Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Functional convergence of structurally distinct thioesterases from cyanobacteria and plants involved in phylloquinone biosynthesis

Fabienne Furt,[a] William J. Allen,[b] Joshua R. Widhalm,[a] Peter Madzelan,[c] Robert C. Rizzo,[b] Gilles Basset[a] and Mark A. Wilson[c]*

[a]Center for Plant Science Innovation and Departments of Agronomy and Horticulture and Department of Biochemistry, University of Nebraska, Lincoln, NE 68588, USA, [b]Department of Applied Mathematics and Statistics, Stony Brook University, Math Tower 1-111, Stony Brook, NY 11794, USA, and [c]Department of Biochemistry and the Redox Biology Center, University of Nebraska, N118 Beadle Center, Lincoln, NE 68588, USA

Correspondence e-mail: mwilson13@unl.edu

Received 8 April 2013
Accepted 6 June 2013

PDB References: Slr0204, 4k00; AtDHNAT1, 4k02;

The synthesis of phylloquinone (vitamin K_1) in photosynthetic organisms requires a thioesterase that hydrolyzes 1,4-dihydroxy-2-naphthoyl-CoA (DHNA-CoA) to release 1,4-dihydroxy-2-naphthoate (DHNA). Cyanobacteria and plants contain distantly related hotdog-fold thioesterases that catalyze this reaction, although the structural basis of these convergent enzymatic activities is unknown. To investigate this, the crystal structures of hotdog-fold DHNA-CoA thioesterases from the cyanobacterium *Synechocystis* (Slr0204) and the flowering plant *Arabidopsis thaliana* (AtDHNAT1) were determined. These enzymes form distinct homotetramers and use different active sites to catalyze hydrolysis of DHNA-CoA, similar to the 4-hydroxybenzoyl-CoA (4-HBA-CoA) thioesterases from *Pseudomonas* and *Arthrobacter*. Like the 4-HBA-CoA thioesterases, the DHNA-CoA thioesterases contain either an active-site aspartate (Slr0204) or glutamate (AtDHNAT1) that are predicted to be catalytically important. Computational modeling of the substrate-bound forms of both enzymes indicates the residues that are likely to be involved in substrate binding and catalysis. Both enzymes are selective for DHNA-CoA as a substrate, but this selectivity is achieved using divergent predicted binding strategies. The Slr0204 binding pocket is predominantly hydrophobic and closely conforms to DHNA, while that of AtDHNAT1 is more polar and solvent-exposed. Considered in light of the related 4-HBA-CoA thioesterases, these structures indicate that hotdog-fold thioesterases using either an active-site aspartate or glutamate diverged into distinct clades prior to the evolution of strong substrate specificity in these enzymes.

1. Introduction

The hotdog fold, named for its distinctive resemblance to a hot dog on a bun (Leesong *et al.*, 1996), is a widely distributed protein architecture that is capable of supporting diverse activities. Thioesterase activity against acyl-coenzyme A (CoA) substrates is most commonly associated with this fold; however, dehydratases as well as non-enzymatic members of the family are also known (Dillon & Bateman, 2004). Several hotdog thioesterases that catalyze the hydrolysis of acylated CoA substrates have been characterized structurally and a unified scheme for categorizing thioesterases based on sequence characteristics shared among various subgroups has been proposed (Cantu *et al.*, 2010). The core motif of hotdog-fold thioesterases comprises an antiparallel β-sheet wrapped around an α-helix in each of two monomers that associate into an obligate dimer across shared strands of the β-sheet (Leesong *et al.*, 1996). Although the dimer is the fundamental folded unit of these thioesterases, they often further associate to form homotetramers, which appears to be their most

common oligomeric state (Benning *et al.*, 1998; Thoden *et al.*, 2003; Cantu *et al.*, 2010; Dillon & Bateman, 2004).

Hotdog-fold acyl-CoA thioesterases possess similar gross active-site characteristics owing to their shared functional constraints, including a tunnel that binds the extended pantothenyl chain of CoA and a cavity that binds the acyl group and contains the key catalytic residues. Despite these broad similarities, the details of their active sites are surprisingly varied (Dillon & Bateman, 2004). Much of what is known about catalysis by these enzymes is based upon the extensive characterization of two archetypal 4-hydroxybenzoyl-CoA thioesterases (4-HBTs) from *Pseudomonas* sp. strain CBS3 (Benning *et al.*, 1998; Chang *et al.*, 1992; Thoden *et al.*, 2002) and *Arthrobacter* sp. strain SU (Thoden *et al.*, 2003). Both of these distantly related enzymes have recently been shown to act through the formation of an enzyme-linked mixed anhydride rather than by the direct activation of water to hydrolyze the thioester (Song *et al.*, 2012; Zhuang *et al.*, 2012). The anhydride intermediate is formed by nucleophilic attack of a carboxylate side chain at the thioester moiety of the substrate and therefore requires an acidic residue in the active site. Given the intrinsic lability of the thioester linkage, it is not clear why a covalent catalytic strategy has evolved in these divergent hotdog-fold thioesterases. Nevertheless, it is observed in structurally distinct enzymes employing different active-site residues to accomplish the same chemical task, suggesting that it may have evolved by convergence.

Hotdog-fold thioesterases play a key and recently discovered role in vitamin K biosynthesis (Widhalm *et al.*, 2009, 2012). Vitamin K comprises two related classes of vitamers: phylloquinone (K_1), which is synthesized by plants and some cyanobacteria, and menaquinone (K_2), which is synthesized by certain types of bacteria (Collins & Jones, 1981; Sakuragi & Bryant, 2006; van Oostende *et al.*, 2011). In both cases, dedicated hotdog-fold CoA thioesterases hydrolyze 1,4-dihydroxynaphthoyl-CoA (DHNA-CoA) to liberate 1,4-dihydroxy-2-naphthoate (DHNA) for subsequent isoprenylation and methylation to form vitamin K (Widhalm *et al.*, 2009, 2012). Various organisms are capable of synthesizing K vitamers; however, they possess distinct distantly related DHNA-CoA thioesterases (DHNATs) for this task. Interestingly, this is true even among organisms that make the same type of vitamin K. For example, both the model flowering plant *Arabidopsis thaliana* and the cyanobacterium *Synechocystis* sp. PCC 6803 make phylloquinone (vitamin K_1), but their DHNA-CoA thioesterases are only ∼13% identical at the amino-acid level (Widhalm *et al.*, 2009, 2012). This functional convergence of distantly related thioesterases is reminiscent of the aforementioned *Pseudomonas* and *Arthrobacter* 4-HBTs. Like these prototypical thioesterases, the DHNATs catalyze the same chemical transformation of the same substrate but share little identity at the amino-acid level.

We have determined the crystal structures of DHNATs from both *A. thaliana* (AtDHNAT1) and *Synechocystis* sp. PCC 6803 (Slr0204). These enzymes are structurally similar at the monomer and dimer levels but form distinct homotetramers. The structures and active sites of plant and cyanobacterial DHNATs differ from each other but are clearly related to the *Pseudomonas* and *Arthrobacter* 4-HBTs. Computational modeling of the bound substrate into the active sites of both of these enzymes identifies residues that are potentially important for substrate binding and catalysis. These residues are conserved within each representative clade but differ between them, supporting the proposal that they are functionally significant. The structural relationship between the 4-HBTs and DHNATs confirms a previous phylogenetic analysis (Widhalm *et al.*, 2012) suggesting that the division between these two clades of hotdog-fold thioesterases preceded the evolution of the stringent substrate specificity of these enzymes.

2. Materials and methods

2.1. Protein expression and purification

The genes for Slr0204 and AtDHNAT1 were cloned from their respective organisms into the bacterial expression vector pET15b (Novagen) between the *Nde*I and *Xho*I restriction sites. These constructs produce recombinant proteins bearing an N-terminal hexahistidine tag that can be removed by cleavage with thrombin, leaving the residual sequence GSH– at the N-termini of the mature proteins. All clones were verified by DNA sequencing (Operon). Proteins were expressed in *Escherichia coli* strain BL21(DE3) (EMD Millipore) by growing cells to mid-log phase (OD_{600} = 0.4–0.6) in LB medium supplemented with 100 µg ml^{-1} ampicillin at 310 K with shaking. Protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 m*M*, followed by incubation with shaking for an additional 3 h. Cells were harvested by centrifugation, frozen using liquid nitrogen and stored at 193 K until needed.

Cell pellets were thawed on ice and resuspended in lysis buffer (50 m*M* HEPES pH 7.5, 300 m*M* NaCl, 10 m*M* imidazole) at a 5:1 volume:mass ratio of buffer to cell pellet. Cell lysis was initiated by adding hen egg-white lysozyme (Thermo Fisher) to a final concentration of 1 mg ml^{-1} and incubating on ice for ∼45 min, after which lysis was completed by sonication. Cell debris was removed from the crude lysate by centrifugation at 12 000*g* for 30 min and the clarified lysate was mixed with His-Select Ni^{2+} metal-affinity resin (Sigma) at 277 K for 15 min in order to bind the hexahistidine-tagged protein. The column was washed with lysis buffer supplemented with 10 m*M* imidazole (20 m*M* final concentration) until no protein was detected in the flowthrough using Bradford's reagent. Resin-bound protein was eluted with lysis buffer supplemented with 200 m*M* imidazole and mixed with high-purity bovine thrombin (MP Biomedicals) at one unit per milligram of eluted protein in order to cleave the N-terminal hexahistidine tag. The mixture was dialyzed against storage buffer (25 m*M* HEPES pH 7.5, 100 m*M* KCl) at 277 K overnight. Thrombin was removed by passage over benzamidine Sepharose resin (GE Healthcare Life Sciences) and the purified proteins were concentrated to 17–20 mg ml^{-1} using

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Synechocystis DHNAT (PDB entry 4k00)	*Arabidopsis* DHNAT (PDB entry 4k02)
Crystallization and crystal data		
Temperature (K)	298	298
Crystal size (mm)	0.4 × 0.1 × 0.1	0.3 × 0.3 × 0.3
Solvent content (%)	47.1	61.8
Data collection		
Space group	$P4_12_12$	$P3_121$
Unit-cell parameters		
$a = b$ (Å)	54.50	99.53
c (Å)	191.09	61.26
$\alpha = \beta$ (°)	90	90
γ (°)	90	120
No. of molecules in unit cell (Z)	2	2
Diffraction source	BioCARS 14-BM-C, APS	GM/CA 23-ID-D, APS
Wavelength (Å)	0.900	1.033
Detector	ADSC Q315 CCD	MAR Mosaic 300 CCD
Temperature (K)	110	110
Resolution range (Å)	37.77–1.90 (1.97–1.90)	28.88–1.90 (1.97–1.90)
Total No. of reflections	183520	951530
No. of unique reflections	23811	26984
Completeness (%)	100.0 (100.0)	96.4 (84.1)
Multiplicity	7.7 (7.9)	35.3 (22.5)
$\langle I\rangle/\langle\sigma(I)\rangle$	17.4 (2.2)	48.8 (2.4)
$R_{\rm merge}$†	0.129 (0.990)	0.058 (0.764)
Model refinement		
Refinement software	*REFMAC* v.5.7.0032	*REFMAC* v.5.7.0032
Resolution range (Å)	35.92–1.90 (1.95–1.90)	28.88–1.90 (1.95–1.90)
Completeness (%)	99.9 (99.8)	95.7 (79.4)
No. of reflections, working set	22535 (1632)	25341 (1539)
No. of reflections, test set	1169 (92)	1361 (91)
$R_{\rm cryst}$‡	0.170 (0.232)	0.183 (0.304)
$R_{\rm free}$§	0.201 (0.248)	0.211 (0.342)
Maximum-likelihood estimated standard uncertainty (Å)	0.091	0.097
No. of non-H atoms		
Protein	2152	1914
Other	4	0
Water	183	52
R.m.s. deviations		
Bonds (Å)	0.006	0.008
Angles (°)	0.927	1.618
Overall average B factor (Å^2)	28.2	66.2
Ramachandran plot analysis, residues in (%)		
Most favored regions	99.6	98.8
Additionally allowed regions	100	100
Disallowed regions	0	0

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where i is the ith observation of a reflection with indices hkl and angle brackets indicate the average over all i observations. ‡ $R_{\rm cryst} = \sum_{hkl}\big||F_{\rm obs}| - |F_{\rm calc}|\big|/\sum_{hkl}|F_{\rm obs}|$, where $F_{\rm calc}$ is the calculated structure-factor amplitude with indices hkl and $F_{\rm obs}$ is the observed structure-factor amplitude with indices hkl. § $R_{\rm free}$ is calculated as $R_{\rm cryst}$, where the $F_{\rm obs}$ are taken from a test set comprising 5% of the data that were excluded from the refinement.

stirred-cell and centrifugal concentrators with 10 kDa molecular-weight cutoffs (EMD Millipore). Protein concentrations were determined from the absorption at 280 nm using extinction coefficients of 18 450 M^{-1} cm^{-1} for Slr0204 and 11 000 M^{-1} cm^{-1} for AtDHNAT1, as calculated from the amino-acid sequences of these proteins using the *ProtParam* tool at the ExPASy Bioinformatics Portal (Gasteiger *et al.*, 2005). Purified Slr0204 migrated as a single species on overloaded Coomassie Blue-stained SDS–PAGE. In contrast, AtDHNAT1 migrated on SDS–PAGE as a mixture of aggregated SDS-resistant high-molecular-weight species and the monomeric species, despite showing no evidence of precipitation during purification and concentration. Both proteins were divided into 50–100 µl aliquots, quickly frozen using liquid nitrogen and stored at 193 K until needed.

2.2. Protein crystallization and data collection

Crystallization conditions for both proteins were determined with commercial sparse-matrix screens in 400 nl drops using an Art Robbins Gryphon liquid-handling robot (Art Robbins Instruments). Conditions delivering crystals were further optimized using sitting-drop vapor diffusion by mixing 2 µl protein solution and 2 µl reservoir solution at room temperature. Crystals of Slr0204 grew from 1.35 *M* NaH_2PO_4/0.8 *M* K_2HPO_4, 400 m*M* Li_2SO_4, 100 m*M* CAPS pH 10.5 in 1–2 d. Crystals of Slr0204 were removed from their drops with a nylon loop, cryoprotected by transfer through solutions of sodium malonate pH 7.4 ranging from 2.1 to 3.4 *M* in ∼0.5 *M* increments (Holyoak *et al.*, 2003) and cryocooled by immersion in liquid nitrogen. Crystals of AtDHNAT1 were grown from 2.7 *M* sodium formate, 100 m*M* Tris–HCl pH 7.0 in 3–5 d. Notably, crystals of AtDHNAT1 could only be obtained after adding 10 m*M* *n*-octyl-β-D-glucoside (β-OG) to the protein before crystallization, which may lead to a more homogeneous sample by solubilizing the high-molecular-weight aggregated species observed on SDS–PAGE (see above). AtDHNAT1 crystals were cryoprotected by transferring the crystals to 10 µl reservoir solution, followed by the sequential addition of six 0.5 µl aliquots of 9 *M* sodium formate (to give a final formate concentration of 4.15 *M*) with an equilibration period of 20–30 s between each addition. More conventional cryoprotection by serial transfer through solutions of elevated sodium formate concentration resulted in crystal cracking. As for Slr0204, the AtDHNAT1 crystals were cryocooled by plunging them into liquid nitrogen.

Diffraction data were collected using the oscillation method from single crystals maintained at 110 K on BioCARS beamline 14-BMC (Slr0204) and GM/CA-CAT beamline 23-ID-D (AtDHNAT1) at the Advanced Photon Source (APS; Argonne, Illinois, USA). The data were recorded on an ADSC Q315 CCD detector (Area Detector Systems Corporation) for crystals of Slr0204 and a MAR Mosaic 300 detector (Rayonix)

for crystals of AtDHNAT1. The diffraction data from crystals of AtDHNAT1 had a large dynamic range and thus the data were collected in separate sweeps with an attenuated incident beam used to record the lower resolution reflections without overloads. These two data sets were collected with a 0.5 Å region of overlap in resolution in order to measure reflections common to both data sets and thus permit final scaling and merging. All diffraction data were indexed and scaled using *HKL*-2000 (Otwinowski & Minor, 1997); the final statistics are provided in Table 1. The $R_{\rm merge}$ values in the highest resolution bins are higher than ideal for both data sets; however, recent work has shown that the commonly used criteria for deciding where to cut the data are overly conservative and discard data that contain useful signal (Diederichs & Karplus, 2012). Therefore, we included these data in refinement. In addition, we note that the overall $R_{\rm merge}$ value of 13% for Slr0204 is higher than expected and high $R_{\rm merge}$ values were observed for multiple data sets collected from these crystals at the APS (Argonne, Illinois, USA) as well as using a rotating-anode X-ray source. The $R_{\rm merge}$ value can be reduced by 1–3% by merging the data in lower symmetry space groups, but the reasonable R factors for models refined against data merged in the higher symmetry space group (Table 1) indicate that either $P4_12_12$ is the true space group or that an NCS operator exists in a lower symmetry space group that is nearly perfectly coincident with a crystallographic symmetry operator in $P4_12_12$.

2.3. Structure determination and refinement

Phases were obtained for Slr0204 by molecular replacement in *Phaser* (McCoy *et al.*, 2007) using the dimeric form of a putative thioesterase from *Prochlorococcus marinus* strain MIT 9313 (PDB entry 2hx5; 33% sequence identity; Joint Center for Structural Genomics, unpublished work) as a search model. A solution could only be obtained from a search model lacking residues 1–10, 54–60 and 135–144. A test set of reflections was sequestered from each data set and used for calculation of the $R_{\rm free}$ value (Brünger, 1992). The initial model was automatically built into model-phased electron-density maps using *PHENIX* (Adams *et al.*, 2010; Terwilliger *et al.*, 2008) and the problematic regions of the resulting model were manually improved using real-space refinement as implemented in *Coot* (Emsley & Cowtan, 2004). The resulting model was refined in *REFMAC*5 (Murshudov *et al.*, 2011), which is part of the *CCP*4 suite (Winn *et al.*, 2011), using restrained maximum-likelihood refinement with an amplitude-based target (Murshudov *et al.*, 1997), local noncrystallographic symmetry restraints between the two monomers in the asymmetric unit, riding H atoms and a translation–libration–screw (TLS) model for atomic displacement parameters (ADPs) that treated each protein monomer as a distinct rigid body (Winn *et al.*, 2001).

Molecular replacement was also used to obtain phases for AtDHNAT1. The monomeric structure of a putative thioesterase from *Haemophilus influenzae* (PDB entry 1sc0; 38% sequence identity; Northeast Structural Genomics Consortium, unpublished work) was used as a search model in *Phaser* (McCoy *et al.*, 2007). The initial AtDHNAT1 model was a homology model generated by *SWISS-MODEL* (Arnold *et al.*, 2006) using 1sc0 as a template, which was then manually improved in *Coot* (Emsley & Cowtan, 2004). The AtDHNAT1 model was subjected to restrained amplitude-based maximum-likelihood refinement in *REFMAC*5 (Murshudov *et al.*, 2011) using local noncrystallographic symmetry restraints between the two monomers in the asymmetric unit, riding H atoms and a TLS model for ADPs (Winn *et al.*, 2001). Each monomer was treated as a separate rigid body for TLS refinement. Notably, there were two elongated features present in 4–5σ $mF_{\rm o} - DF_{\rm c}$ electron density near Glu24 of symmetry-related molecules that could not be modeled. It is possible that this electron density could result from partially ordered β-octyl glucoside molecules, but this is speculative. The final models for both Slr0204 and AtDHNAT1 were validated using *Coot* (Emsley & Cowtan, 2004) and *MolProbity* (Chen *et al.*, 2010) and final model statistics are reported for both structures in Table 1. All figures were produced using *POVScript+* (Fenn *et al.*, 2003).

2.4. Modeling of bound DHNA-CoA in the substrate-free enzymes

The tetrameric apo forms of Slr0204 and AtDHNAT1 were superimposed with the crystal structures of tetrameric 4-HBTs from *Pseudomonas* sp. strain CBS3 (PDB entry 1lo9; Thoden *et al.*, 2002) and 4-HBT from *Arthrobacter* sp. strain SU (PDB code 1q4u; Thoden *et al.*, 2003), respectively. 1lo9 is the structure of the catalytically crippled D17N mutant of *Pseudomonas* 4-HBT bound to its substrate, hydroxybenzoyl-CoA. 1q4u is the structure of wild-type *Arthrobacter* 4-HBT bound to a competitive inhibitor, hydroxybenzyl-CoA, which is a nonhydrolyzable thioether analogue of its 4-hydroxybenzoyl-CoA (4-HBA-CoA) thioester substrate. Both represent the probable structures of the Michaelis complexes for these enzymes. Subsequently, the protein atoms from 1lo9 and 1q4u were removed, thus generating initial models of the complexes between tetrameric Slr0204 and four copies of 4-HBA-CoA (from 1lo9) and of tetrameric AtDHNAT1 bound to four copies of 4-hydroxybenzyl-CoA (from 1q4u). As these are not the natural substrates for the DHNA-CoA thioesterases, the 4-HBA or 4-hydroxybenzyl group of each ligand was changed to a 1,4-dihydroxy-2-naphthoyl (DHNA) moiety using the modeling software *Molecular Operating Environment* (*MOE*; Chemical Computing Group), thereby forming initial models of the DHNA-CoA-bound proteins. Using the *AMBER*11 tools *leap* and *antechamber* (Case *et al.*, 2010), FF99SB parameters (Hornak *et al.*, 2006) were assigned to the proteins and GAFF parameters (Wang *et al.*, 2004) with AM1-BCC charges (Jakalian *et al.*, 2000, 2002) were assigned to the ligands. Gas-phase energy minimizations were performed in *AMBER*11 in order to eliminate any unfavorable interactions that may have resulted from model construction or protonation. In the initial 1000 steps of minimization, all non-H protein atoms were restrained with a 5.0 kcal mol^{-1} Å^{-2} harmonic restraint (1 cal = 4.184 J), while protein H atoms and all ligand atoms

were allowed to move. Following the restrained minimization, an additional 1000 steps of minimization were performed with no harmonic restraints. Single-point energy calculations were performed on the minimized complexes in the program *DOCK* v.6.5 (Lang *et al.*, 2009) and were decomposed into per-residue van der Waals and electrostatic interactions as described previously (Balius *et al.*, 2011).

2.5. Sedimentation-equilibrium centrifugation for molecular-mass determination

Sedimentation-equilibrium ultracentrifugation was used to determine the solution molecular mass of both Slr0204 and AtDHNAT1 as described previously (Lakshminarasimhan *et al.*, 2010) with minor modifications. Briefly, both samples were dialysed into 25 m*M* HEPES pH 7.5, 100 m*M* KCl at 277 K overnight. This buffer was supplemented with 1 m*M* β-octyl glucoside for AtDHNAT1. The sample absorbance at 275 nm as a function of radius was measured at three rotor speeds (1.0×10^4, 1.5×10^4 and 2.0×10^4 rev min^{-1}) after 30 h of equilibration at each speed, followed by final data acquisition 2 h later. No differences were observed between the 30 and 32 h data sets, confirming that equilibrium had been reached at each speed. Two different concentrations of each sample were used (0.1 and 0.2 mg ml^{-1} for Slr0204 and 0.2 and 0.4 mg ml^{-1} for AtDHNAT1), thus producing six total data sets per protein for global fitting of molecular mass in *Origin*.

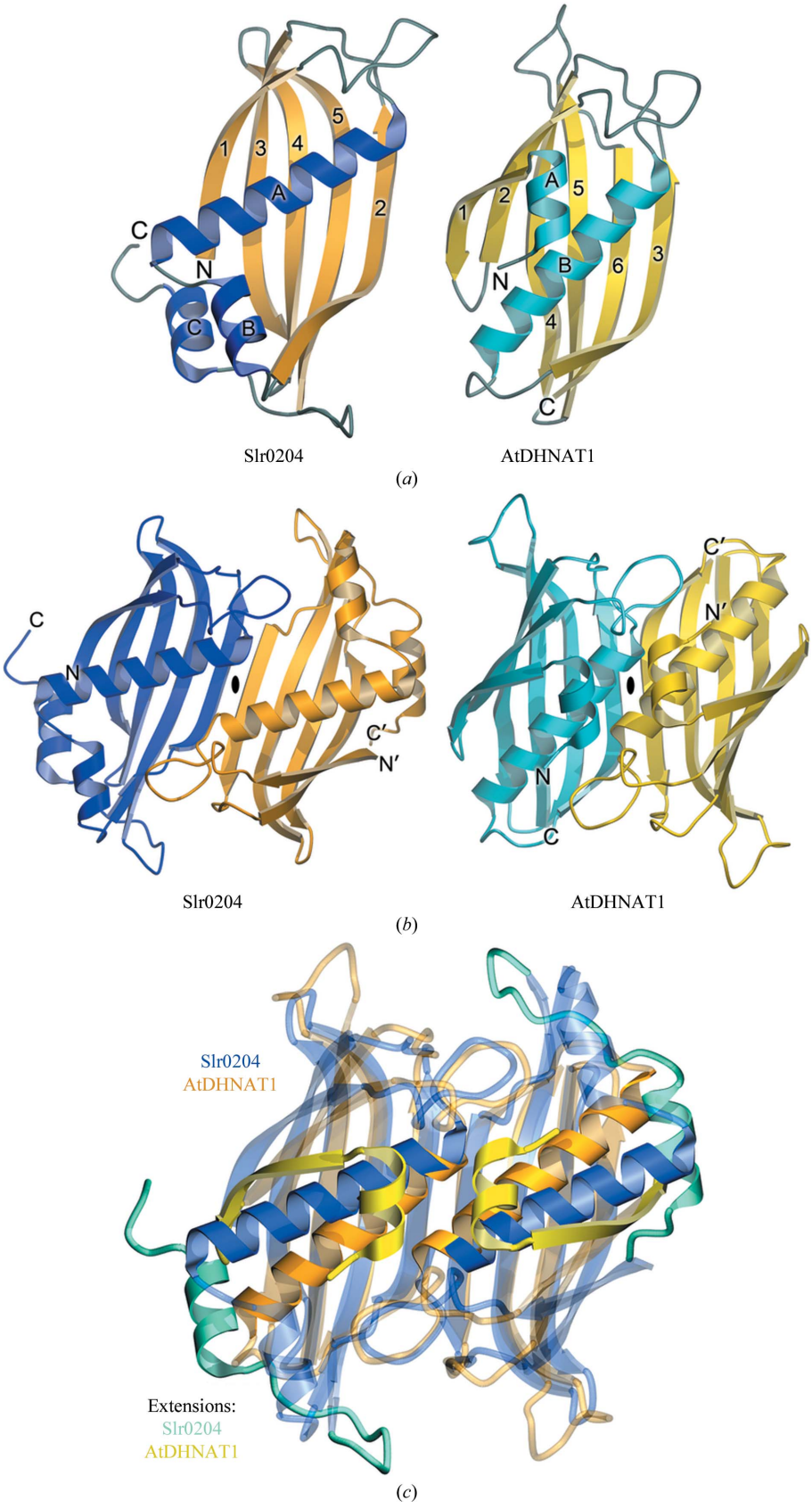

Figure 1
Structural comparison of the cyanobacterial and plant DHNA-CoA thioesterase monomers and dimers. In (*a*), ribbon diagrams for the cyanobacterial thioesterase Slr0204 and the plant thioesterase AtDHNAT1 are shown with sequentially numbered strands and lettered helices. In (*b*), the dimers of these two proteins (labeled) are shown with each monomer represented in a different color and the location of the twofold axis indicated by a solid ellipse. (*c*) shows the superimposition of the Slr0204 (blue/cyan) and AtDHNAT1 (orange/yellow) dimers, with the variable structural elements in each protein rendered as opaque and the conserved elements as semi-transparent. The N-terminal extension of AtDHNAT1 (yellow) displaces the 'hotdog' helix (orange) relative to its position in Slr0204 (blue).

3. Results and discussion

3.1. Comparison of Slr0204 and AtDHNAT1 monomer and dimer structures

Despite a low shared sequence identity of ~13%, the crystal structures of Slr0204 and AtDHNAT1 are similar, with a core monomeric C^{α} r.m.s.d. value of 1.6 Å as calculated using secondary-structural matching (Krissinel & Henrick, 2004) in *Coot* (Emsley & Cowtan, 2004). As in other hotdog-fold proteins, the core structures of both proteins are defined by a 5–6-stranded antiparallel β-sheet that partially wraps around a prominent five-turn α-helix (Fig. 1*a*). Despite this expected similarity in overall fold, AtDHNAT1 has a 27-amino-acid extension at its N-terminus and Slr0204 has a 25-amino-acid extension at its C-terminus (residues 113–138 in Slr0204). The C-terminal extension of Slr0204 forms a loop region with some β-hairpin character followed by an α-helix, while the N-terminal extension of AtDHNAT1 forms an additional α-helix followed by a β-strand (Fig. 1*a*). The significance of these variable extensions for oligomerization of the proteins is discussed below.

As observed in other thioesterases, both proteins dimerize across a β-sheet spanning two monomers (Dillon & Bateman, 2004; Fig. 1*b*). The dimeric form of these proteins places the hotdog α-helices (helix *A* in Slr0204, helix *B* in AtDHNAT1) in the middle of a β-sheet bun that wraps around them (Fig. 1*b*). While both proteins form the same type of dimer, the variable extensions at the termini of these proteins result in considerable structural differences between them, as manifested by a dimeric C^{α} r.m.s.d. value of 2.4 Å. The most notable structural difference, apart from the terminal extensions themselves, is the displacement of the C-terminal end of the 'hotdog' α-helix *B* of AtDHNAT1 by ~9 Å relative to its orientation in Slr0204 (Fig. 1*c*). This displacement is caused by the N-terminal α-helix/β-strand extension in AtDHNAT1, which would sterically conflict with the C-proximal end of this helix if it were in the same orientation as in Slr0204 (Fig. 1*c*).

3.2. Slr0204 and AtDHNAT1 form distinct homotetramers

The structurally similar dimers of Slr0204 and AtDHNAT1 further associate to form tetramers using entirely different interfaces (Figs. 2*a* and 2*b*). Slr0204 is a facial tetramer that features a dimer–dimer interface near the hotdog α-helices, which point inward towards the center of the tetramer (Fig. 2*a*). In contrast, AtDHNAT1 is a back-to-back tetramer in which dimer–dimer association is mediated by association across the β-sheet on the other side of the protein, pointing the hotdog helices outwards (Fig. 2*b*). In both cases, the tetramer interface is solvated and largely polar in character. The crystallographic tetramers for both proteins are the only probable assemblies identified by *PISA* (Krissinel & Henrick, 2007), burying a total of 7180 $Å^2$ of surface area for Slr0204 and 7440 $Å^2$ of surface area for AtDHNAT1. The solution oligomerization states of both proteins were determined using sedimentation-equilibrium ultracentrifugation (Supplementary Fig. S1[1]), which is insensitive to the frictional coefficient and thus provides the molecular mass of the particle without other hydrodynamic complications. For Slr0204, the best-fit molecular mass was 77.4 kDa, which is in fair agreement with the mass of 63.0 kDa expected for the tetramer. For AtDHNAT1, sedimentation-equilibrium centrifugation gives a mass of 62.5 kDa, which agrees reasonably well with the calculated mass of 68.0 kDa for the tetramer. In both cases, the single tetrameric species model fit the data well, with no systematic trends in the residuals (Supplementary Fig. S1), indicating that both proteins exist predominantly as tetramers in solution.

The aforementioned terminal extensions of Slr0204 and AtDHNAT1 select against the alternative tetramerization mode by steric occlusion. In AtDHNAT1, the helix–strand N-terminal extension prevents facial tetramerization of the type observed in Slr0204, as these residues would be placed into conflict with their symmetry mates in the facial tetramer. Likewise, the C-terminal extension of Slr0204 would conflict with its symmetry mates if placed in an AtDHNAT1-type back-to-back tetramer, as well as generating additional clashes with residues 85–87. Therefore, steric conflicts in both Slr0204 and AtDHNAT1 play an important role in selecting against the alternative choice of tetramerization interface, although stabilizing contacts across these interfaces are also present in each protein.

3.3. Similarity to the 4-hydroxybenzoyl-CoA thioesterases

The details of the dimeric structures and the distinct modes of tetramerization observed in the DHNATs are strongly reminiscent of the 4-HBTs from *Arthrobacter* (Thoden *et al.*, 2003) and *Pseudomonas* (Benning *et al.*, 1998; Thoden *et al.*, 2002). Like the DHNATs, the 4-HBTs are enzymes from distantly related organisms that catalyze the hydrolysis of the same CoA thioester substrate using structurally related but oligomerically distinct hotdog-fold assemblies. At the structural level, Slr0204 is similar to the *Pseudomonas* 4-HBT (PDB entry 1lo9; Thoden *et al.*, 2002), with a C^{α} r.m.s.d. of 1.9 Å for the monomer and 2.3 Å for the tetramer, while AtDHNAT1 and the *Arthrobacter* 4-HBT (PDB entry 1q4u; Thoden *et al.*, 2003) are similar, with a C^{α} r.m.s.d. of 1.2 Å for the monomer and 1.5 Å for the tetramer. The superposition of these proteins (Figs. 3*a* and 3*b*) illustrates that *Pseudomonas* 4-HBT and the cyanobacterial DHNAT have structurally similar core regions although somewhat divergent external loop conformations (Fig. 3*a*), while *Arabidopsis* DHNAT1 and *Arthrobacter* 4-HBT share quite similar backbone structures (Fig. 3*b*). Importantly, the divergent modes of tetramerization in the *Pseudomonas* and *Arthrobacter* 4-HBTs are clearly conserved in Slr0204 and AtDHNAT1 (Figs. 3*a* and 3*b*) despite the modest 30% sequence identity between AtDHNAT1 and *Arthrobacter* 4-HBT and the lower 20% sequence identity between Slr0204 and *Pseudomonas* 4-HBT.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: DW5053). Services for accessing this material are described at the back of the journal.

The structural similarities between the 4-HBTs and the DHNATs are consistent with a prior secondary-structure-based phylogenetic alignment of these sequences with other thioesterases (Widhalm *et al.*, 2012) using the classification scheme proposed by Reilly and coworkers (Cantu *et al.*, 2010) and curated at the ThYme database (Cantu *et al.*, 2011). In the ThYme classification scheme, Slr0204 is a member of the TE12 clade (1,4-dihydroxy-2-naphthoyl-CoA hydrolases), while AtDHNAT1 is a member of the TE11 clade (4-HBT-II/EntH) (Widhalm *et al.*, 2012). These assignments were confirmed by *BLAST* searches of the Slr0204 and AtDHNAT1 amino-acid sequences against the downloaded ThYme database, which

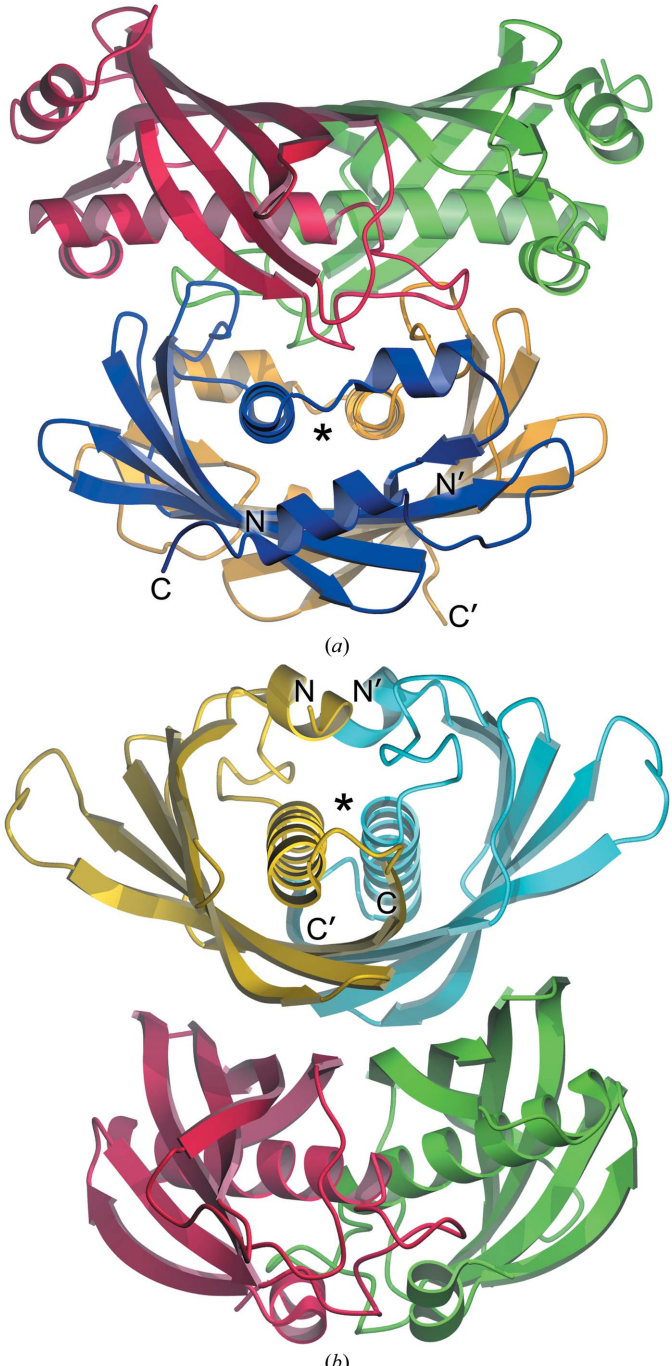

Figure 2
Slr0204 and AtDHNAT1 form distinct tetramers. In (*a*), the Slr0204 tetramer is shown with the orange/blue dimer oriented such that the hotdog α-helices point out of the plane of the page (marked by an asterisk). Slr0204 tetramerizes in the 'facial' mode, with these α-helices pointing inwards and the β-sheets oriented outwards. In (*b*), the AtDHNAT1 tetramer is shown with the yellow/blue dimer oriented in approximately the same way as the Slr0204 dimer in (*a*). AtDHNAT1 forms a 'back-to-back' tetramer, with the hotdog α-helices on the exterior of the multimer and the β-sheets mediating the dimer–dimer contacts in the interior of the tetramer.

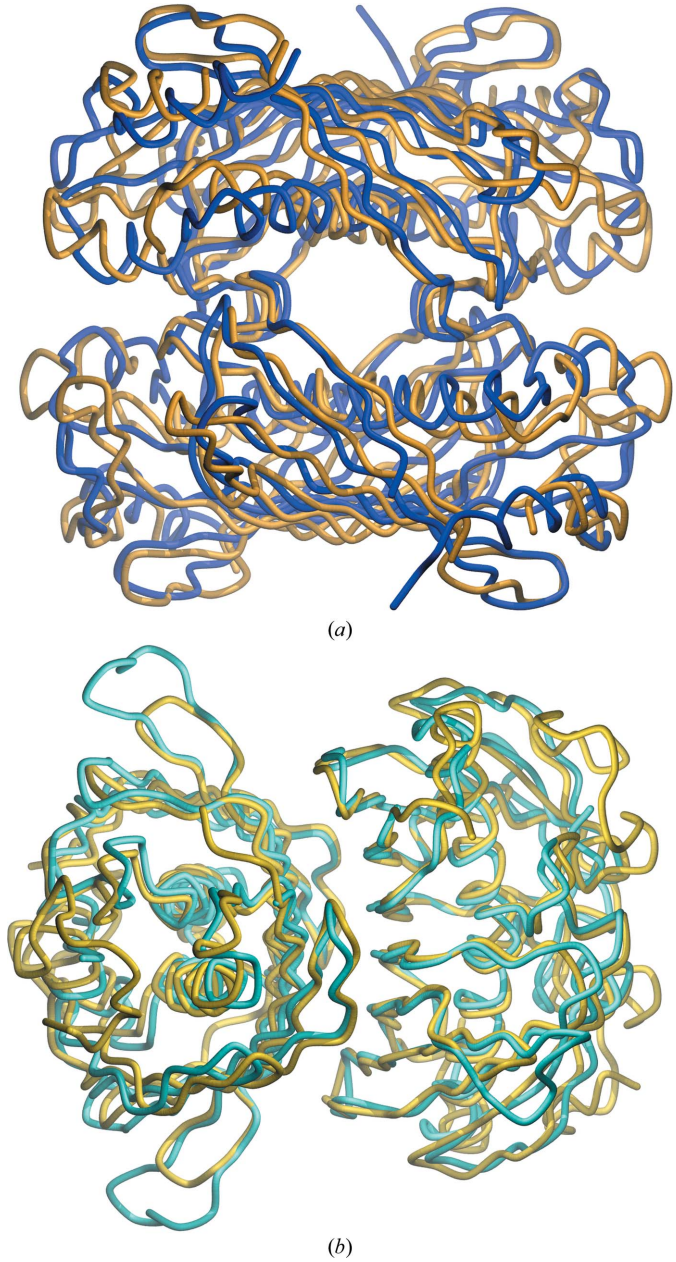

Figure 3
Slr0204 and AtDHNAT1 are structurally similar to the *Pseudomonas* and *Arthrobacter* 4-HBTs. The tetramers of Slr0204 (blue) and *Pseudomonas* 4-HBT (orange; PDB entry 1lo9) are superimposed in (*a*), showing the overall conservation of the facial tetramerization mode and the placement of secondary-structural elements. Despite their overall similarity, these proteins diverge in the conformations of the loops on the surface of the tetramers. (*b*) shows the superposition of AtDHNAT1 (cyan) and *Arthrobacter* 4-HBT (yellow; PDB entry 1q4u), illustrating the high degree of structural similarity between these two proteins.

exclusively identify curated sequences in the assigned clades among the top 100 hits for each protein. Considered in light of the structural similarities reported in this work, the inclusion of AtDHNAT1 in the TE11 clade is consistent with its clear structural similarity to *Arthrobacter* 4-HBT, which is a charter member of the TE11 group. In contrast, Slr0204 is not a member of the TE10 clade that contains the *Pseudomonas* 4-HBT, consistent with their lower overall structural similarity. The only member of the TE12 clade for which a structure has previously been determined is the *P. marinus* strain MIT 9313 enzyme (PDB entry 2hx5), which was used for the molecular-replacement phasing of Slr0204 (see §2). No functional characterization of the *Prochlorococcus* enzyme has been reported, but it superimposes with a monomeric C^{α} r.m.s.d. value of 1.1 Å with Slr0204, confirming that TE12 is the best choice of clade for Slr0204 and also indicating likely DHNA-CoA thioesterase activity for the *Prochlorococcus* enzyme. Additionally, the structural similarities between Slr0204 and *Pseudomonas* 4-HBT reported here suggest that structures of enzymes in clades TE10 and TE12 are all likely to be similar, which will require the determination of additional TE12 structures to properly test.

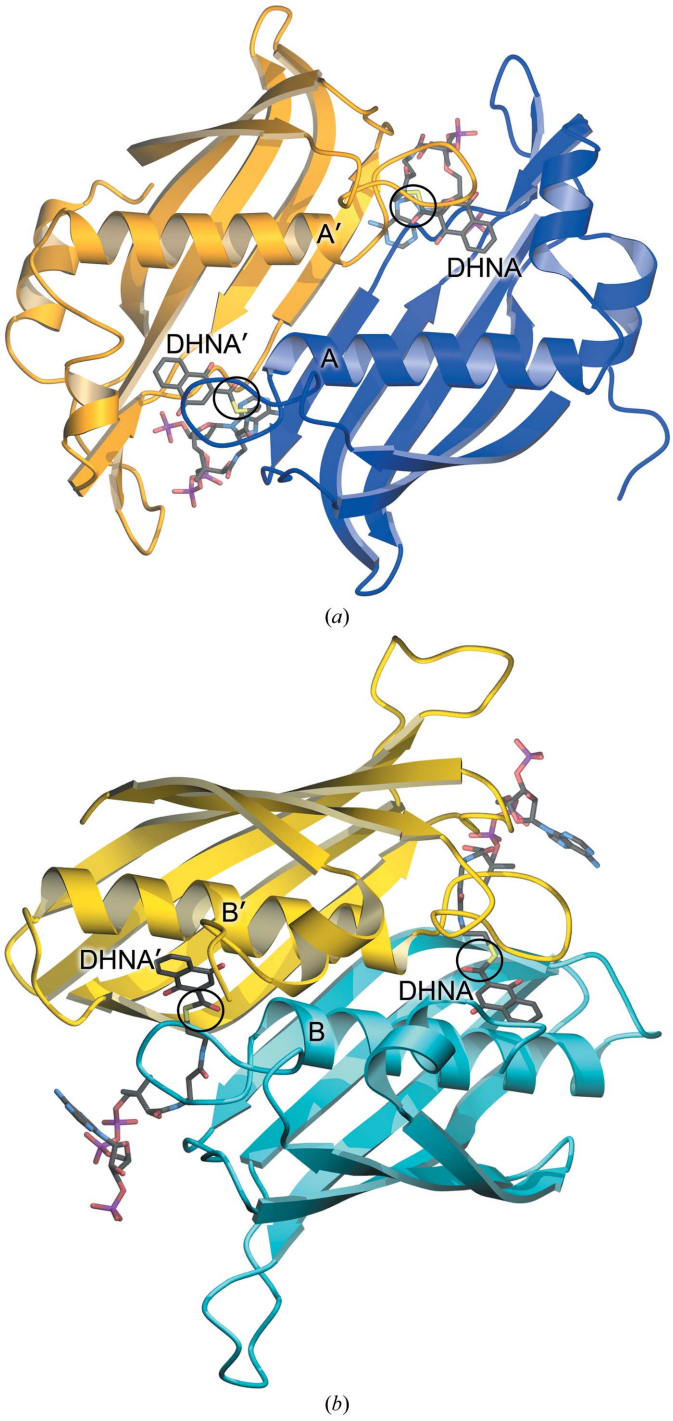

Figure 4
Orientation of the modeled DHNA-CoA in Slr0204 and AtDHNAT1. The DHNA-CoA thioesterase dimers are shown with modeled DHNA-CoA. In both proteins, the labile thioester bond is indicated by a black circle. Both Slr0204 (*a*) (blue and orange) and AtDHNAT1 (*b*) (cyan and yellow) position the thioester linkage near the N-termini of the hotdog α-helices. Although the relative locations of the DHNA moiety in these proteins are conserved, the bound orientations of the remainder of the CoA moiety differ.

3.4. Modeling of bound DHNA-CoA in plant and cyanobacterial thioesterases

Owing to the instability of DHNA-CoA, repeated attempts to obtain crystal structures of this substrate bound to active-site mutants of these proteins (D16N Slr0204 and E57Q AtDHNAT1) either by soaking in the substrate or cocrystallizing the complexes were unsuccessful. These mutations were chosen because they are predicted to greatly reduce the catalytic rate of these thioesterases based on observations made for the 4-HBTs (Song *et al.*, 2012; Zhuang *et al.*, 2012), thereby increasing the likelihood of crystallizing a stable Michaelis complex. In order to model the DHNA-Co-bound Michaelis complexes, substrate-bound or inhibitor-bound complexes of the *Pseudomonas* and *Arthrobacter* 4-HBTs were used as starting points for computational modeling of Slr0204 and AtDHNAT1 bound to DHNA-CoA (see §2). The resulting computational models are thus derived from previous experimentally determined complexes of related proteins, increasing the likelihood that they represent plausible binding poses for the substrate. It must be borne in mind, however, that these models are not directly fitted to experimental data and thus could deviate from the actual structures of the substrate complexes. Protein atoms were initially harmonically restrained to their positions in the crystal structures, while the DHNA-CoA ligand was subjected to energy minimization using *AMBER*11 (Case *et al.*, 2010). Ligand energy minimization was followed by removal of the harmonic restraints on protein atoms and further minimization of the entire system, thereby allowing both protein and ligand atoms to move without explicit restraint of the crystal structures. As a consequence, the computationally modeled complex differs from the crystal structure in some regions, particularly side chains in the active site that make direct contact with the substrate. The movement of these side chains in the active sites of both proteins is required to alleviate steric conflicts between the DHNA moiety and these residues in the crystal structure and therefore reflects changes that must

occur during ligand binding. No large changes were observed in either case, however, as the C^{α} r.m.s.d. values between the crystal structure and the substrate model are 0.6 Å for AtDHNAT1 and 0.3 Å for Slr0204. PDB files for the modeled structures of AtDHNAT1 and Slr0204 bound to DHNA-CoA are provided as Supplementary Material and the predicted contributions of each residue to the van der Waals and electrostatic interaction energies with modeled substrate are tabulated in the Supplementary Material.

As expected based on prior work (Thoden *et al.*, 2002, 2003), the binding site for DHNA-CoA spans the dimer interface in both DHNATs (Fig. 4). The modeled DHNA-CoA

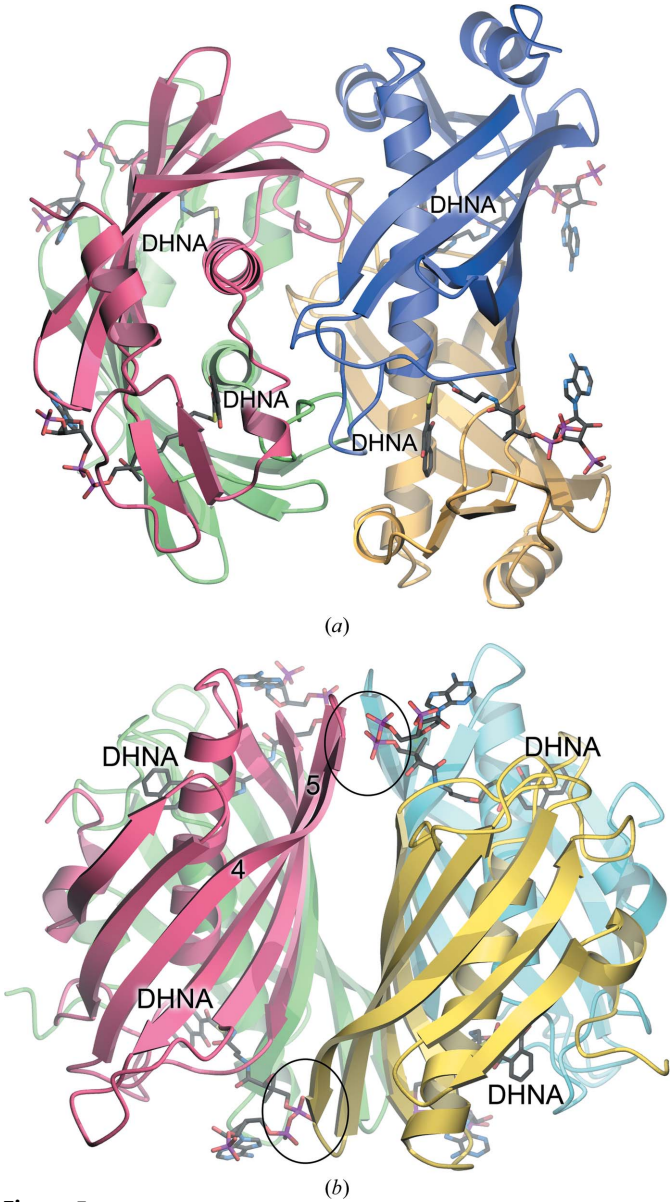

Figure 5
DHNA-CoA mediates dimer–dimer contacts in AtDHNAT1 but not in Slr0204. (*a*) shows that the Slr0204 facial tetramer directs the CoA portion of bound DHNA-CoA away from tetramerization interface, while in (*b*) the back-to-back AtDHNAT1 tetramer places substrate atoms at the dimer–dimer interface. Predicted contacts are made between the α-phosphate and β-phosphate of CoA and the loop between β-strands 4 and 5 (circled).

structures place the labile thioester bond near the N-termini of the hotdog α-helices, which would orient the positive end of a proposed helical macrodipole moment toward the thioester O atom. Because this O atom accumulates negative charge upon the formation of a tetrahedral intermediate during thioester cleavage, macrodipolar electrostatic stabilization has been previously suggested to potentially facilitate catalysis (Thoden *et al.*, 2002; Zhuang *et al.*, 2002). However, the magnitude of the energetic contribution made by helix macrodipole moments remains a subject of active debate (Roos *et al.*, 2013; Sengupta *et al.*, 2005; Wada, 1976), and hydrogen bonding has been suggested to be the dominant contributor to anion stabilization near the N-termini of α-helices in some systems (Roos *et al.*, 2013).

Despite similar modeled dimeric structures of the bound substrate in Slr0204 and AtDHNAT1, the different tetramers formed by these two proteins results in correspondingly distinct predicted interactions between the oligomers with DHNA-CoA. In particular, the modeled DHNA-CoA–Slr0204 complex is not predicted to contain any substrate-mediated interactions that span the tetramer interface (Fig. 5*a*), while the AtDHNAT1 complex does (Fig. 5*b*). In AtDHNAT1, the α-phosphate and β-phosphate of the CoA moiety are likely to accept hydrogen bonds donated by the backbone amide H atoms of residues 103–105 in the computational model. These residues are located in the loop between β-strands 4 and 5 of the adjacent monomer, thus creating four symmetry-related intermonomer contacts mediated by the bound substrate (Fig. 5*b*).

3.5. Active site of Slr0204

Although they catalyze the same reaction on the same substrate, AtDHNAT1 and Slr0204 have different active-site residues (Benning *et al.*, 1998; Thoden *et al.*, 2002, 2003). As with the tetramerization interfaces, the active sites of the DHNA-CoA thioesterases show a strong similarity to the distinct clades of 4-HBTs. The active sites of Slr0204 and *Pseudomonas* 4-HBT both feature a conserved aspartic acid that has been established as the catalytic nucleophile in *Pseudomonas* 4-HBT (Zhuang *et al.*, 2012). In other more distantly related thioesterases, acidic active-site residues have been proposed to act as general bases that activate water for direct attack at the thioester (Cao *et al.*, 2009). Therefore, we cannot state with certainty whether Asp16 acts as a general base or a catalytic nucleophile in Slr0204 as either would be consistent with the determined structure. In Slr0204, this residue (Asp16) is located on a loop at the bottom of the DHNA-binding pocket and is in well defined $2mF_o - DF_c$ electron density (Fig. 6*a*). The active-site region is defined by several hydrophobic residues (Fig. 6*a*), consistent with the largely hydrophobic character of the 1,4-dihydroxy 2-naphthoyl moiety of the substrate that this region binds. Despite the hydrophobic nature of the pocket, the electron-density maps for free Slr0204 show that this pocket is also rich in ordered water molecules that must be displaced upon substrate binding (Fig. 6*a*), potentially making a favorable

solvent entropic contribution that would partially offset the loss of substrate entropy upon binding.

The computational model of the bound substrate places the carboxylate side chain of the conserved Asp16 residue near the thioester C atom, consistent with the location and established essential role for this residue in catalysis by related thioesterases (Fig. 6*b*; Zhuang *et al.*, 2012). Pro57 is in the middle of a hydrophobic stretch of residues (55–59) that are positioned above the DHNA moiety and pack against the bicyclic aromatic ring system of the substrate (Fig. 6*b*). This hydrophobic cavity is narrow and closely conforms to the planar ring structure of the DHNA moiety in the model (Fig. 6*b*). A computational decomposition of the various energetic contributions to the predicted DHNA-CoA binding mode suggests that these hydrophobic residues would make multiple favorable van der Waals interactions with substrate, likely contributing to the high degree of selectivity of Slr0204 for DHNA-CoA (see Supplementary Material). The thioester

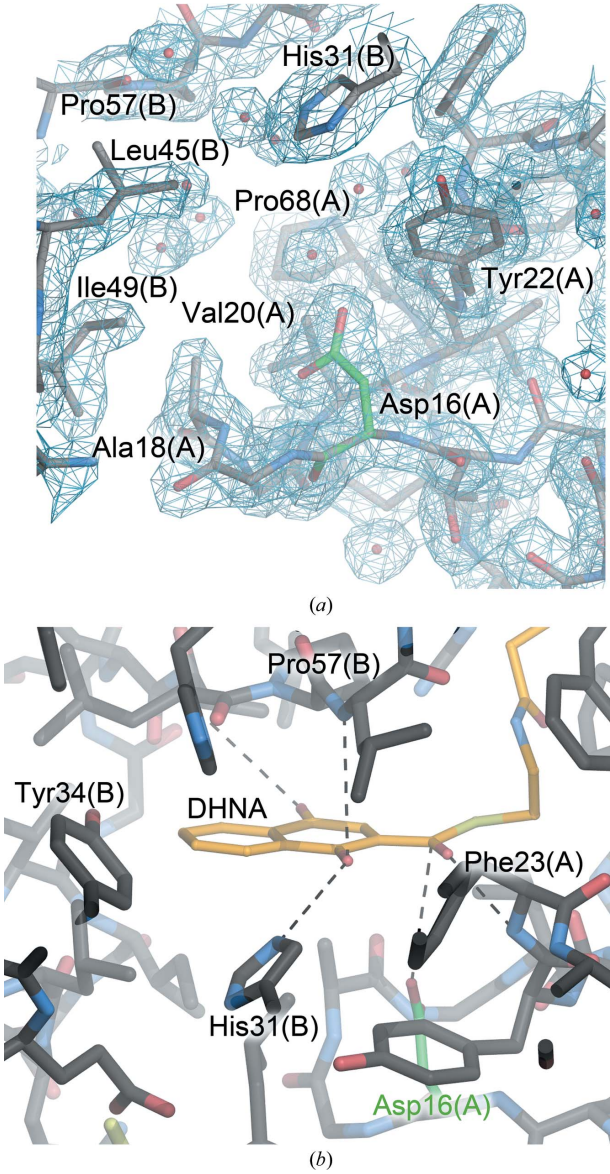

Figure 6
The active-site region of Slr0204. (*a*) shows the region around the active-site Asp16 residue (green), with $2mF_o - DF_c$ electron density contoured at 1.0σ shown in blue. Ordered water molecules are depicted as red spheres and selected residues are labeled with the monomer chain indicated in parentheses. The pocket is dominated by hydrophobic residues and contains substantial ordered solvent that must be displaced during substrate binding. The computationally modeled structure of the bound substrate is shown in (*b*), with candidate contacts indicated by dashed lines.

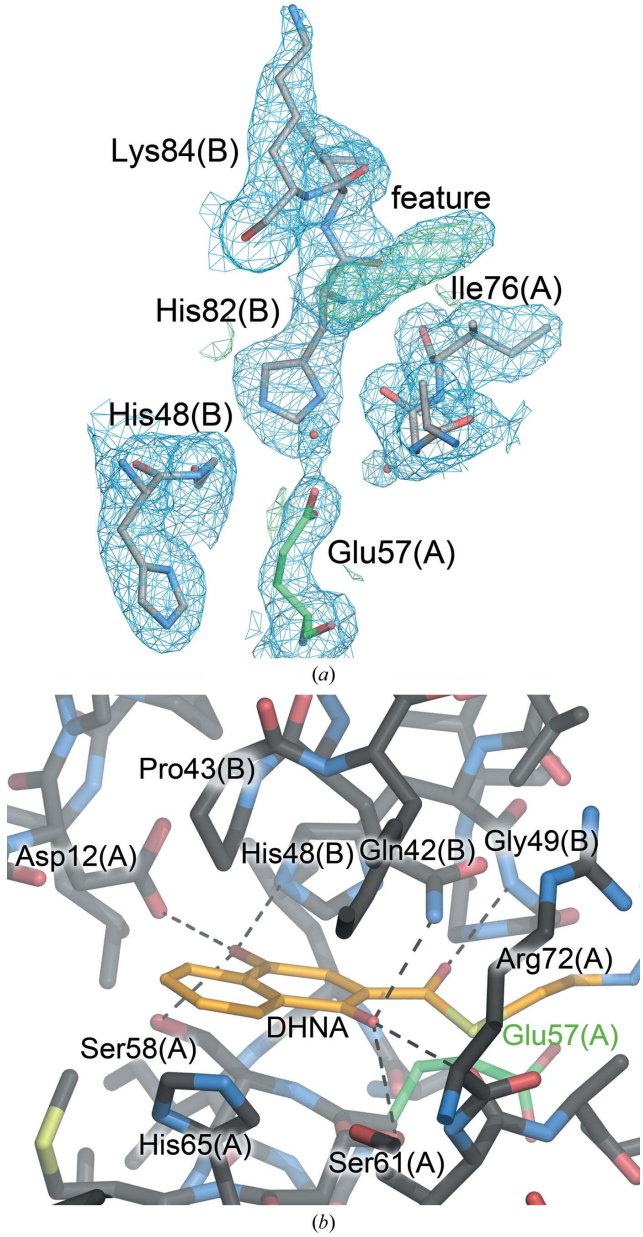

Figure 7
The active-site region of AtDHNAT1. (*a*) shows $2mF_o - DF_c$ electron density contoured at 1.0σ (blue) in a region near the predicted DHNA-CoA binding site. A prominent feature in $mF_o - DF_c$ difference electron density contoured at 3.5σ (green) was not modeled, but coincides with the modeled location of the pantothenyl moiety of CoA in the substrate-bound complex. In (*b*), predicted contacts between the modeled substrate and the protein are indicated by dashed lines. Compared with Slr0204, AtDHNAT1 makes more hydrogen bonds to the exocyclic O atoms of the DHNA moiety and possesses a more polar DHNA-binding cavity.

O atom makes a potential hydrogen bond to the amide H atom of Phe23 at the N-terminal end of the hotdog helix. This hydrogen bond would be expected to stabilize the accumulation of negative charge on this atom during the formation of a tetrahedral intermediate and thus may play an important role in catalysis (Cao *et al.*, 2009; Thoden *et al.*, 2002). This role is similar to that proposed for the macrodipole moment of the helix containing Phe23 (discussed above), but hydrogen bonding has been proposed to be more important than macrodipolar electrostatic effects in several systems (Roos *et al.*, 2013). Candidate contacts between the protein and one of the hydroxyl groups of DHNA involve the imidazole side chain of His31 and the amide H atom of Leu58 (Fig. 6*b*). The other quinol O atom (O4) is directed towards an opening in the binding pocket and thus is partially solvent-accessible. While we note that this model of the Slr0204 Michaelis complex has some intrinsic uncertainty that is attendant on all such computational models, it was calculated based on experimentally determined starting crystal structures and conservative energetic optimization.

3.6. Active site of AtDHNAT1

AtDHNAT1 and *Arthrobacter* 4-HBA-CoA thioesterase have similar active sites centered on a conserved glutamic acid (Glu57 in AtDHNAT1) that is essential for *Arthrobacter* 4-HBA-CoA thioesterase activity (Song *et al.*, 2012; Thoden *et al.*, 2003). The $2mF_o - DF_c$ electron-density map for AtDHNAT1 has an elongated feature that is also present as prominent (5–6σ) $mF_o - DF_c$ difference electron density in all four monomers in the asymmetric unit (Fig. 7*a*). This electron density was ambiguous and was therefore not modeled; however, we note that it overlaps closely with the predicted location of the pantothenyl moiety of DHNA-CoA in the computational model of the substrate complex. Speculatively, it is possible that this feature corresponds to a partially ordered portion of the β-OG detergent that was required for crystal formation. This is reminiscent of the observation of bound Jeffamine in the CoA-binding pocket of the CalE7 hotdog thioesterase (Kotaka *et al.*, 2009).

Analogous to the Slr0204–DHNA-CoA complex model, the AtDHNAT1–DHNA-CoA model contains a potential hydrogen bond between the N-terminal amide (Gly49) of the hotdog helix and the thioester O atom (Fig. 7*b*), possibly stabilizing formation of the tetrahedral intermediate resulting from nucleophilic attack at the thioester. Additionally, a proline residue (Pro43) is positioned above the DHNA moiety in AtDHNAT1 (Fig. 7*b*), similar to the predicted interaction between Pro57 and the substrate in Slr0204 (Fig. 6*b*). Furthermore, the active-site glutamic acid (Glu57) is not modeled in an optimal conformation for attack on the substrate thioester bond, reminiscent of the unproductive conformation observed for Asp16 in Slr0204. Unlike Slr0204, both exocyclic O atoms of the DHNA moiety are predicted to participate in multiple potential hydrogen-bonding interactions with surrounding residues in AtDHNAT1 (Fig. 7*b*), which is a consequence of a more polar active-site pocket in AtDHNAT1 than in Slr0204. Potential hydrogen bonding between active-site residues and both exocyclic O atoms of DHNA may contribute to the selectivity of AtDHNAT for DHNA-CoA as a substrate and contrasts with the sterically restricted and hydrophobic character of the Slr0204 active site. An additional difference between the AtDHNAT1 and Slr0204 active sites is the comparatively open acyl-binding cavity in AtDHNAT1, which directs the edge of the benzyl ring of DHNA out towards the solvent. This contrasts with the more enclosed DHNA-binding pocket in Slr0204 and suggests that AtDHNAT1 may be less selective for DHNA-CoA as substrate than is Slr0204.

4. Conclusions

This comparative structural analysis of DHNA-CoA thioesterases from *Synechocystis* and *Arabidopsis* provides a rationale for the high substrate selectivity of these enzymes (Widhalm *et al.*, 2009, 2012), which contrasts with the more permissive substrate profiles of some other hotdog-fold thioesterases (Cao *et al.*, 2009; Song *et al.*, 2012; Zhuang *et al.*, 2012). Based on computational modeling of the substrate-bound complexes, both enzymes possess binding pockets that accommodate the bulky planar DHNA moiety and contain residues that are predicted to hydrogen bond to one (Slr0204) or both (AtDHNAT1) exocyclic quinol O atoms of DHNA. The active-site pocket of Slr0204 tightly sandwiches the bicyclic aromatic ring system of DHNA and thus is likely to contribute to the high substrate selectivity of this enzyme. In contrast, the binding pocket of AtDHNAT1 is more polar and accessible to solvent at the C8 and C9 positions of DHNA than is the predominantly hydrophobic pocket of Slr0204. Therefore, AtDHNAT1 may be able to accommodate more polar substrates or substrates that bear larger acyl groups created by substitution of the DHNA moiety at these positions.

The crystal structures of Slr0204 and AtDHNAT1 demonstrate that these two distantly related proteins use distinct active-site architectures to catalyze the hydrolysis of the same DHNA-CoA thioester, recapitulating the divergent active sites and modes of tetramerization observed in the 4-hydroxybenzoyl CoA thioesterases of *Pseudomonas* and *Arthrobacter* (Benning *et al.*, 1998; Thoden *et al.*, 2002, 2003). Beyond the shared hotdog fold, the parallels between the DHNAT and 4-HBT enzymes are surprising and support a model whereby thioesterases containing either an active-site aspartate or glutamate residue diverged into distinct phylogenetic lineages prior to the evolution of strong substrate specificity in these enzymes, as Slr0204 and AtDHNAT1 are more similar to their corresponding 4-HBTs than they are to each other. This is supported by the unusually high degree of substrate specificity of the DHNAT enzymes (Widhalm *et al.*, 2009, 2012), which must have evolved after the key active-site residues in these distinct enzymes were already in place. Therefore, our structural results indicate that the distinct Asp/Glu active-site clades are an ancient bifurcation in the hotdog-fold thioesterases. Since plants appear to have obtained the AtDHNAT1-type gene from a horizontal gene-transfer event

from a species within the Lactobacillales order (Widhalm *et al.*, 2012), it should be further explored whether DHNATs from this order of bacteria also display strong substrate specificity. Very recently, a DHNAT activity from *E. coli* was assigned to YdiI, which appears to be a member of the *Arthrobacter* 4-HBT/*Arabidopsis* DHNAT1 family that employ a catalytic glutamate (Chen *et al.*, 2013). This enzyme is moderately selective for DHNA-CoA; however, it also has activity against related aromatic CoA thioesters such as salicylyl-CoA, 3,4-dihydroxybenzoyl-CoA and 3,5-dihydroxybenzoyl-CoA. This substrate profile is consistent with expectations based on the structure of the distantly related AtDHNAT1 enzyme, which has a more open and polar binding pocket that may accommodate a greater diversity of ligands than the more restricted Slr0204-type pocket. From a chemical standpoint, the diversity of active sites observed in the CoA thioesterases is likely to be a consequence of the comparative instability of the thioester linkage, the hydrolysis of which can be catalysed by multiple active-site architectures (Cantu *et al.*, 2010; Cao *et al.*, 2009; Kotaka *et al.*, 2009; Li *et al.*, 2000; Thoden *et al.*, 2003). Evolution has acted on this active-site diversity by generating (at least) two distinct classes of DHNA-CoA thioesterases, providing a new example of functional convergence in structurally distinct members in the hotdog-fold superfamily.

Portions of this work were funded by National Science Foundation Grant MCB-0918258 to GB, NIH grant R01 GM083669 to RCR and NIH grant R01 GM092999 to MAW. GM/CA at APS has been funded in whole or in part with Federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Sciences (Y1-GM-1104). Use of the BioCARS Sector 14 was supported by grants from the National Center for Research Resources (5P41RR007707) and the National Institute of General Medical Sciences (8P41GM103543) from the National Institutes of Health. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science under contract No. DE-AC02-06CH11357.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.

Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). *Bioinformatics*, **22**, 195–201.

Balius, T. E., Mukherjee, S. & Rizzo, R. C. (2011). *J. Comput. Chem.* **32**, 2273–2289.

Benning, M. M., Wesenberg, G., Liu, R., Taylor, K. L., Dunaway-Mariano, D. & Holden, H. M. (1998). *J. Biol. Chem.* **273**, 33572–33579.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Cantu, D. C., Chen, Y., Lemons, M. L. & Reilly, P. J. (2011). *Nucleic Acids Res.* **39**, D342–D346.

Cantu, D. C., Chen, Y. & Reilly, P. J. (2010). *Protein Sci.* **19**, 1281–1295.

Cao, J., Xu, H., Zhao, H., Gong, W. & Dunaway-Mariano, D. (2009). *Biochemistry*, **48**, 1293–1304.

Case, D. A. *et al.* (2010). *AMBER*11. University of California, San Francisco, USA.

Chang, K. H., Liang, P. H., Beck, W., Scholten, J. D. & Dunaway-Mariano, D. (1992). *Biochemistry*, **31**, 5605–5610.

Chen, M., Ma, X., Chen, X., Jiang, M., Song, H. & Guo, Z. (2013). *J. Bacteriol.* **195**, 2768–2775.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Collins, M. D. & Jones, D. (1981). *Microbiol. Rev.* **45**, 316–354.

Diederichs, K. & Karplus, P. A. (2012). *Science*, **336**, 1030–1033.

Dillon, S. C. & Bateman, A. (2004). *BMC Bioinformatics*, **5**, 109.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Fenn, T. D., Ringe, D. & Petsko, G. A. (2003). *J. Appl. Cryst.* **36**, 944–947.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.

Holyoak, T., Fenn, T. D., Wilson, M. A., Moulin, A. G., Ringe, D. & Petsko, G. A. (2003). *Acta Cryst.* D**59**, 2356–2358.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. (2006). *Proteins*, **65**, 712–725.

Jakalian, A., Bush, B. L., Jack, D. B. & Bayly, C. I. (2000). *J. Comput. Chem.* **21**, 132–146.

Jakalian, A., Jack, D. B. & Bayly, C. I. (2002). *J. Comput. Chem.* **23**, 1623–1641.

Kotaka, M., Kong, R., Qureshi, I., Ho, Q. S., Sun, H., Liew, C. W., Goh, L. P., Cheung, P., Mu, Y., Lescar, J. & Liang, Z.-X. (2009). *J. Biol. Chem.* **284**, 15739–15749.

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Lakshminarasimhan, M., Madzelan, P., Nan, R., Milkovic, N. M. & Wilson, M. A. (2010). *J. Biol. Chem.* **285**, 29651–29661.

Lang, P. T., Brozell, S. R., Mukherjee, S., Pettersen, E. F., Meng, E. C., Thomas, V., Rizzo, R. C., Case, D. A., James, T. L. & Kuntz, I. D. (2009). *RNA*, **15**, 1219–1230.

Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M. & Smith, J. L. (1996). *Structure*, **4**, 253–264.

Li, J., Derewenda, U., Dauter, Z., Smith, S. & Derewenda, Z. S. (2000). *Nature Struct. Biol.* **7**, 555–559.

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

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Oostende, C. van, Widhalm, J. R., Furt, F., Ducluzeau, A.-L. & Basset, G. J. (2011). *Adv. Bot. Res.* **59**, 229–261.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Roos, G., Foloppe, N. & Messens, J. (2013). *Antioxid. Redox Signal.* **18**, 94–127.

Sakuragi, Y. & Bryant, D. A. (2006). *Photosystem I: The Light-Driven Plastocyanin:Ferredoxin Oxidoreductase*, edited by J. H. Golbeck, pp. 205–222. Dordrecht: Springer.

Sengupta, D., Behera, R. N., Smith, J. C. & Ullmann, G. M. (2005). *Structure*, **13**, 849–855.

Song, F., Thoden, J. B., Zhuang, Z., Latham, J., Trujillo, M., Holden, H. M. & Dunaway-Mariano, D. (2012). *Biochemistry*, **51**, 7000–7016.

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.

Thoden, J. B., Holden, H. M., Zhuang, Z. & Dunaway-Mariano, D. (2002). *J. Biol. Chem.* **277**, 27468–27476.

Thoden, J. B., Zhuang, Z., Dunaway-Mariano, D. & Holden, H. M. (2003). *J. Biol. Chem.* **278**, 43709–43716.

Wada, A. (1976). *Adv. Biophys.* **9**, 1–63.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. (2004). *J. Comput. Chem.* **25**, 1157–1174.
Widhalm, J. R., Ducluzeau, A. L., Buller, N. E., Elowsky, C. G., Olsen, L. J. & Basset, G. J. (2012). *Plant J.* **71**, 205–215.
Widhalm, J. R., van Oostende, C., Furt, F. & Basset, G. J. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 5599–5603.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.
Zhuang, Z., Latham, J., Song, F., Zhang, W., Trujillo, M. & Dunaway-Mariano, D. (2012). *Biochemistry*, **51**, 786–794.
Zhuang, Z., Song, F., Zhang, W., Taylor, K., Archambault, A., Dunaway-Mariano, D., Dong, J. & Carey, P. R. (2002). *Biochemistry*, **41**, 11152–11160.