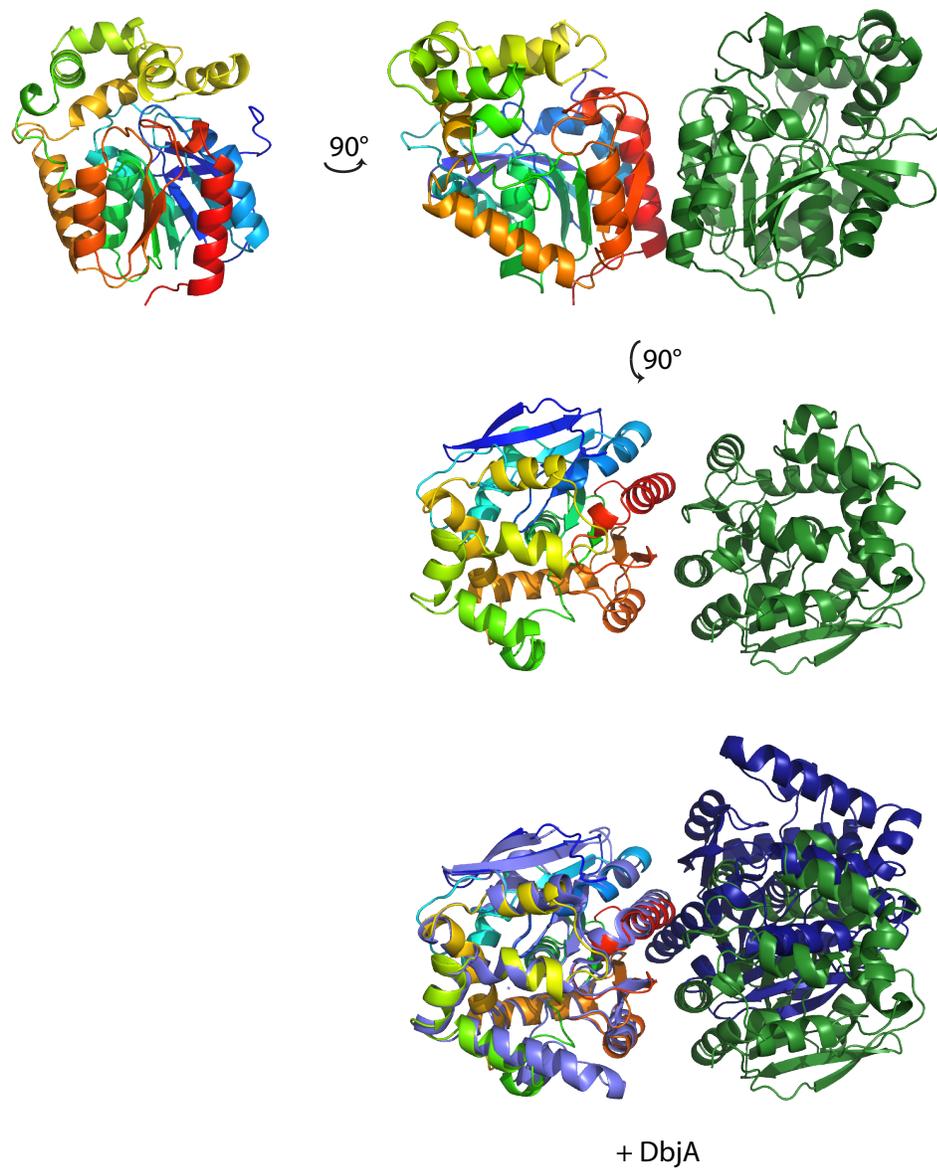
### Figure 5.4 Structure of DmmA

**A.** DmmA polypeptide. The stereo ribbon diagram is colored as a rainbow from blue at the N-terminus to red at the C-terminus with catalytic pentad residues in stick form with magenta. **B.** Topology diagram. The  $\alpha/\beta$  hydrolase core is conserved among all  $\alpha/\beta$  hydrolases and the lid is conserved among haloalkane dehalogenases. Residues of the catalytic pentad are labeled. **C.** Superposition of DmmA<sub>long</sub> (gray) and DmmA<sub>short</sub> (rainbow main-chain, magenta active site). Six additional residues on the N-terminus of DmmA<sub>long</sub> are visible. **D.** Protein-protein contact in the crystal structures. Only the left molecules of DmmA<sub>short</sub> (rainbow main-chain, magenta active site) and DmmA<sub>long</sub> (gray, purple N-terminal extension) are superimposed. The near superposition of the right molecules highlights the similar interfaces. Contacts of  $\alpha 10$ ,  $\beta 8$ , and N-terminal extension are labeled.

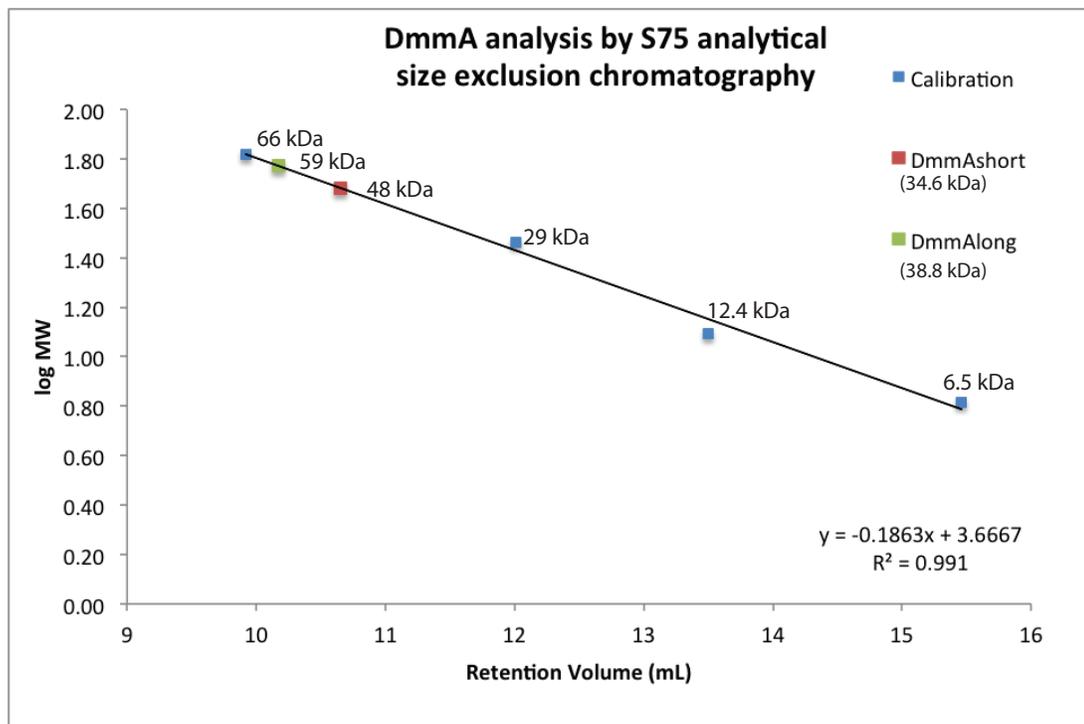
### *Dmma oligomeric state*

A common protein-protein contact exists in the unrelated crystal forms of Dmma<sub>long</sub> and Dmma<sub>short</sub>. The contact has twofold (non-crystallographic) symmetry, suggesting that the protein may be dimeric (Figure 5.4D). The primary interaction occurs within the HLD core between  $\alpha$ 10 and  $\beta$ 8, and is composed of both hydrophobic and hydrophilic contacts and buries only four water molecules. The additional six ordered residues in Dmma<sub>long</sub> (38-43) associate with the partner subunit at this interface, resulting in a larger buried surface area in Dmma<sub>long</sub> compared to Dmma<sub>short</sub> (980 Å<sup>2</sup> vs. 610 Å<sup>2</sup>) (Figure 5.4D). The small size of the interface led us to question whether it exists in solution. Other HLDs have been characterized as monomers, dimers, or dimers only under low-salt conditions. Among HLDs of known structure, only DbjA is dimeric in solution (158). The helix  $\alpha$ 10 is central to the subunit contacts in both Dmma and DbjA, but the interfaces otherwise differ (Figure 5.5).

The quaternary structure in solution was probed by analytical gel filtration chromatography. Both Dmma<sub>short</sub> and Dmma<sub>long</sub> eluted with apparent molecular weights intermediate between calculated values for the monomer and the dimer (48 kDa observed vs. 34.6 kDa for a Dmma<sub>short</sub> monomer, and 59 kDa vs. 38.8 kDa for Dmma<sub>long</sub>) (Figure 5.6). This result implies a dynamic equilibrium between monomeric and dimeric forms, however the retention volume did not change over a 4-fold concentration range of either Dmma<sub>short</sub> or Dmma<sub>long</sub>. Additionally, the larger buried surface in the putative Dmma<sub>long</sub> dimer compared to Dmma<sub>short</sub> does not result in a more dimeric retention volume. Thus, we conclude that the interface observed in the crystal structures does not represent an interface in solution and that Dmma is monomeric. The aberrant elution volumes may result from the dynamic behavior of a monomeric Dmma.



**Figure 5.5 Comparison of DbjA dimer interface and DmmA protein-protein contact**  
 DmmA molecules are in rainbow and green and DbjA is in light and dark purple. The same side of DmmA and DbjA forms the contact, but the interactions differ substantially.

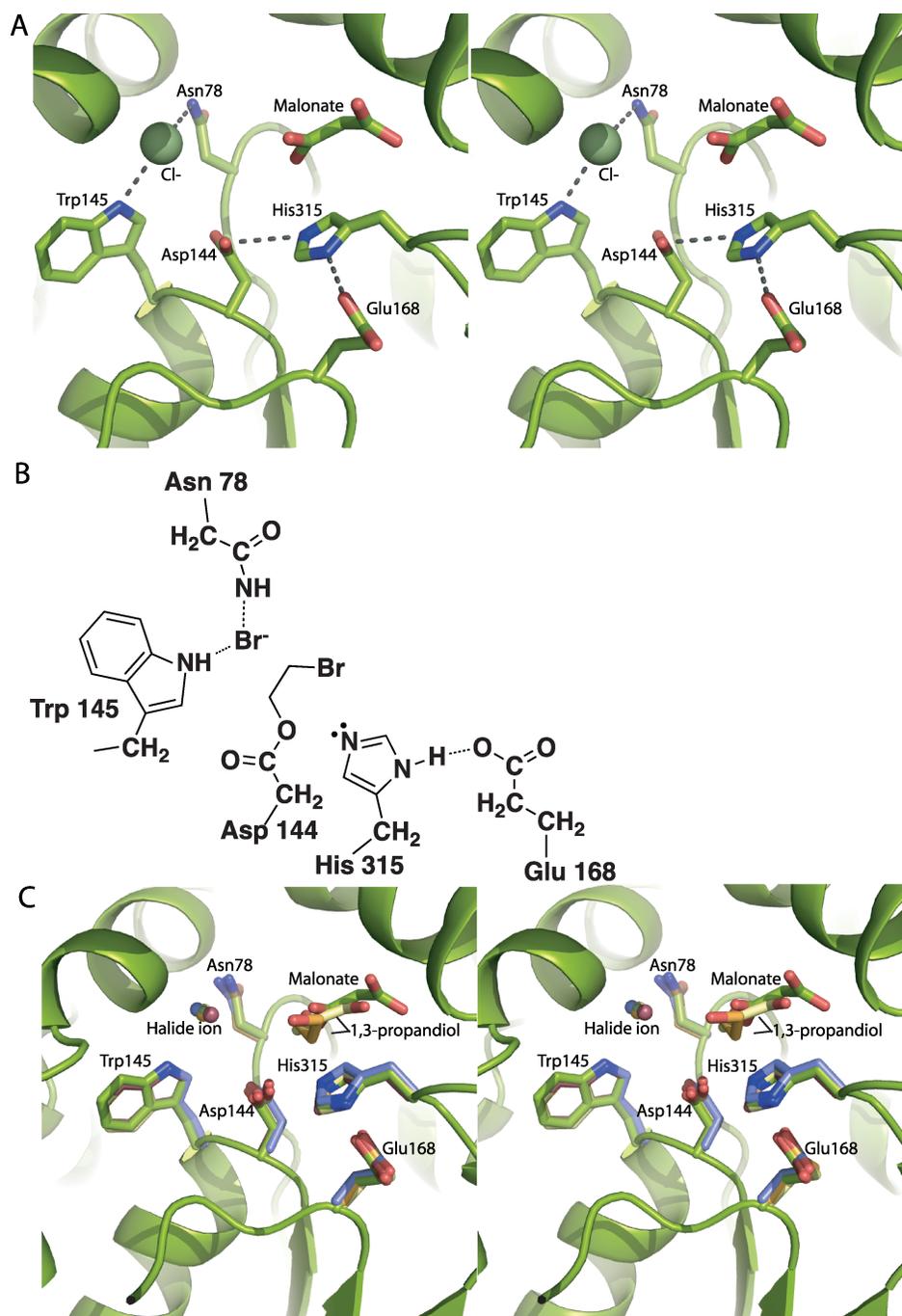


**Figure 5.6 Analysis of DmmA oligomeric state by analytical size exclusion**

DmmA<sub>short</sub> and DmmA<sub>long</sub> (15 mg/mL) were eluted from a Superdex 75 10/300 GL (GE Healthcare) analytical size exclusion column, and compared against standards of known molecular weight and oligomeric state. DmmA<sub>long</sub> and DmmA<sub>short</sub> both elute with an apparent molecular weight (DmmA<sub>short</sub>: 48 kDa, DmmA<sub>long</sub>: 59 kDa) between that of the monomer (DmmA<sub>short</sub>: 34.6 kDa, DmmA<sub>long</sub>: 38.8 kDa) and the dimer (DmmA<sub>short</sub>: 69.2 kDa, DmmA<sub>long</sub>: 77.6 kDa). The elution volumes did not shift when the protein concentrations were decreased 4-fold.

*DmmA active site*

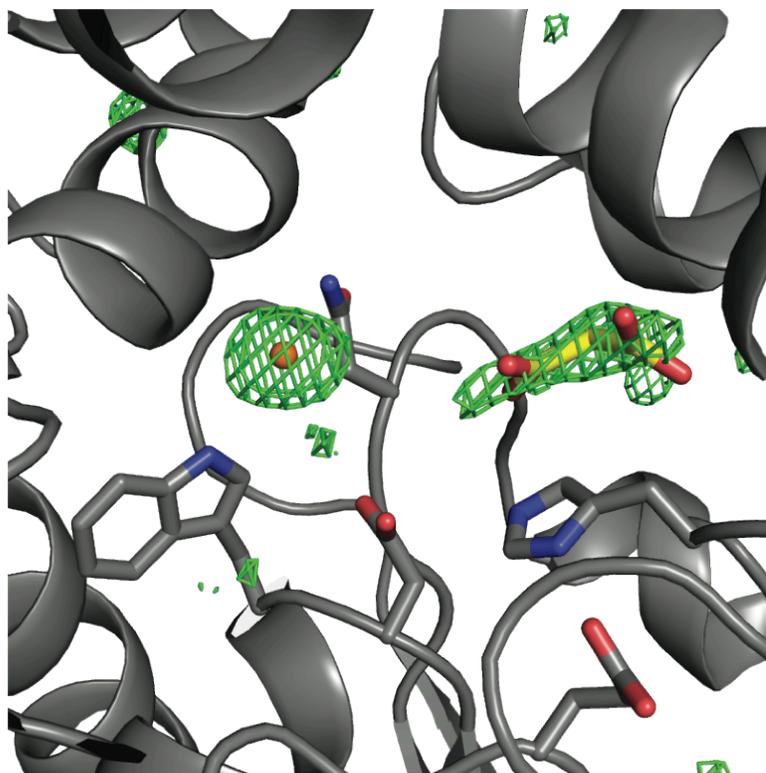
The active site has an intact and well ordered catalytic pentad, consisting of the nucleophile (Asp144), base (His315), acid (Glu168), and two halide-stabilizing residues (Trp145 and Asn78) (Figure 5.7A,B). The nucleophile (Asp144), base (His 315) and Trp145 are conserved in all HLDs. Asn78 is conserved in subfamilies II and III, and Glu168 is present only in subfamily II. The catalytic pentad resides are in an otherwise hydrophobic chamber that is accessed through an entrance tunnel. This feature and the location of the pentad are well conserved in structures of all subfamily II HLDs (Figure 5.7C). Density resembling malonate, a component of the crystallization solution, was



**Figure 5.7 DmmA active site**

**A.** DmmA active site showing catalytic pentad (sticks), Cl<sup>-</sup> and malonate. **B.** Schematic of the DmmA active site showing the acyl-enzyme intermediate. The halide-stabilizing residues (Asn78 and Trp145) bind the halide released when generating the acyl enzyme intermediate. The acyl enzyme on Asp144 will subsequently be hydrolyzed by a water activated by His315. **C.** Active site of subfamily II haloalkane dehalogenases, based on superposition of core C $\alpha$  atoms. The catalytic pentad is shown in sticks and bound halide ion as a sphere for DmmA (green C), LinB 11Z8 (143) (orange C), Rv2578 2O2I

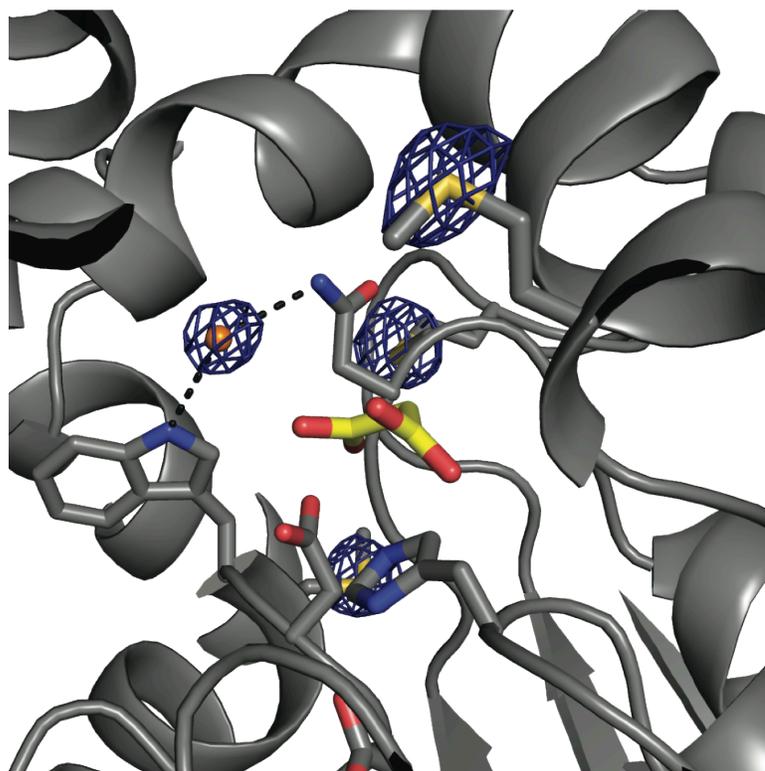
(146) (yellow C), DbjA 3A2M (145) (purple C) and DhaA 1CQW (139) (mauve C). Bound malonate is shown for DmmA and bound product, 1,3-propanediol, for Rv2578 (yellow C) and LinB (orange C).



**Figure 5.8 Omit map of DmmA active site**

Simulated annealing omit map ( $F_o - F_c$ ) contoured at 3.5 sigma generated with  $Cl^-$  and malonate removed from the atomic model.

found in the active site cleft (Figure 5.7A, Figure 5.8). A product alcohol is bound at this position in structures of several other HLDs (Figure 5.7C). Difference density for a stronger scatterer than water (peak height  $7.7\sigma$ ) appeared in the putative halide-binding site within hydrogen bonding distance of Trp 145 and Asn78, the halide-stabilizing residues (Figure 5.7A, Figure 5.8). A halide occupies this position in the structures of several other HLDs (Figure 5.7C). The halide-binding site was confirmed by Br anomalous scattering from a crystal of  $DmmA_{short}$  soaked in 1,5-dibromopentane prior to data collection at the energy of the bromine K absorption edge.  $Br^-$  binding resulted in an anomalous difference peak of height  $9\sigma$ ; no density was observed for the hydroxyl product (Figure 5.9).

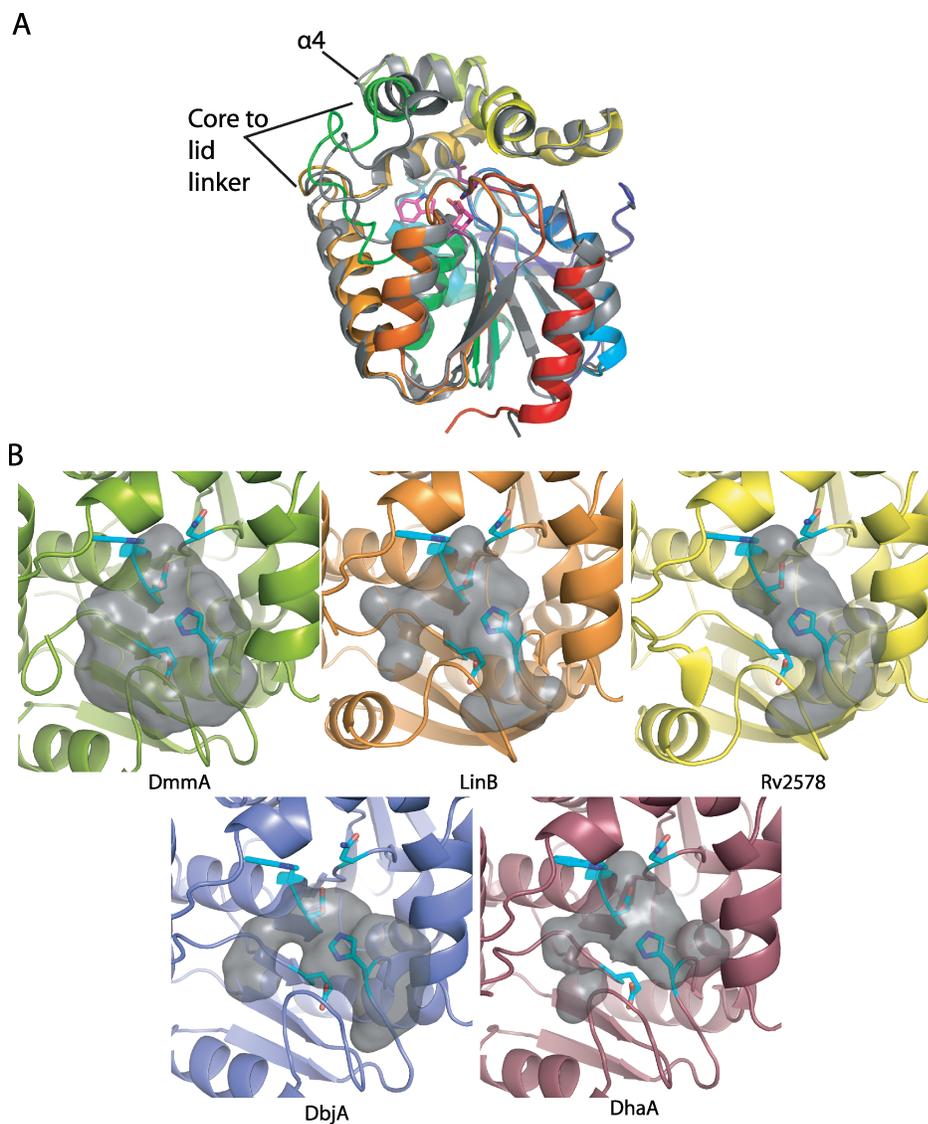


**Figure 5.9 Anomalous difference Fourier map of DmmA active site**

Anomalous difference Fourier map using Friedel data collected at the bromine edge. A crystal of SeMet DmmA<sub>short</sub> was soaked with 1,5-dibromopentane and data were collected at 13.4761 keV. This X-ray energy is 3 eV above the theoretical Br K absorption edge and 800 eV above the Se K edge. The anomalous difference Fourier map contoured at 5.0 sigma shows density for Br<sup>-</sup> and for Se atoms in three SeMet side chains.

*Comparison to other HLDs*

The DmmA structure is similar to structures of other subfamily II HLDs (0.5-0.7 Å RMSD for 219-226 C $\alpha$  atoms of proteins that are 38-48% identical overall to DmmA). The largest difference between the most similar structure, DhaA, and DmmA is in lid helix  $\alpha$ 4 and the core-to-lid linker from  $\beta$ 6 to  $\alpha$ 4 (Figure 5.10A). As expected, DmmA is less similar to the subfamily I HLDs, Dh1A (1.4 Å RMSD for 203 C $\alpha$ ) and DppA (1.9 Å for 209 C $\alpha$ ), which have lower overall sequence identity to DmmA.



**Figure 5.10 DmmA and other HLDs**

**A.** Superposition of DmmA (rainbow main-chain, magenta active site) and the HLD DhaA (1BN7 (139)). Differences in the lid-to-core linker and  $\alpha 4$  are labeled. **B.** Active site cavity of DmmA and other type II haloalkane dehalogenases: DmmA (green), LinB 1IZ8 (143) (orange), Rv2578 2O2I (146) (yellow), DbjA 3A2M (145) (purple) and DhaA 1CQW (139) (mauve). The gray surface represents the internal volume of the cavity. Residues that form the forward surface are omitted. Active site residues are shown in sticks (cyan C, blue N, red O).

Structure-function studies of other HLDs have demonstrated that the size of the active site cavity and entry tunnel play a role in substrate specificity (132) and chiral selectivity (145). DmmA has a larger entry tunnel and wider substrate cleft than other HLDs (Figure 5.10B). This difference is due primarily to the position of lid helix  $\alpha 4$  and

the preceding loop, and to a lesser extent to amino acid substitutions within the cavity. The loop and helix  $\alpha 4$  are the parts of the DmmA structure that differ most from other HLDs. Other HLDs with large active sites (141, 145) have very broad substrate specificity and react well with larger substrates (132). DmmA, with an even larger active site, should accommodate substrates of greater size and shape diversity. To test this, we assayed DmmA for dehalogenase activity with several halogenated molecules larger than the 1,3-dibromopropane standard substrate (1,6-dibromohexane, 1,6-dichlorohexane, 1,6-diiodohexane, and bromocyclohexane) (Table 5.6). In contrast to those tested previously (132), DmmA has greater activity for the bromocyclohexane compared to the larger straight-chain substrates. This observation is consistent with the wide cavity of DmmA and suggests that it has robust activity towards other bulky substrates.

## Discussion

This work establishes that DmmA is a haloalkane dehalogenase. The kinetic properties of DmmA with a standard substrate, 1,3-dibromopropane, are within the range reported for other HLDs. However, as for most HLDs, the natural substrates are unknown. Of the two protein variants studied, DmmA<sub>short</sub> was significantly faster than DmmA<sub>long</sub>, exhibiting tenfold greater activity. Although most of the additional 43 residues on the N-terminus of DmmA<sub>long</sub> are disordered, they may interfere with substrate access to the active site, resulting in the reduced activity observed for the longer form of the protein.

The DmmA structure has the same fold and active site position as other HLDs, as expected. Nevertheless, DmmA has unique structural features that set it apart from other enzymes within this class. The active site cavity of DmmA is significantly larger than the cavity of other homologs, which was not predictable from straightforward sequence comparisons to other HLDs of known structure. Instead, sequence differences led to subtle structural variation in the position of helix  $\alpha 4$  and the core-to-lid linker, which resulted in a large cavity volume. The size of amino acid side chains lining the active site pocket makes a secondary contribution to the cavity size. The larger cavity volume may confer an ability to accept a wide range of large, bulky substrates compared to other HLDs. Among the larger substrates tested, DmmA had a preference for the bulky bromocyclohexane over corresponding linear molecules. This suggests that the natural

DmmA substrate may bear a halogenated ring system. The large active site also suggests the value of DmmA as a potential chemoenzymatic tool for large substrates.

The ultimate biological source of DmmA is unknown, as attempts to amplify *dmmA* from *M. producens* (formerly *L. majuscula*) DNA or to find *dmmA* in the genome sequence have been unsuccessful. *dmmA* (originally annotated as *curN* (2)) was originally found in the cosmid library developed from an *M. producens* field isolate, which also contained bacterial symbionts or associants as a microbial consortium. To our knowledge no other HLD has been found in a cyanobacterium, which also suggests an origin outside *M. producens*, in a symbiont or associated bacterium. Presumably, the natural function of DmmA is to dehalogenate molecules present in the native marine environment of *M. producens*. In this respect, *M. producens* has been a rich source of halogenated metabolites, including in this strain, barbamide, which possesses a trichloromethyl group (6), and in another strain, jamaicamide, which possesses vinyl chloride and alkynyl bromide groups (159). Despite its orphan status, the large active-site cavity poises DmmA for development as a new biotechnology tool.

#### **Author contributions to the manuscript**

JGM, L. Gu, TWG, DHS, and JLS designed the research. L. Gu created the *dmmA* expression plasmid. WCB analyzed the promoter sequence. L. Gerwick searched for *dmmA* in genomic DNA. JGM performed all other experiments. JGM, TWG, and JLS analyzed data. JGM and JLS wrote the paper. JGM, L. Gu, L. Gerwick, WCB, WHG, DHS and JLS revised and edited the paper.

## Chapter 6

### Conclusions and Future Directions

#### Conclusions

##### *Overview*

Presented here are the structural and biochemical details of the sulfotransferase (ST) and thioesterase (TE) domains from both a polyketide synthase (PKS) and an olefin synthase (OLS) system. This study revealed two enzymes working together: the ST for substrate activation, and the TE for a unique sulfonation-assisted decarboxylative elimination reaction producing a terminal alkene. Structural and biochemical investigations showed how the CurM and OLS ST and TE domains differ from canonical ST and TE folds to perform chemistry unique to the terminal-alkene producing systems. The CurM and OLS enzymes are similar to each other in structure and activity, but also exhibit subtle differences in structure, substrate specificity, and catalytic efficiency.

##### *Activity*

The investigation of OLS ST and OLS TE was the first demonstration of OLS activity *in vitro*. OLS ST has sulfotransferase activity with  $\beta$ -hydroxyacyl-ACP substrates but does not tolerate substrates with a C5-methoxy (Chapter 3). CurM ST, however, accepts C5-methoxy substrates and has great activity than OLS ST under the conditions tested (Chapter 3). OLS TE, like CurM TE, has thioesterase activity specifically with  $\beta$ -sulfated substrates and accepts substrates with and without a C5-methoxy (Chapter 4). OLS ST and TE show a tolerance for the exogenous ACPs by acting on the ACP from CurM (Chapters 3 & 4). The knowledge of reaction rates and specificities gained in this work will help guide development of the ST and TE enzymes as chemical tools.

### *Thioesterase*

The OLS and CurM TE structures show how the common TE fold is adapted to perform the specialized functions of thioester hydrolysis, sulfate elimination, and decarboxylation. From the structure of CurM TE, an explanation was proposed for the intriguing specificity for substrates with a  $\beta$ -sulfate group (Chapter 2). The novel lid structure and dimer interface create an open active site and a conserved Arg, shown to be vital for  $\beta$ -sulfate recognition, is a critical element to position the substrate for catalysis by interacting with the  $\beta$ -sulfate group. The structure of the active site in CurM TE also led to a proposed mechanism for the concerted decarboxylation/sulfate elimination step (Chapter 2). The OLS TE structure supports all the conclusions drawn from CurM TE, and the greater protein stability of OLS TE allows further probing of the protein by amino acid substitutions that were not possible for CurM TE (Chapter 4). A pronounced positive surface patch in the lid of OLS TE is predicted to be an ACP docking site, a hypothesis supported by the reduced activity of OLS TE with charge-altering amino acid substitutions (Chapter 4). The OLS and CurM TE structures show many novel features including the dimer interface, lid structure, and ACP binding site that are unique to decarboxylating TEs.

### *Sulfotransferase*

OLS and CurM ST represent a distinct branch of the ST superfamily that is used for functional-group activation. OLS and CurM ST have similar structures with unique peripheral loops and helices compared to the canonical ST fold (Chapter 3). The details of the PAPS binding site were identified by co-crystallization with PAP and  $\text{ZnCl}_3^-$  (a mimic of the PAPS sulfate) (Chapter 3). Proposals for the catalytic base, for amino acids that interact with PAPS, and for amino acids that hydrogen bond in the active site were supported by reduced activity of amino acid substitutions (Chapter 3). The structures also reveal a dynamic active-site flap that presumably closes over the substrate during sulfonate transfer. OLS/CurM flap chimeras demonstrate that the active-site flap is a determinant of substrate specificity for the C5-methoxy (Chapter 3). Knowledge of the structures and the specificity-determining active-site flap will assist in the development of the STs into chemical tools.

### *ST-TE didomain*

The structure and activity of the OLS ST-TE didomain was investigated to probe the interaction of the two domains. The ST and TE on the same polypeptide do not interact in the didomain structure, and ST to TE crystal contacts do not appear to be biologically relevant (Chapter 4). The ST-TE linker, which is not conserved in length or amino acid content, is disordered in the structure and appears to be flexible (Chapter 4). There was no change to the reaction rate for the ST and TE when they acted *in trans*, *in cis*, or *in cis* with a shortened linker (Chapter 4). From the investigation of the ST-TE didomain, there is no evidence of any direct substrate channeling between the two domains.

### *DmmA*

The original annotation of the 3' terminus of the curacin gene cluster included the coding sequence for a haloalkane dehalogenase (HLD) DmmA (CurN), but upon resequencing the cluster was found to lack a *curN*. The origin of DmmA is currently unknown, as it does not appear in the *M. producens* genome sequence and cannot be amplified from genomic DNA. The genetic material for DmmA may have come from another bacterium in the *M. producens* marine microbial consortium. Despite its unknown origins, DmmA was characterized as a HLD and may be a potential chemical tool. The DmmA structure is similar to structures of other HLDs but has a larger substrate-binding pocket (Chapter 5). DmmA was shown to have haloalkane dehalogenase activity and to act upon small (3-carbon) as well as large and bulky (linear and cyclic 6-carbon) substrates (Chapter 5). HLDs have previously been developed as chemical tools for many different applications (71, 70). DmmA, with its large active site, is a promising candidate for applications that utilize larger substrates that may not work with other HLDs. A systematic survey of DmmA substrate specificity may identify optimal substrates and open up the door for chemical tool development.

## **Future directions**

### *Structures with substrates or products*

Currently, modeling of substrates into both the ST and TE active sites was used to make hypotheses about substrate affinity, binding, and catalysis. Crystal structures of

bound substrates, products, or transition state analogs would allow for more accurate and detailed interpretations of the active site than is available with the present models. Details of substrate specificity could be deduced from structures of different substrates. Structures might be obtained either with an ACP-tethered substrate or by soaking or co-crystallizing a small-molecule substrate. If the structures contained ACP, the ACP binding interface could also be characterized in detail. Obtaining structures of PKS enzymes complexes presents a challenge since the binding affinity for substrates is often low. Strategies to overcome this include capturing the TE acyl-enzyme intermediate and making amino acids substitutions catalytic residues in the ST to prevent the turnover and release of a substrate.

#### *Hydrocarbon product detection*

The studies of OLS activity presented in Chapters 3 and 4 were performed by detecting modifications of the ACP [sulfated ACP (ST) and holo ACP (TE)]. Detection of the ACP products allowed for rapid analysis of catalytic activity by an HPLC based assay but did not allow for the detection of the small molecule products. Detection of hydrocarbons from OLS ST-TE *in vitro* using a GCMS based assay has thus far eluded us. However, we know that OLS produces hydrocarbons *in vivo* from gene deletion and feeding studies (52). The development of new workflows or implementation of new detection methods could address this challenge. Reconstitution of hydrocarbon production *in vitro* would be a major step forward for the study of the OLS system and enable further studies focusing on the mechanism of hydrocarbon production.

#### *Mechanism and substrate specificity*

Many unanswered questions remain about the mechanism and substrate specificity of the ST and TE. Currently we know that CurM and OLS TE act on substrates with  $\beta$ -sulfate groups but not with  $\beta$ -hydroxy groups. Do the TEs accept other negatively charged leaving groups? What effect do different leaving groups have on substrate binding, catalysis, and the production of the terminal alkene? These questions can be probed by using additional substrates with other negatively charged leaving groups at the  $\beta$ -position such as phosphate. Screening should include both the rate of TE hydrolysis and the ability of the TE to eliminate the leaving group to generate the

terminal alkene. These studies may also shed light into why sulfate is used as a leaving group in this instance instead of the more common phosphate.

The ST and TE were tested with substrates both containing and lacking a C5-methoxy substituent. The results show contrasting substrate specificity in CurM ST (accepting C5-methoxy substrates) and OLS ST (not accepting C5-methoxy substrates). However, there is no knowledge of specificity for any other functional group past the  $\beta$ -position. How flexible is substrate specificity for CurM ST? Does OLS ST only accept substituent-free acyl chains? How broad a range of substrates do CurM and OLS TE accept? Questions about substrate flexibility can be addressed by probing with substrates containing additional functional groups beyond the  $\beta$ -hydroxy. Priority should be given to functional groups found in other PKS systems (such as methyl, hydroxy, enoyl, or internal double bonds) to probe for the potential of the ST and TE to be used in combinatorial biosynthesis. Other potential substrates to test include bulky groups or amino acids from NRPS systems. Combined knowledge of the substrate preferences and the structures could help in elucidating other residues involved in substrate specificity and catalysis.

#### *The full OLS module*

Studies of the OLS ST and TE domains have begun to shed light into the OLS system but the six domains N-terminal of the ST-TE domains have not yet been characterized. Cloning, recombinant expression, and purification of full-length OLS would open up study to the OLS system as a whole. The activity of the full-length construct may be able to be reconstituted *in vitro*. The substrates tested are commercially available fatty acids that with the addition of the appropriate cofactors (ATP, malonyl-CoA, NADPH, PAPS) should be converted to hydrocarbons. In addition, the architecture of the full OLS module could be pursued by electron microscopy or X-ray crystallography. Characterization of the full-length OLS would directly open the door to opportunities for bioengineering for hydrocarbon production.

#### *The acyl activating (AA) domain*

A domain of interest in OLS is the acyl activating (AA) domain, which is not present in the better-studied PKS and FAS systems. AA domains activate fatty acids by

5'-adenylation from ATP and load them onto phosphopantethienyl-ACP (holo-ACP) (53). When no ACP is present the AAs will release the acyl-AMP intermediate (160, 128). The AA domain has the vital role in the OLS system of activating a fatty acid (presumably produced by the FAS machinery) and directing it into the OLS system. The AA could be cloned individually or in concert with the tandem ACP and studied biochemically and structurally. The substrate tolerance could be tested with an array of commercially available fatty acids varying in length and functionalization. The crystal structure of the OLS AA domain (32% sequence identity to other AA structures) could be pursued either alone or with the tandem ACP to visualize the ACP interaction.

The AA domain has the potential to be developed as a tool for the study of PKS and FAS reactions. A challenge in the production of substrate mimics for assay is the coupling of substrates to CoA. CoA coupling is necessary for substrate loading onto ACP by a phosphopantetheinyltransferase such as SVP (94). CoA coupling is often the bottleneck in producing substrates to assay PKS enzymes. The AA has the potential to prepare ACP-linked substrates from fatty acids, which completely bypasses the need for the tedious step of producing the CoA linked intermediate. The specificity of both the fatty acid and ACP substrates should be examined to determine how widely the OLS AA could be applied.

#### *Investigation of new pathways*

Several interesting new PKS and NRPS pathways have been identified during the investigation of ST-TE offloading. Two pathways containing the ST-TE didomain were identified from the proteobacteria *Haliangium ochraceum* and *Pseudomonas entomophila* (Chapter 1). The small-molecules produced by these pathways, which presumably contain terminal alkenes, have not been identified. Five pathways were identified (2 cyanobacterial and 3 microalgal) that contain a gene for the ST but lack a gene for the decarboxylating TE (Chapter 3). With the exception of one pathway from *Planktothrix*, the natural products from the discovered pathways have not been identified but presumably contain a sulfate moiety. The pathways mentioned above have unusual domains and domain arrangements that merit closer investigation and identification of the products. The ST and TE genes can continue to be a probe to identify interesting new pathways as more microbial genomes are sequenced.

## **Applications**

### *Combinatorial biosynthesis*

Combinatorial biosynthesis uses genetic engineering to exchange domains or modules within PKS or NPRS pathways to produce novel “semi-natural” products (100, 101). With this technique, chemical space that is challenging to access using traditional organic synthesis methods can be explored. Current successes of combinatorial biosynthesis include incorporation of tailoring enzymes such as halogenases into pathways (161, 100), the reconstitution of PKS pathways into genetically tractable hosts (162), and exchanging domains between closely related pathways (163). Although promising, more methods development must occur before this technique is able to reach its full potential.

The ST-TE system provides the ability to produce two different chemical moieties through combinatorial biosynthesis. The terminal alkene could be produced using the ST and TE from CurM, which have both greater activity and a greater substrate range than the OLS enzymes. The ST and TE could presumably be substituted in place of a canonical offloading TE in any pathway that produces a  $\beta$ -hydroxy in its final module or in which the DH and ER, if present, are removed or inactivated. A sulfate group could also be generated using the ST alone as a chemical tool. The ST could either be inserted at the end of the pathway preceding a “canonical” TE in a module that produces a  $\beta$ -hydroxy or be inserted in an internal module. CurM ST may be a better target for the generation of sulfate moieties as it accepts more substrates than OLS ST.

### *Biofuel engineering*

The most widely used biofuels are currently ethanol and biodiesel. Both have drawbacks to being used in the present petroleum-based infrastructure. Challenges with transportation, corrosion of engines, and links between food and fuel prices have encouraged the development of biofuel with similar properties as petroleum (131). OLS, naturally producing C<sub>17</sub> and C<sub>19</sub> alkenes from fatty acids, has potential to be utilized for diesel (C<sub>9</sub>-C<sub>23</sub>) or jet fuel (C<sub>8</sub>-C<sub>16</sub>). Cyanobacteria (the OLS source organism) have several advantages for biofuel production including fast growth compared to plants, the ability to be genetically engineered, and the simple nutrient requirements of CO<sub>2</sub>, light,

and water (164). Manipulation of the fatty acid synthase in cyanobacteria can produce high levels of fatty acids (130), which can subsequently be converted to the hydrophobic molecules useful for fuels. Other routes to biofuels from fatty acids include fatty alcohols produced by an acyl-CoA synthase and acyl-CoA reductase, methyl ketones produced by the first three steps of  $\beta$ -oxidation, alkanes produced by an acyl-ACP reductase and aldehyde decarbonylase, and alkenes produced by a cytochrome P450 (131). The OLS pathway, being a single gene of cyanobacterial origin, makes a unique addition to the arsenal of chemical tools for generating hydrocarbons from fatty acids.

CurM ST and TE also show potential to be developed for biofuel production since they act on long acyl chain substrates. The CurM ST and TE enzymes, which have greater activity than the OLS ST and TE enzyme in the conditions tested, could be incorporated into the OLS pathway. Alternatively, the full length CurM module could be modified to act on hydrocarbons with the addition of acyl-activating and ACP domains. Both CurM and OLS ST-TE mediated decarboxylative offloading systems show great promise for biofuel engineering.

## References

1. Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod* 75:311-335.
2. Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH (2004) Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 67:1356-1367.
3. Gerwick WH, Proteau PJ, Nagle DG, Hamel E, Blokhin A, Slate DL (1994) Structure of Curacin a, a Novel Antimitotic, Antiproliferative, and Brine Shrimp Toxic Natural Product from the Marine Cyanobacterium *Lyngbya majuscula*. *J Org Chem* 59:1243-1245.
4. Engene N, Rottacker EC, Kastovsky J, Byrum T, Choi H, Ellisman MH, Komarek J, Gerwick WH (2012) *Moorea producens* gen. nov., sp. nov. and *Moorea bouillonii* comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. *Int J Syst Evol Microbiol* 62:1171-1178.
5. Jones AC, Monroe EA, Podell S, Hess WR, Klages S, Esquenazi E, Niessen S, Hoover H, Rothmann M, Lasken RS, Yates JR, 3rd, Reinhardt R, Kube M, Burkart MD, Allen EE, Dorrestein PC, Gerwick WH, Gerwick L (2011) Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. *Proc Natl Acad Sci U S A* 108:8815-8820.
6. Chang Z, Flatt P, Gerwick WH, Nguyen VA, Willis CL, Sherman DH (2002) The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* 296:235-247.
7. Jones AC, Gerwick L, Gonzalez D, Dorrestein PC, Gerwick WH (2009) Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins. *BMC Microbiol* 9:247.
8. Sorrels CM, Proteau PJ, Gerwick WH (2009) Organization, evolution, and expression analysis of the biosynthetic gene cluster for scytonemin, a cyanobacterial UV-absorbing pigment. *Appl Environ Microbiol* 75:4861-4869.
9. Edwards DJ, Gerwick WH (2004) Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. *J Am Chem Soc* 126:11432-11433.
10. Akey DL, Gehret JJ, Khare D, Smith JL (2012) Insights from the sea: Structural biology of marine polyketide synthases. *Nat Prod Rep* 29:1038-1049.
11. Blokhin AV, Yoo HD, Gerald RS, Nagle DG, Gerwick WH, Hamel E (1995) Characterization of the interaction of the marine cyanobacterial natural product curacin A with the colchicine site of tubulin and initial structure-activity studies with analogues. *Mol Pharmacol* 48:523-531.

12. Verdier-Pinard P, Lai JY, Yoo HD, Yu J, Marquez B, Nagle DG, Nambu M, White JD, Falck JR, Gerwick WH, Day BW, Hamel E (1998) Structure-activity analysis of the interaction of curacin A, the potent colchicine site antimetabolic agent, with tubulin and effects of analogs on the growth of MCF-7 breast cancer cells. *Mol Pharmacol* 53:62-76.
13. Austin MB, Saito T, Bowman ME, Haydock S, Kato A, Moore BS, Kay RR, Noel JP (2006) Biosynthesis of *Dictyostelium discoideum* differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase. *Nat Chem Biol* 2:494-502.
14. Gokhale RS, Saxena P, Chopra T, Mohanty D (2007) Versatile polyketide enzymatic machinery for the biosynthesis of complex mycobacterial lipids. *Nat Prod Rep* 24:267-277.
15. Callahan B, Thattai M, Shraiman BI (2009) Emergent gene order in a model of modular polyketide synthases. *Proc Natl Acad Sci U S A* 106:19410-19415.
16. Fischbach MA, Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* 106:3468-3496.
17. Tsai SC, Ames BD (2009) Structural enzymology of polyketide synthases. *Methods Enzymol* 459:17-47.
18. Sherman DH, Smith JL (2006) Clearing the skies over modular polyketide synthases. *ACS Chem Biol* 1:505-509.
19. Gu L, Geders TW, Wang B, Gerwick WH, Hakansson K, Smith JL, Sherman DH (2007) GNAT-like strategy for polyketide chain initiation. *Science* 318:970-974.
20. Gu L, Jia J, Liu H, Hakansson K, Gerwick WH, Sherman DH (2006) Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: functional identification of a mechanistically diverse enzyme pair. *J Am Chem Soc* 128:9014-9015.
21. Geders TW, Gu L, Mowers JC, Liu H, Gerwick WH, Hakansson K, Sherman DH, Smith JL (2007) Crystal structure of the ECH2 catalytic domain of CurF from *Lyngbya majuscula*. Insights into a decarboxylase involved in polyketide chain beta-branching. *J Biol Chem* 282:35954-35963.
22. Gu L, Wang B, Kulkarni A, Geders TW, Grindberg RV, Gerwick L, Hakansson K, Wipf P, Smith JL, Gerwick WH, Sherman DH (2009) Metamorphic enzyme assembly in polyketide diversification. *Nature* 459:731-735.
23. Khare D, Wang B, Gu L, Razelun J, Sherman DH, Gerwick WH, Hakansson K, Smith JL (2010) Conformational switch triggered by alpha-ketoglutarate in a halogenase of curacin A biosynthesis. *Proc Natl Acad Sci U S A* 107:14099-14104.
24. Gu L, Eisman EB, Dutta S, Franzmann TM, Walter S, Gerwick WH, Skiniotis G, Sherman DH (2011) Tandem acyl carrier proteins in the curacin biosynthetic pathway promote consecutive multienzyme reactions with a synergistic effect. *Angew Chem Int Ed Engl* 50:2795-2798.
25. Busche A, Gottstein D, Hein C, Ripin N, Pader I, Tufar P, Eisman EB, Gu L, Walsh CT, Sherman DH, Lohr F, Guntert P, Dotsch V (2012) Characterization of molecular interactions between ACP and halogenase domains in the Curacin A polyketide synthase. *ACS Chem Biol* 7:378-386.

26. Akey DL, Razelun JR, Tehranisa J, Sherman DH, Gerwick WH, Smith JL (2010) Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* 18:94-105.
27. Gu L, Wang B, Kulkarni A, Gehret JJ, Lloyd KR, Gerwick L, Gerwick WH, Wipf P, Hakansson K, Smith JL, Sherman DH (2009) Polyketide decarboxylative chain termination preceded by *O*-sulfonation in curacin a biosynthesis. *J Am Chem Soc* 131:16033-16035.
28. Tsai SC, Miercke LJ, Krucinski J, Gokhale R, Chen JC, Foster PG, Cane DE, Khosla C, Stroud RM (2001) Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proc Natl Acad Sci U S A* 98:14808-14813.
29. Tsai SC, Lu H, Cane DE, Khosla C, Stroud RM (2002) Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry* 41:12598-12606.
30. Akey DL, Kittendorf JD, Giraldes JW, Fecik RA, Sherman DH, Smith JL (2006) Structural basis for macrolactonization by the pikromycin thioesterase. *Nat Chem Biol* 2:537-542.
31. Giraldes JW, Akey DL, Kittendorf JD, Sherman DH, Smith JL, Fecik RA (2006) Structural and mechanistic insights into polyketide macrolactonization from polyketide-based affinity labels. *Nat Chem Biol* 2:531-536.
32. Scaglione JB, Akey DL, Sullivan R, Kittendorf JD, Rath CM, Kim ES, Smith JL, Sherman DH (2010) Biochemical and Structural Characterization of the Tautomycetin Thioesterase: Analysis of a Stereoselective Polyketide Hydrolase. *Angew Chem Int Ed Engl* 49:5726-5730.
33. Bruner SD, Weber T, Kohli RM, Schwarzer D, Marahiel MA, Walsh CT, Stubbs MT (2002) Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* 10:301-310.
34. Samel SA, Wagner B, Marahiel MA, Essen LO (2006) The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. *J Mol Biol* 359:876-889.
35. Samel SA, Schoenafinger G, Knappe TA, Marahiel MA, Essen LO (2007) Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* 15:781-792.
36. Frueh DP, Arthanari H, Koglin A, Vosburg DA, Bennett AE, Walsh CT, Wagner G (2008) Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* 454:903-906.
37. Koglin A, Lohr F, Bernhard F, Rogov VV, Frueh DP, Strieter ER, Mofid MR, Guntert P, Wagner G, Walsh CT, Marahiel MA, Dotsch V (2008) Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature* 454:907-911.
38. Liu Y, Zheng T, Bruner SD (2011) Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chem Biol* 18:1482-1488.

39. Claxton HB, Akey DL, Silver MK, Admiraal SJ, Smith JL (2009) Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway. *J Biol Chem* 284:5021-5029.
40. Whicher JR, Florova G, Sydor PK, Singh R, Alhamadsheh M, Challis GL, Reynolds KA, Smith JL (2011) Structure and function of the RedJ protein, a thioesterase from the prodiginine biosynthetic pathway in *Streptomyces coelicolor*. *J Biol Chem* 286:22558-22569.
41. Rounge TB, Rohrlack T, Tooming-Klunderud A, Kristensen T, Jakobsen KS (2007) Comparison of cyanopeptolin genes in *Planktothrix*, *Microcystis*, and *Anabaena* strains: evidence for independent evolution within each genus. *Appl Environ Microbiol* 73:7322-7330.
42. Rounge TB, Rohrlack T, Kristensen T, Jakobsen KS (2008) Recombination and selectional forces in cyanopeptolin NRPS operons from highly similar, but geographically remote *Planktothrix* strains. *BMC Microbiol* 8:141.
43. Bick MJ, Banik JJ, Darst SA, Brady SF (2010) Crystal structures of the glycopeptide sulfotransferase Teg12 in a complex with the teicoplanin aglycone. *Biochemistry* 49:4159-4168.
44. Chapman E, Best MD, Hanson SR, Wong CH (2004) Sulfotransferases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew Chem Int Ed Engl* 43:3526-3548.
45. Bowman KG, Bertozzi CR (1999) Carbohydrate sulfotransferases: mediators of extracellular communication. *Chem Biol* 6:R9-R22.
46. Liu J, Pedersen LC (2007) Anticoagulant heparan sulfate: structural specificity and biosynthesis. *Appl Microbiol Biotechnol* 74:263-272.
47. Schelle MW, Bertozzi CR (2006) Sulfate metabolism in mycobacteria. *Chembiochem* 7:1516-1524.
48. Ivanova N, Daum C, Lang E, Abt B, Kopitz M, Saunders E, Lapidus A, Lucas S, Glavina Del Rio T, Nolan M, Tice H, Copeland A, Cheng JF, Chen F, Bruce D, Goodwin L, Pitluck S, Mavromatis K, Pati A, Mikhailova N, Chen A, Palaniappan K, Land M, Hauser L, Chang YJ, Jeffries CD, Detter JC, Brettin T, Rohde M, Goker M, Bristow J, Markowitz V, Eisen JA, Hugenholtz P, Kyrpides NC, Klenk HP (2010) Complete genome sequence of *Haliangium ochraceum* type strain (SMP-2). *Stand Genomic Sci* 2:96-106.
49. Fudou R, Iizuka T, Yamanaka S (2001) Haliangicin, a novel antifungal metabolite produced by a marine myxobacterium. 1. Fermentation and biological characteristics. *J Antibiot (Tokyo)* 54:149-152.
50. Kundim BA, Ito Y, Sakagami Y, Fudou R, Iizuka T, Yamanaka S, Ojika M (2003) New haliangicin isomers, potent antifungal metabolites produced by a marine myxobacterium. *J Antibiot (Tokyo)* 56:630-638.
51. Vodovar N, Vallenet D, Cruveiller S, Rouy Z, Barbe V, Acosta C, Cattolico L, Jubin C, Lajus A, Segurens B, Vacherie B, Wincker P, Weissenbach J, Lemaitre B, Medigue C, Boccard F (2006) Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium *Pseudomonas entomophila*. *Nat Biotechnol* 24:673-679.

52. Mendez-Perez D, Begemann MB, Pflieger BF (2011) Modular Synthase-Encoding Gene Involved in  $\alpha$ -Olefin Biosynthesis in *Synechococcus sp.* Strain PCC 7002. *Appl Environ Microbiol* 77:4264-4267.
53. Leger M, Gavalda S, Guillet V, van der Rest B, Slama N, Montrozier H, Mourey L, Quemard A, Daffe M, Marrakchi H (2009) The dual function of the *Mycobacterium tuberculosis* FadD32 required for mycolic acid biosynthesis. *Chem Biol* 16:510-519.
54. Donia MS, Fricke WF, Ravel J, Schmidt EW (2011) Variation in tropical reef symbiont metagenomes defined by secondary metabolism. *PLoS One* 6:e17897.
55. Metzger P, Largeau C (2005) *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids. *Appl Microbiol Biotechnol* 66:486-496.
56. Chan Young T, Largeau C, Casadevall E (1986) Biosynthesis of non-isoprenoid hydrocarbons by the microalga *Botryococcus braunii*. Evidences for an elongation-decarboxylation mechanism. Activation of decarboxylation. *New Journal of Chemistry* 10:701-707.
57. Rieley G, Teece MA, Peakman TM, Raven AM, Greene KJ, Clarke TP, Murray M, Leftley JW, Campbell C, Harris RP, Parkes RJ, Maxwell JR (1998) Long-chain alkenes of the haptophytes *Isochrysis galbana* and *Emiliania huxleyi*. *Lipids* 33:617-625.
58. Byres E, Alphey MS, Smith TK, Hunter WN (2007) Crystal structures of *Trypanosoma brucei* and *Staphylococcus aureus* mevalonate diphosphate decarboxylase inform on the determinants of specificity and reactivity. *J Mol Biol* 371:540-553.
59. Voynova NE, Fu Z, Battaile KP, Herdendorf TJ, Kim JJ, Mizioroko HM (2008) Human mevalonate diphosphate decarboxylase: characterization, investigation of the mevalonate diphosphate binding site, and crystal structure. *Arch Biochem Biophys* 480:58-67.
60. Bonanno JB, Edo C, Eswar N, Pieper U, Romanowski MJ, Ilyin V, Gerchman SE, Kycia H, Studier FW, Sali A, Burley SK (2001) Structural genomics of enzymes involved in sterol/isoprenoid biosynthesis. *Proc Natl Acad Sci U S A* 98:12896-12901.
61. Fortman JL, Chhabra S, Mukhopadhyay A, Chou H, Lee TS, Steen E, Keasling JD (2008) Biofuel alternatives to ethanol: pumping the microbial well. *Trends Biotechnol* 26:375-381.
62. Peralta-Yahya PP, Keasling JD (2010) Advanced biofuel production in microbes. *Biotechnol J* 5:147-162.
63. Rude MA, Schirmer A (2009) New microbial fuels: a biotech perspective. *Curr Opin Microbiol* 12:274-281.
64. Hill J, Nelson E, Tilman D, Polasky S, Tiffany D (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proc Natl Acad Sci U S A* 103:11206-11210.
65. Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, Del Cardayre SB, Keasling JD (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463:559-562.
66. Menzella HG, Reeves CD (2007) Combinatorial biosynthesis for drug development. *Curr Opin Microbiol* 10:238-245.

67. Wong FT, Khosla C (2012) Combinatorial biosynthesis of polyketides--a perspective. *Curr Opin Chem Biol* 16:117-123.
68. Gehret JJ, Gu L, Geders TW, Brown WC, Gerwick L, Gerwick WH, Sherman DH, Smith JL (2012) Structure and activity of DmmA, a marine haloalkane dehalogenase. *Protein Sci* 21:239-248.
69. Stucki G, Thuer M (1995) Experiences of a Large-Scale Application of 1,2-Dichloroethane Degrading Microorganisms for Groundwater Treatment. *Environ Sci Technol* 29:2339-2345.
70. Alcalde M, Ferrer M, Plou FJ, Ballesteros A (2006) Environmental biocatalysis: from remediation with enzymes to novel green processes. *Trends Biotechnol* 24:281-287.
71. Swanson PE (1999) Dehalogenases applied to industrial-scale biocatalysis. *Curr Opin Biotechnol* 10:365-369.
72. Prokop Z, Oplustil F, DeFrank J, Damborsky J (2006) Enzymes fight chemical weapons. *Biotechnol J* 1:1370-1380.
73. Campbell DW, Muller C, Reardon KF (2006) Development of a fiber optic enzymatic biosensor for 1,2-dichloroethane. *Biotechnol Lett* 28:883-887.
74. Bidmanova S, Chaloupkova R, Damborsky J, Prokop Z (2010) Development of an enzymatic fiber-optic biosensor for detection of halogenated hydrocarbons. *Anal Bioanal Chem* 398:1891-1898.
75. Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohana RF, Urh M, Simpson D, Mendez J, Zimmerman K, Otto P, Vidugiris G, Zhu J, Darzins A, Klaubert DH, Bulleit RF, Wood KV (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* 3:373-382.
76. Newman DJ, Cragg GM (2007) Natural products as sources of new drugs over the last 25 years. *J Nat Prod* 70:461-477.
77. Roongsawang N, Washio K, Morikawa M (2007) In vivo characterization of tandem C-terminal thioesterase domains in arthrofactin synthetase. *Chembiochem* 8:501-512.
78. DelProposto J, Majmudar CY, Smith JL, Brown WC (2009) Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein Expr Purif* 63:40-49.
79. Fischetti RF, Xu S, Yoder DW, Becker M, Nagarajan V, Sanishvili R, Hilgart MC, Stepanov S, Makarov O, Smith JL (2009) Mini-beam collimator enables microcrystallography experiments on standard beamlines. *J Synchrotron Radiat* 16:217-225.
80. Otwinowski Z, aM, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode *Methods in Enzymology* 276:307-326.
81. Zwart PH, Afonine PV, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, McKee E, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Storoni LC, Terwilliger TC, Adams PD (2008) Automated structure solution with the PHENIX suite. *Methods Mol Biol* 426:419-435.
82. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building

- new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58:1948-1954.
83. Terwilliger TC (2000) Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* 56:965-972.
  84. Terwilliger TC (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D Biol Crystallogr* 59:38-44.
  85. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126-2132.
  86. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240-255.
  87. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760-763.
  88. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 62:439-450.
  89. Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by C $\alpha$  geometry: phi,psi and C $\beta$  deviation. *Proteins* 50:437-450.
  90. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403-410.
  91. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
  92. DeLano WL. The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC. (2002). DeLano Scientific, Palo Alto, CA, USA.
  93. Schuttelkopf AW, van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* 60:1355-1363.
  94. Sanchez C, Du L, Edwards DJ, Toney MD, Shen B (2001) Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. *Chem Biol* 8:725-738.
  95. Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15:305-308.
  96. Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB (2010) Microbial biosynthesis of alkanes. *Science* 329:559-562.
  97. Rath VL, Verdugo D, Hemmerich S (2004) Sulfotransferase structural biology and inhibitor discovery. *Drug Discov Today* 9:1003-1011.
  98. Bojarova P, Williams SJ (2008) Sulfotransferases, sulfatases and formylglycine-generating enzymes: a sulfation fascination. *Curr Opin Chem Biol* 12:573-581.
  99. Gehret JJ, Gu L, Gerwick WH, Wipf P, Sherman DH, Smith JL (2011) Terminal alkene formation by the thioesterase of curacin A biosynthesis: structure of a decarboxylating thioesterase. *J Biol Chem* 286:14445-14454.
  100. Goss RJ, Shankar S, Fayad AA (2012) The generation of "unNatural" products: Synthetic biology meets synthetic chemistry. *Nat Prod Rep* 29:870-889.

101. Zabala AO, Cacho RA, Tang Y (2012) Protein engineering towards natural product synthesis and diversification. *J Ind Microbiol Biotechnol* 39:227-241.
102. Allali-Hassani A, Pan PW, Dombrovski L, Najmanovich R, Tempel W, Dong A, Loppnau P, Martin F, Thornton J, Edwards AM, Bochkarev A, Plotnikov AN, Vedadi M, Arrowsmith CH (2007) Structural and chemical profiling of the human cytosolic sulfotransferases. *PLoS Biol* 5:e97.
103. Cleland WW, Hengge AC (2006) Enzymatic mechanisms of phosphate and sulfate transfer. *Chem Rev* 106:3252-3278.
104. Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D (2007) Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci* 16:1569-1576.
105. Derewenda ZS, Vekilov PG (2006) Entropy and surface engineering in protein crystallization. *Acta Crystallogr D Biol Crystallogr* 62:116-124.
106. Bricogne G, BE, Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P SA, Smart O.S., Vonrhein C., Womack T.O. (2011) BUSTER version 2.10.0. Cambridge, United Kingdom: Global Phasing Ltd.
107. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66:12-21.
108. Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38:W545-549.
109. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. (2010).
110. Moon AF, Edavettal SC, Krahn JM, Munoz EM, Negishi M, Linhardt RJ, Liu J, Pedersen LC (2004) Structural analysis of the sulfotransferase (3-*O*-sulfotransferase isoform 3) involved in the biosynthesis of an entry receptor for herpes simplex virus 1. *J Biol Chem* 279:45185-45193.
111. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754-755.
112. Nelson SG, Peelen TJ, Wan Z (1999) Catalytic asymmetric acyl halide-aldehyde cyclocondensations. A strategy for enantioselective catalyzed cross aldol reactions. *J Am Chem Soc* 121:9742-9743.
113. Teramoto T, Sakakibara Y, Liu MC, Suiko M, Kimura M, Kakuta Y (2009) Snapshot of a Michaelis complex in a sulfonyl transfer reaction: Crystal structure of a mouse sulfotransferase, mSULT1D1, complexed with donor substrate and acceptor substrate. *Biochem Biophys Res Commun* 383:83-87.
114. Hitomi Y, Outten CE, O'Halloran TV (2001) Extreme zinc-binding thermodynamics of the metal sensor/regulator protein, ZntR. *J Am Chem Soc* 123:8614-8615.
115. Outten CE, O'Halloran TV (2001) Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. *Science* 292:2488-2492.
116. Moon AF, Xu Y, Woody SM, Krahn JM, Linhardt RJ, Liu J, Pedersen LC (2012) Dissecting the substrate recognition of 3-*O*-sulfotransferase for the biosynthesis of anticoagulant heparin. *Proc Natl Acad Sci U S A* 109:5265-5270.
117. Cook I, Wang T, Falany CN, Leyh TS (2012) A Nucleotide-Gated Molecular Pore Selects Sulfotransferase Substrates. *Biochemistry*.
118. Caffrey P (2005) The stereochemistry of ketoreduction. *Chem Biol* 12:1060-1062.

119. Pedersen LC, Petrotchenko E, Shevtsov S, Negishi M (2002) Crystal structure of the human estrogen sulfotransferase-PAPS complex: evidence for catalytic role of Ser137 in the sulfuryl transfer reaction. *J Biol Chem* 277:17928-17932.
120. Vaillancourt FH, Yeh E, Vosburg DA, O'Connor SE, Walsh CT (2005) Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* 436:1191-1194.
121. McCarthy JG, Eisman EB, Kulkarni S, Gerwick L, Gerwick WH, Wipf P, Sherman DH, Smith JL (2012) The Structural Basis of Functional Group Activation by Sulfotransferases in Complex Metabolic Pathways. *ACS Chem Biol*.
122. Zhang YM, Rao MS, Heath RJ, Price AC, Olson AJ, Rock CO, White SW (2001) Identification and analysis of the acyl carrier protein (ACP) docking site on beta-ketoacyl-ACP synthase III. *J Biol Chem* 276:8231-8238.
123. Anand S, Mohanty D (2012) Modeling holo-ACP:DH and holo-ACP:KR complexes of modular polyketide synthases: a docking and molecular dynamics study. *BMC Struct Biol* 12:10.
124. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98:10037-10041.
125. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22:195-201.
126. Tang Y, Kim CY, Mathews, II, Cane DE, Khosla C (2006) The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc Natl Acad Sci U S A* 103:11124-11129.
127. Maier T, Leibundgut M, Ban N (2008) The crystal structure of a mammalian fatty acid synthase. *Science* 321:1315-1322.
128. Arora P, Goyal A, Natarajan VT, Rajakumara E, Verma P, Gupta R, Yousuf M, Trivedi OA, Mohanty D, Tyagi A, Sankaranarayanan R, Gokhale RS (2009) Mechanistic and functional insights into fatty acid activation in *Mycobacterium tuberculosis*. *Nat Chem Biol* 5:166-173.
129. Zhang Z, Zhou R, Sauder JM, Tonge PJ, Burley SK, Swaminathan S (2011) Structural and functional studies of fatty acyl adenylate ligases from *E. coli* and *L. pneumophila*. *J Mol Biol* 406:313-324.
130. Liu X, Sheng J, Curtiss R, 3rd (2011) Fatty acid production in genetically modified cyanobacteria. *Proc Natl Acad Sci U S A* 108:6899-6904.
131. Peralta-Yahya PP, Zhang F, del Cardayre SB, Keasling JD (2012) Microbial engineering for the production of advanced biofuels. *Nature* 488:320-328.
132. Koudelakova T, Chovancova E, Brezovsky J, Monincova M, Fortova A, Jarkovsky J, Damborsky J (2011) Substrate specificity of haloalkane dehalogenases. *Biochem J* 435:345-354.
133. Janssen DB (2004) Evolving haloalkane dehalogenases. *Curr Opin Chem Biol* 8:150-159.
134. Chovancova E, Kosinski J, Bujnicki JM, Damborsky J (2007) Phylogenetic analysis of haloalkane dehalogenases. *Proteins* 67:305-316.
135. Verschueren KH, Kingma J, Rozeboom HJ, Kalk KH, Janssen DB, Dijkstra BW (1993) Crystallographic and fluorescence studies of the interaction of haloalkane

- dehalogenase with halide ions. Studies with halide compounds reveal a halide binding site in the active site. *Biochemistry* 32:9031-9037.
136. Verschueren KH, Seljee F, Rozeboom HJ, Kalk KH, Dijkstra BW (1993) Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. *Nature* 363:693-698.
  137. Pikkemaat MG, Ridder IS, Rozeboom HJ, Kalk KH, Dijkstra BW, Janssen DB (1999) Crystallographic and kinetic evidence of a collision complex formed during halide import in haloalkane dehalogenase. *Biochemistry* 38:12052-12061.
  138. Hesseler M, Bogdanovic X, Hidalgo A, Berenguer J, Palm GJ, Hinrichs W, Bornscheuer UT (2011) Cloning, functional expression, biochemical characterization, and structural analysis of a haloalkane dehalogenase from *Plesiocystis pacifica* SIR-1. *Appl Microbiol Biotechnol* 91:1049-1060.
  139. Newman J, Peat TS, Richard R, Kan L, Swanson PE, Affholter JA, Holmes IH, Schindler JF, Unkefer CJ, Terwilliger TC (1999) Haloalkane dehalogenases: structure of a *Rhodococcus* enzyme. *Biochemistry* 38:16105-16114.
  140. Stsiapanava A, Dohnalek J, Gavira JA, Kutý M, Koudelakova T, Damborsky J, Kuta Smatanova I (2010) Atomic resolution studies of haloalkane dehalogenases DhaA04, DhaA14 and DhaA15 with engineered access tunnels. *Acta Crystallogr D Biol Crystallogr* 66:962-969.
  141. Marek J, Vevodova J, Smatanova IK, Nagata Y, Svensson LA, Newman J, Takagi M, Damborsky J (2000) Crystal structure of the haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26. *Biochemistry* 39:14082-14086.
  142. Oakley AJ, Prokop Z, Bohac M, Kmunicek J, Jedlicka T, Monincova M, Kuta-Smatanova I, Nagata Y, Damborsky J, Wilce MC (2002) Exploring the structure and activity of haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26: evidence for product- and water-mediated inhibition. *Biochemistry* 41:4847-4855.
  143. Streltsov VA, Prokop Z, Damborsky J, Nagata Y, Oakley A, Wilce MC (2003) Haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* UT26: X-ray crystallographic studies of dehalogenation of brominated substrates. *Biochemistry* 42:10104-10112.
  144. Oakley AJ, Klvana M, Otyepka M, Nagata Y, Wilce MC, Damborsky J (2004) Crystal structure of haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* UT26 at 0.95 Å resolution: dynamics of catalytic residues. *Biochemistry* 43:870-878.
  145. Prokop Z, Sato Y, Brezovsky J, Mozga T, Chaloupkova R, Koudelakova T, Jerabek P, Stepankova V, Natsume R, van Leeuwen JG, Janssen DB, Florian J, Nagata Y, Senda T, Damborsky J (2010) Enantioselectivity of haloalkane dehalogenases and its modulation by surface loop engineering. *Angew Chem Int Ed Engl* 49:6111-6115.
  146. Mazumdar PA, Hulecki JC, Cherney MM, Garen CR, James MN (2008) X-ray crystal structure of *Mycobacterium tuberculosis* haloalkane dehalogenase Rv2579. *Biochim Biophys Acta* 1784:351-362.
  147. Guerrero SA, Hecht HJ, Hofmann B, Biebl H, Singh M (2001) Production of selenomethionine-labelled proteins using simplified culture conditions and generally applicable host/vector systems. *Appl Microbiol Biotechnol* 56:718-723.

148. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67:271-281.
149. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62:72-82.
150. Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 55:849-861.
151. Ho BK, Gruswitz F (2008) HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct Biol* 8:49.
152. Schindler JF, Naranjo PA, Honabberger DA, Chang CH, Brainard JR, Vanderberg LA, Unkefer CJ (1999) Haloalkane dehalogenases: steady-state kinetics and halide inhibition. *Biochemistry* 38:5772-5778.
153. Marvanova S, Nagata Y, Wimmerova M, Sykorova J, Hynkova K, Damborsky J (2001) Biochemical characterization of broad-specificity enzymes using multivariate experimental design and a colorimetric microplate assay: characterization of the haloalkane dehalogenase mutants. *J Microbiol Methods* 44:149-157.
154. Gordon L, Chervonenkis AY, Gammerman AJ, Shahmuradov IA, Solovyev VV (2003) Sequence alignment kernel for recognition of promoter regions. *Bioinformatics* 19:1964-1971.
155. Bosma T, Pikkemaat MG, Kingma J, Dijk J, Janssen DB (2003) Steady-state and pre-steady-state kinetic analysis of halopropane conversion by a rhodococcus haloalkane dehalogenase. *Biochemistry* 42:8047-8053.
156. Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, Taly JF, Notredame C (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* 39:W13-17.
157. Heinig M, Frishman D (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res* 32:W500-502.
158. Hasan K, Fortova A, Koudelakova T, Chaloupkova R, Ishitsuka M, Nagata Y, Damborsky J, Prokop Z (2011) Biochemical characteristics of the novel haloalkane dehalogenase DatA, isolated from the plant pathogen *Agrobacterium tumefaciens* C58. *Appl Environ Microbiol* 77:1881-1884.
159. Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick WH (2004) Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem Biol* 11:817-833.
160. Trivedi OA, Arora P, Sridharan V, Tickoo R, Mohanty D, Gokhale RS (2004) Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. *Nature* 428:441-445.
161. Olano C, Mendez C, Salas JA (2010) Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Nat Prod Rep* 27:571-616.

162. Zhang H, Boghigian BA, Armando J, Pfeifer BA (2011) Methods and options for the heterologous production of complex natural products. *Nat Prod Rep* 28:125-151.
163. Fisch KM, Bakeer W, Yakasai AA, Song Z, Pedrick J, Wasil Z, Bailey AM, Lazarus CM, Simpson TJ, Cox RJ (2011) Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite. *J Am Chem Soc* 133:16635-16641.
164. Machado IM, Atsumi S (2012) Cyanobacterial biofuel production. *J Biotechnol*.