Crystal Structure of 4-Chlorobenzoate:CoA Ligase/Synthetase in the Unliganded and Aryl Substrate-Bound States[†,‡]

Andrew M. Gulick,*[,§,‖] Xuefeng Lu,[⊥] and Debra Dunaway-Mariano[⊥]

Hautpman-Woodward Medical Research Institute, State University of New York at Buffalo, Buffalo, New York 14203,
Department of Structural Biology, State University of New York at Buffalo, Buffalo, New York 14203, and
Department of Chemistry, University of New Mexico, Albuquerque, New Mexico 87131

Received March 30, 2004; Revised Manuscript Received May 6, 2004

ABSTRACT:　4-Chlorobenzoate:CoA ligase (CBAL) is a member of a family of adenylate-forming enzymes that catalyze two-step adenylation and thioester-forming reactions. In previous studies, we have provided structural evidence that members of this enzyme family (exemplified by acetyl-CoA synthetase) use a large domain rotation to catalyze the respective partial reactions [A. M. Gulick, V. J. Starai, A. R. Horswill, K. M. Homick, and J. C. Escalante-Semerena, (2003) *Biochemistry 42*, 2866−2873]. CBAL catalyzes the synthesis of 4-chlorobenzoyl-CoA, the first step in the 4-chlorobenzoate degradation pathway in PCB-degrading bacteria. We have solved the 2.0 Å crystal structure of the CBAL enzyme from *Alcaligenes sp. AL3007* using multiwavelength anomalous dispersion. The results demonstrate that in the absence of any ligands, or bound to the aryl substrate 4-chlorobenzoate, the enzyme adopts the conformation poised for catalysis of the adenylate-forming half-reaction. We hypothesize that coenzyme A binding is required for stabilization of the alternate conformation, which catalyzes the 4-CBA-CoA thioester-forming reaction. We have also determined the structure of the enzyme bound to the aryl substrate 4-chlorobenzoate. The aryl binding pocket is composed of Phe184, His207, Val208, Val209, Phe249, Ala280, Ile303, Gly305, Met310, and Asn311. The structure of the 4-chlorobenzoate binding site is discussed in the context of the binding sites of other family members to gain insight into substrate specificity and evolution of new function.

In this paper we report the three-dimensional structure of the bacterial enzyme 4-chlorobenzoyl-CoA ligase (CBAL,[1] EC 6.2.1.33), both unliganded and bound to its substrate 4-chlorobenzoate (4CBA). 4CBA is an environmental pollutant derived from the breakdown of man-made polychlorinated biphenyls by soil dwelling microbes. In *Alcaligenes sp. AL3007* and in other opportunistic 4CBA degrading bacteria, CBAL catalyzes the first step of the 4CBA dehalogenation pathway (*1*). It has been speculated that

Scheme 1
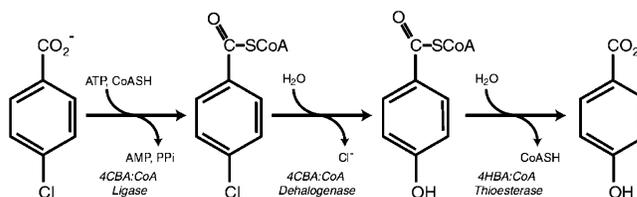
CBAL, as well as the pathway dehalogenase and thioesterase (Scheme 1), are products of adaptive evolution aimed at the utilization of environmental 4CBA as an alternate carbon source (*2*). In *Alcaligenes sp. AL3007*, the 4CBA pathway genes are carried on a large plasmid which also contains the genes encoding PCB breakdown (*3*).

Previous mechanistic studies (*4, 5*), which focused on the *Pseudomonas sp CBS3* CBAL, showed that catalysis occurs in two kinetically independent partial reactions utilizing a 4CBA-adenylate intermediate (Scheme 2). The first partial

* Corresponding author. Address: Hautpman-Woodward Medical Research Institute, Department of Structural Biology, State University of New York at Buffalo, 73 High St., Buffalo, NY 14203-1196. Phone: (716) 856-9600, ext. 327. Fax: (716) 852-6086. E-mail: gulick@hwi.buffalo.edu.
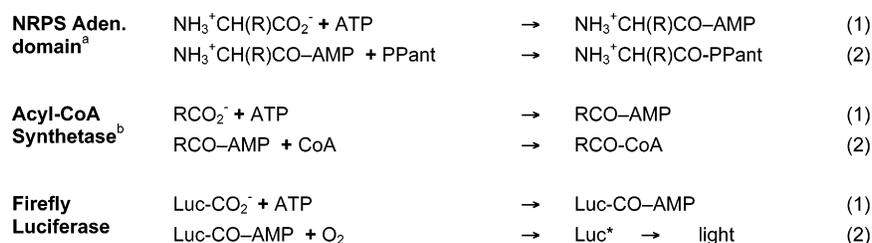
§ Hautpman-Woodward Medical Research Institute, State University of New York at Buffalo.

‖ Department of Structural Biology, State University of New York at Buffalo.

⊥ University of New Mexico.

Scheme 2: Two-Step Reactions Catalyzed by the Three Subfamilies of the Adenylate-Forming Family of Enzymes

NRPS Aden. domain[a]	$NH_3^+CH(R)CO_2^- + ATP$	$\rightarrow$	$NH_3^+CH(R)CO{-}AMP$ (1)
	$NH_3^+CH(R)CO{-}AMP + PPant$	$\rightarrow$	$NH_3^+CH(R)CO\text{-}PPant$ (2)
Acyl-CoA Synthetase[b]	$RCO_2^- + ATP$	$\rightarrow$	$RCO{-}AMP$ (1)
	$RCO{-}AMP + CoA$	$\rightarrow$	$RCO\text{-}CoA$ (2)
Firefly Luciferase	$Luc\text{-}CO_2^- + ATP$	$\rightarrow$	$Luc\text{-}CO{-}AMP$ (1)
	$Luc\text{-}CO{-}AMP + O_2$	$\rightarrow$	$Luc^* \rightarrow$ light (2)

[a] PPant represents the phosphopantetheine cofactor for the NRPS carrier protein domains which follow the adenylation domain. [b] $R = Cl{-}C6H4$ for CBAL.

reaction involves adenylation of 4CBA by MgATP ($k = 130$ s^{-1}) while the second is the nucleophilic attack of the CoA thiol on the mixed anhydride to form the CoA thioester ($k = 100$ s^{-1}).

CBAL is a member of the adenylate-forming enzyme superfamily (*6*). The enzymes are 500−700 residues in length and are usually soluble however membrane-bound proteins have also been identified that are involved in fatty-acid uptake and metabolism (*7, 8*). The adenylate-forming family contains three subfamilies, the acyl- and aryl-CoA synthetases or ligases, the adenylation domains of the nonribosomal peptide synthetases (NRPSs) (*9*), and firefly luciferase (*10*).

All adenylate-forming enzymes catalyze a two-step ping-pong reaction (*11, 12*) in which the first half-reaction is an adenylation of a carboxylic acid, creating an acyl- or aryl-adenylate associated with the release of pyrophosphate (Scheme 2). The acyl and aryl-CoA synthetases catalyze CoA thioester formation from adenylated short or long chain fatty acids or adenylated benzoic acid derivatives. The nonribosomal peptide synthetases (NRPSs) are composed of numerous catalytic domains linked in a modular fashion, often within a single polypeptide that can be thousands of residues in length (*9, 13, 14*). The adenylation domains of the NRPSs serve as the entry point for amino acid binding and therefore dictate most, though not all (*15*), of the sequence of the final peptide. During peptide synthesis, the nascent peptide remains covalently attached to the NRPS through the pantetheine cofactor of the adjacent carrier protein domains. Usually, the carrier protein domain is located immediately C-terminal to the adenylation domain in the multidomain NRPS protein however examples of intermolecular transfer to a carrier protein domain located on a separate polypeptide have also been reported (*16−18*). The third subfamily, containing firefly luciferase, catalyzes the oxidative decarboxylation of the adenylate in the second half-reaction.

The crystal structures of at least one representative from each subfamily have been reported. Specifically, the current library includes the structures of luciferase (*19*), the phenylalanine activating domain (PheA) from gramicidin synthetase S (*20*), the 2,3-dihydroxybenzoate activating domain (DhbE) from the *Bacillus subtilis* bacillibactin NRPS system (*21*), and the acetyl-CoA synthetases (Acs) from *Salmonella enterica* (*22*) and yeast (*23*). The enzymes are composed of two α,β-domains: the large N-terminal domain and the smaller C-terminal domain, which has been observed in two distinct conformations in different crystal structures (Figure 1). Analysis of the structure of acetyl-CoA synthetase (Acs) from *Salmonella enterica* bound to CoA and adenosine-5′-propyl phosphate, a mimic of the adenylate intermediate (*22*), illustrates that a rotation of the C-terminal domain through

an angle of ∼140° is needed to superimpose this domain on the C-terminal domains of the NRPS adenylation domains. We have proposed that conformation 1 (Figure 1A) catalyzes the adenylate-forming reaction while conformation 2 (Figure 1B) catalyzes the thioester-forming reaction. The kinetically independent catalytic sites implicated by the transient kinetic studies of the CBAL reaction (*4*) could therefore be rationalized not as fully unique active sites for the two half-reactions but, rather, as mediated by the opposing faces of the C-terminal domain. The adenylate-forming enzymes therefore appear to utilize domain alternation, in which an enzyme adopts two different conformations to catalyze individual steps of a multistep reaction (*24*), in the formation of 4CBA-CoA.

In this paper we present the structures of CBAL in the unliganded state and bound to 4CBA. The aryl substrate binding pocket is identified and analyzed in comparison to other aryl-CoA ligases (*25*). The CBAL structure will be used, together with structures of the dehalogenase and thioesterase (*26, 27*), in future studies to launch the redesign of the 4CBA binding site for an expanded range of substituted benzoates as a first step toward adaptation of the 4CBA pathway for general bioremediation applications.

METHODS

Materials. PEG4000, AMP, and 4-CBA were purchased from SIGMA. All other chemicals were of reagent grade.

Protein Purification. C-terminal His-tagged recombinant *Alcaligenes sp. Al3007* CBAL ($k_{cat} = 9$ s^{-1}, Mg·ATP $K_m = 0.12$ mM, CoA $K_m = 0.31$ mM, and 4CBA $K_m = 0.001$ mM at pH 7.5 and 25 °C) was prepared by IPTG-induced expression of the SphI-BglII-pQE-70-CBAL in transformed *Escherichia coli* JM109 cells followed by Ni−NTA Agarose column chromatography of the cell lysate as described in S. Y. Lai, X. Lu, W. Zhang, A. C. Layton, G. S. Sayler, and D. Dunaway-Mariano (manuscript in preparation). The protein was concentrated to ∼10 mg/mL in 20 mM K⁺Hepes (pH 7.5)/200 mM KCl/1 mM DTT and frozen by pipetting directly into liquid nitrogen (*28*). Aliquots of CBAL were thawed daily for crystallization experiments. The SeMet labeled enzyme was prepared by IPTG induction of transformed cells grown in 50 μg/mL ampicillin-containing minimal media, which was supplemented with L-lysine, L-phenylalanine, L-threonine at 100 mg/L each, and L-isoleucine, L-leucine, L-valine, and L-selenomethionine at 50 mg/L each (*29*). The purification procedure was modified to include 2 mM DTT in all purification buffers and 10 mM DTT in the storage buffer. The specific activity of the SeMet CBAL was the same as the native enzyme.
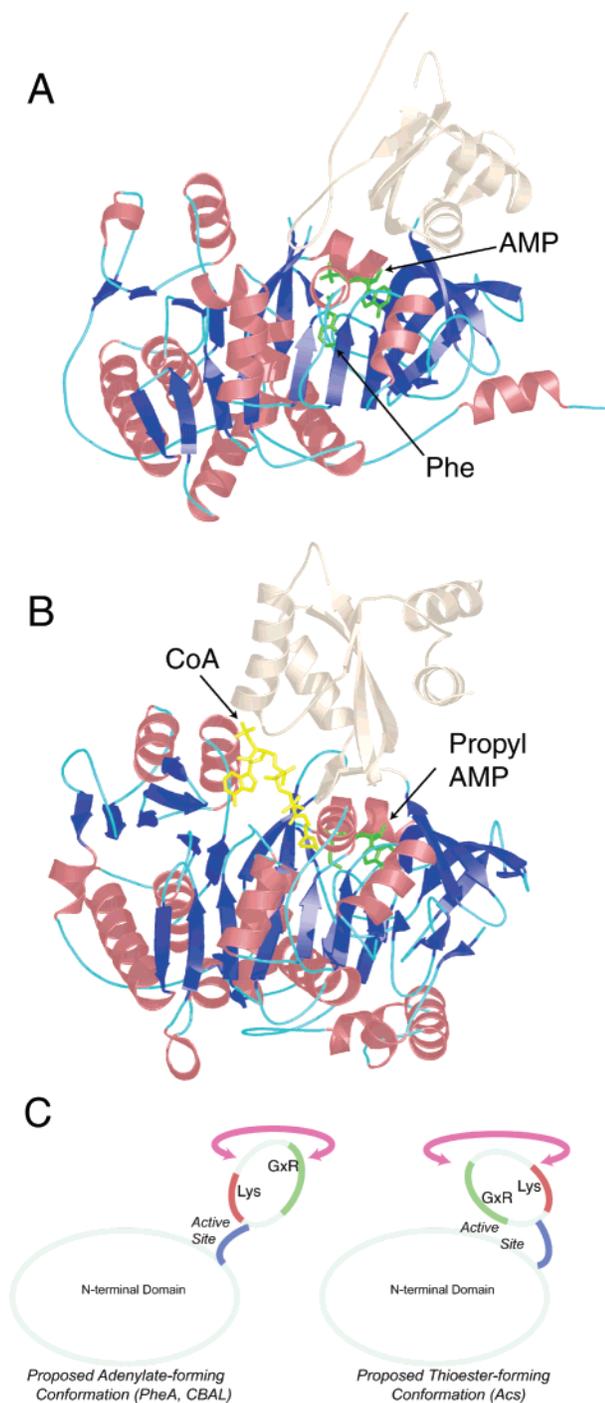
FIGURE 1: Distinct conformations in the adenylate forming family of enzymes. Ribbon diagrams of (A) PheA domain bound to AMP and phenylalanine, PDB entry 1AMU (*20*), and (B) *S. enterica* Acs bound to adenosine-5′-propyl phosphate and CoA, PDB entry 1PG4 (*22*). For both proteins, the larger N-terminal domains are shown with blue strands and red helices and the C-terminal domains are shown in tan. The active site contents are shown in ball-and-stick configuration. For the PheA structure, the AMP and Phe molecules mimic the adenylate at the end of the first half-reaction. The Acs structure, in contrast, contains the alkyl-adenylate, adenosine-5′-propyl phosphate, and the CoA and has been proposed to represent the structure of the enzyme at the start of the second half-reaction. The C-terminal domains contain the same secondary structural elements however the Acs structure contains an additional two helices. (C) A schematic diagram depicting the rotation of the C-terminal domain leading to the conformational differences observed in the bacterial Acs enzyme and the PheA and CBAL structures.

Crystallization of CBAL. The CBAL protein was screened for crystallization using sparse matrix screening (*30*, *31*) using 120 unique crystallization cocktails. Showers of small crystals were observed within 5 h of setup of the hanging drop experiment. The conditions were refined and crystals suitable for data collection were obtained from hanging drop experiments at 4 °C using a precipitant of 6−8% PEG 4000, 10% glycerol, 0.24−0.28 M $CaCl_2$, and 50 mM CHES (pH 9.0). Crystals mounted in a quartz capillary or cryocooled at −170 °C in a stream of N_2 gas after cryoprotection with increased glycerol concentration routinely diffracted to a maximum resolution of ∼3.5 Å. The space group was determined to be either $P3_121$ or $P3_221$ with $a = b = 129.3$ Å, $c = 71.5$ Å. Slight improvements were observed by soaking the crystals for 1−2 min in solutions containing a higher concentration of PEG4000. The final cryoprotection protocol involved transferring the crystals through five solutions containing increasing salt, glycerol, and PEG concentrations for one minute each. The final cryoprotectant contained 20% PEG4000, 25% glycerol, 0.3 M $CaCl_2$, and 50 mM CHES (pH 9.0). Through this protocol, the diffraction consistently improved to as high as 2.2 Å using in-house X-radiation; associated with the improvement in resolution was a shortening of all three cell lengths of the crystal lattice by ∼3%.

Structure Determination and Refinement. Attempts were made to solve the structure of CBAL using molecular replacement using a low-resolution data set. Search models included the full length and N-terminal domains of Acs and PheA. Despite high structural similarity between family members, no suitable solutions were obtained. The final structure of CBAL was determined through multiwavelength anomalous dispersion using SeMet labeled protein. MAD data were collected at beamline ×12C of the National Synchrotron Light Source at Brookhaven National Laboratory. Data were collected at three wavelengths using 0.5° rotations and 20 s exposures per frame. The detector was a Brandeis 2 × 2 CCD-based detector and the crystal to detector distance was 225 mm. Additional data collection statistics are presented in Table 1.

The heavy-atom substructure was solved using *BnP*, an automated interface for the determination of protein structure (*32*). All three wavelengths were used, and the default settings were applied. The *SnB* algorithm was able to locate all 10 of the Se sites. The positions of the ten Se atoms were found among the top 11 peaks identified by *SnB*. Because the true space group was unknown, the enantiomorph determination feature of *BnP* was utilized. Using unrefined protein phases, the ratio of the standard deviations of the electron density in the protein and solvent regions were 1.15 for $P3_121$ and 1.74 for $P3_221$, clearly identifying the latter as the true space group. The refined phases were used for automated chain tracing with RESOLVE (*33*). Approximately 300 residues (of 504) were positioned by RESOLVE; an additional 175 residues were built manually into Fourier maps created with the refined, solvent-flattened phases output from both RESOLVE and PHASES (*34*).

The traced model was refined initially with CNS (*35*) against a moderate resolution data set collected in-house using Cu/Kα X-radiation. The *R*-factor dropped from 39.2% to 28.3% while the Rfree went from 39.0% to 30.6%. To obtain an atomic model refined against the highest resolution

Table 1: Crystallographic Data for SeMet Data Sets

	SeMet peak	SeMet inflection	SeMet remote	SeMet #2	native + 4CBA
wavelength	0.97934 Å	0.97959 Å	0.95 Å	0.95 Å	1.54 Å
resolution	2.5 Å	2.5 Å	2.5 Å	2.0 Å	2.2 Å
unit cell ($a = b$)	124.7 Å	124.7 Å	124.7 Å	124.2 Å	124.7 Å
unit cell (c)	68.9 Å	68.9 Å	68.8 Å	69.0 Å	68.9 Å
R_{merge}[a,b]	7.3% (17.1%)	8.5% (27.4%)	9.2% (30.9%)	8.5% (51.6%)	6.6% (37.3%)
completeness[a]	99.9% (100%)	99.9% (100%)	99.8% (100%)	100% (99.7%)	96.1% (78.5%)
I/σ	15.1	11.5	10.2	9.1	13.6
no. of observations	93567	103670	112553	239238	74606
no. of reflections	21617	21616	21591	41581	30276

[a] Values for the highest resolution shell are given in parentheses. [b] $R_{merge} = [\sum(|I_{hi} - I_h|)/\sum I_{hi}]$, where I_{hi} and I_h are individual and mean intensities of all equivalent reflections, respectively.

Table 2: Refinement Statistics

	unliganded	4CBA-bound
resolution range[a]	25.0−2.0 Å	25.0−2.2 Å
R_{cryst} (overall/highest resolution shell)[b]	18.3% (23.2%)	17.8% (25.6%)
R_{free} (overall/highest resolution shell)[b]	20.6% (25.7%)	22.1% (33.4%)
Wilson B-factor	25.2 Å^2	40.8 Å^2
average B-factor, overall	16.0 Å^2	32.3 Å^2
average B-factor, protein (all, main chain, side chain)	14.6, 14.0, 15.2 Å^2	31.1, 30.8, 31.5 Å^2
average B-factor, solvent[c]	32.5 Å^2 (340)	45.6 Å^2 (325)
average B-Factor, ligands and metal ion[c]	31.5 Å^2 (1)	48.4 Å^2 (11)
RMS deviation bond lengths, angles	0.009 Å, 1.063°	0.011 Å, 1.206°

[a] The highest-resolution shell is from 2.05 to 2.00 Å for the unliganded protein and 2.26 to 2.20 Å for the protein bound to 4CBA. [b] R_{cryst} and $R_{free} = [\sum_h(|F_{obs} - F_{calc}|)/\sum_h F_{obs}]$, where F_{obs} and F_{calc} are observed and calculated structure factors. The R_{free} data set included 5% of the reflections. [c] Total number of atoms used in calculation in parentheses.

data available, we used the remote wavelength data set of a second crystal grown from SeMet labeled protein (SeMet #2). These data were also collected at beamline ×12C of Brookhaven National Laboratory on a crystal that diffracted to 2.0 Å.

An additional data set was collected using an in-house X-ray source and a crystal that was grown in the presence of 1 mM AMP and 1 mM 4CBA. The crystal was cryoprotected using the same protocol for the unliganded crystals with solutions containing 1 mM AMP and 4CBA. Data collection statistics are presented in Table 1. The structure was solved by difference Fourier methods using the native model from which all solvent molecules had been removed. Despite the presence of both ligands in all solutions, examination of the active site density revealed clear density for the 4CBA molecule only. Refinement of both models were performed with CNS (*35*) and REFMAC 5.0 (*36*). The refined model of the 4CBA molecule is in close agreement to the geometric parameters obtained from the crystal structures of 4CBA alone (*37, 38*). Final refinement statistics are presented in Table 2.

Because of the distinct conformational orientations of the C-terminal domain, we applied the TLS (translation, libration, screw-rotation displacements) feature of REFMAC in the final stages of refinement (*39*). The two domains, residues 1−401 and 402−504, were considered as distinct domains with TLS to model anisotropic displacements of each domain as a rigid body. Incorporation into refinement resulted in a ~1% decrease in both the crystallographic and free R-factors.

Computational Analysis. Structural alignments were performed with ALIGN (*40*) and LSQKAB (*41*). Figures were generated with PYMOL (*42*). Surface area calculations were performed with SURFACE (*43*), which uses the algorithm

of Lee and Richards (*44*), with a probe sphere of 1.4 Å and a *z* step of 0.25 Å.

Deposition of Atomic Coordinates. The final atomic coordinates and structure factors for both the unliganded SeMet-labeled and 4CBA-bound proteins have been deposited with the Protein Data Bank (accession codes 1T5H and 1T5D, respectively).

RESULTS

Structure Determination and Refinement. The structure of CBAL was determined by MAD phasing using SeMet-labeled protein. The protein crystallized in the trigonal space group $P3_x21$. Maximum resolution data were obtained by transferring the crystals incrementally to higher concentrations of salt, glycerol, and PEG4000, which decreased the lengths of the unit cell axes and improved the diffraction considerably. The heavy atom substructure was solved, and phases were refined, using *BnP* (*32*). The enantiomorph determination feature of *BnP* was able to identify the correct space group, $P3_221$. The atomic model was refined against data obtained from crystals of SeMet-labeled protein. The final model for the unliganded protein contains 498 residues; the model for the protein bound to 4CBA contains 502 residues. The crystallized protein also contains a C-terminal RSHHHHHH tag which was used for purification however the entire tag is disordered. Additionally, there are two disordered regions in the protein chain. One of these loops, residues 162−165, is part of a highly conserved TSG(S/T)-TGxPKG motif that is disordered in several of the previously determined structures of adenylate-forming enzyme. In Acs and DhbE, the two structures where this loop was ordered, it adopts different conformations; this region may interact with the β- and γ-phosphates of ATP in a manner similar to
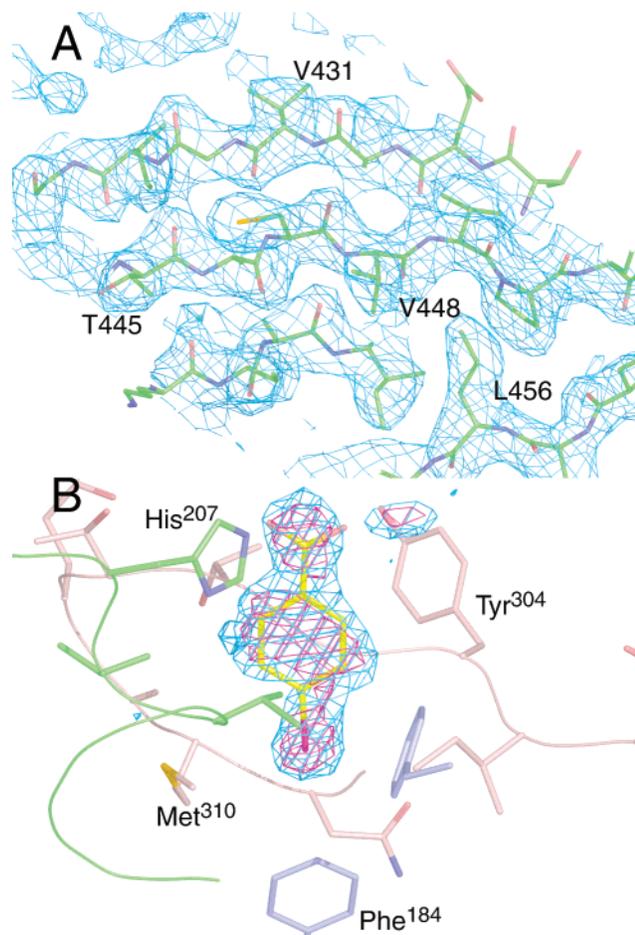
FIGURE 2: Representative electron density. (A) A region of the multiple anomalous dispersion experimental electron density map contoured at 1σ is shown for the main β-sheet present in the C-terminal domain. The map was generated using data from *BnP* (*32*) and RESOLVE (*61*). (B) An unbiased electron density map calculated with coefficients of the form *Fo-Fc* is shown contoured at 2.5σ (cyan) and 3.5σ (magenta). The map was generated prior to the addition of the atomic coordinates of the 4CBA ligand. Present near the carboxylate and the phenolic hydroxyl of Tyr304 is partial electron density for a water molecule.

FIGURE 3: Ribbon diagrams of CBAL. (A) A ribbon representation of the CBAL protein bound to 4CBA. The molecule of 4CBA is shown in green ball-and-stick representation. The three sheets of the N-terminal domain are shown in blue (sheet A), cyan (sheet B), and magenta (sheet C). Bound to 4CBA, the protein is in the conformation 1 seen previously in PheA, DhbE, and the yeast Acs1 enzyme. The Lys492 residue is directed into the active site, though the side chain is disordered. The two strands shown in yellow, and the loop connecting these strands contain residues that interact with the active site in the alternate conformation seen in the bacterial Acs structure (*22*). (B) A ribbon representation of the CBAL dimer illustrating that the N-terminal domain is used as the dimer interface. The figure is shown viewed down the crystallographic 2-fold. The upper subunit is rotated approximately 20° from the view in pane A.

traditional P-loops (*45*, *46*). Mutation of CBAL loop residues Gly163, Gly166, Pro168, and Lys169 results in impaired catalysis of the CBA adenylation partial reaction (*5*). A second loop, located at residues 110−111, is disordered in the unliganded structure however sufficient density was present to allow modeling of this region of the structure for the protein bound to 4CBA. In the liganded structure, the density is still weak and the overall B-factor for this region is higher (60.9 Å^2) than for the N-terminal domain average (38.9 Å^2). The homologous loop in Acs is considerably longer and contains an Arg residue (Arg191) that interacts with the phosphate of CoA in the structure of Acs (*22*). This loop of CBAL contains an Arg residue at position 111, which may serve a similar role. The unliganded protein contains 14 residues for which the side chains were disordered (Arg87, Val109, Gln112, Val113, Asp115, Phe118, Gln149, Ser162, Asn347, Asn491, Lys492, Arg496, Gln500, and Ser503). The protein bound to 4CBA contains seven residues with disordered side chains (Arg111, Thr165, Glu230, Asn491, Lys492, Arg496, and Gln500). These residues are all located on the surface of the protein and are included in the final model as alanine residues. Each model contained a Ca^{2+} ion

coordinated between Asp483 and Gln484 of one subunit and Asp472 of a crystallographically related molecule. The final refinement statistics are presented in Table 2. Representative experimental electron density is shown (Figure 2).

The CBAL protein is considerably smaller than other members of the adenylate-forming family. The N-terminal domain consists of residues 1−401, with Asp402 serving as the hinge for the proposed domain motion described above. The overall structure of the CBAL protein is similar to previously determined structures for adenylate-forming enzymes (Figure 3). The pairwise alignment of all six adenylate forming enzyme structures was determined using ALIGN (*40*). With the exception of the two Acs enzymes, the overall rms deviations for pairs of different N- and C-terminal domains are similar (Table 3). It is interesting to note that enzymes of the adenylate-forming family are not more conserved in sequence identity or structural homology within a single subfamily than between different subfamilies. CBAL,

Table 3: Sequence Identity (%) and rms Deviation (Å) between Structurally Characterized Members of the Adenylate-Forming Family[a]

	Acs (S. ent.)	Acs1 (yeast)	CBAL	DhbE	PheA
Acs1 (yeast)	42.5% 0.9 Å/1.1 Å				
CBAL	20.9% 2.0 Å/1.5 Å	21.2% 1.9 Å/1.6 Å			
DhbE	21.9% 1.9 Å/1.5 Å	18.7% 1.9 Å/1.3 Å	24.4% 2.0 Å/1.2 Å		
PheA	20.7% 2.1 Å/1.5 Å	19.4% 2.1 Å/1.7 Å	18.6% 1.9 Å/1.4 Å	24.0% 1.9 Å/1.5 Å	
Luc	21.2% 2.0 Å/1.6 Å	20.9% 2.0 Å/1.5 Å	23.0% 2.0 Å/1.3 Å	25.2% 1.9 Å/1.5 Å	20.4% 1.8 Å/1.6 Å

[a] RMS deviations represent values for the N- and C-terminal domains.
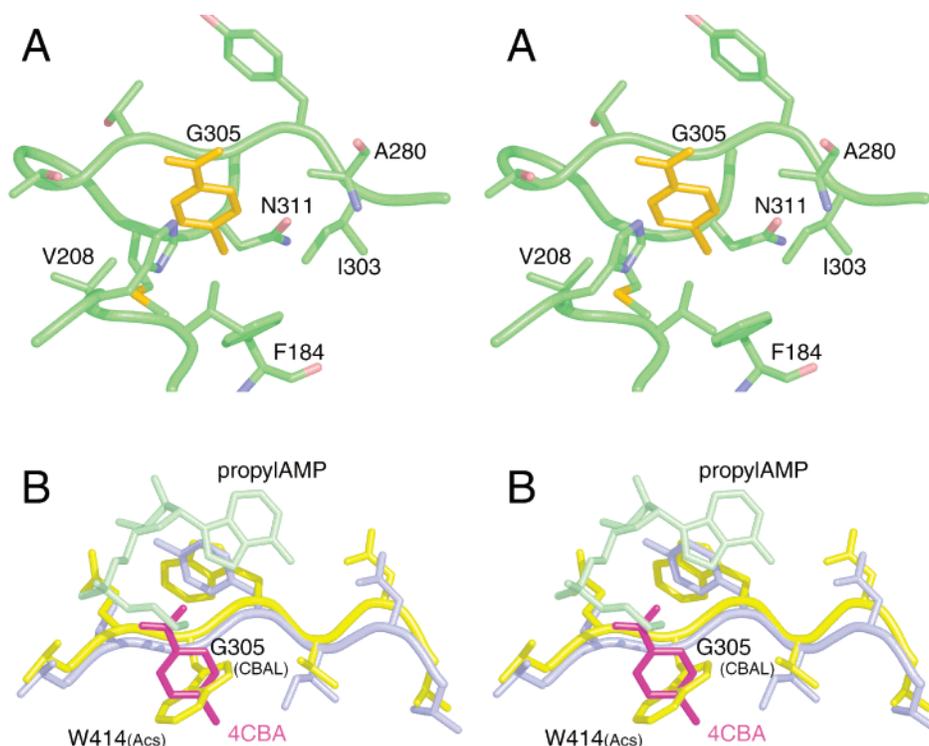
FIGURE 4: (A) The binding pocket for the 4CBA substrate. The CBAL main chain is represented with a ribbon; select side chains are shown with carbon atoms colored green, oxygen atoms red, and nitrogen atoms blue. The 4CBA molecule is shown in gold. Gly305, without a side chain, is positioned between Tyr304 and Thr306. (B) A comparison of the active sites of bacterial Acs and CBAL. Atoms shown in yellow and green are from the Acs structure; the atoms in magenta and blue are from CBAL. The adenosine-5′-propyl phosphate molecule (propylAMP) illustrates the position of the adenylate intermediate in Acs. The ribbon represents a single region from the enzymes (V[410]DTWWQT of Acs and V[301]NIYGTT of CBAL) and illustrates how the Trp414 of the Acs enzyme closes off the active site to larger acyl substrates. The replacement of this residue with Gly305 in CBAL and nearly all other family members allows the enzymes to accommodate larger substrates.

for example, exhibits higher sequence identity with luciferase and DhbE than with the Acs enzymes.

The quaternary structure of CBAL demonstrates that the protein is a dimer (Figure 3B). This is consistent with biochemical analyses for the enzyme from *Pseudomonas* (*1*) or *Alcaligenes* (X. L. and D. D. M., data not shown). In the crystalline lattice, the dimer aligns around a crystallographic 2-fold along the major axes. The dimer interface was determined to be 1380Å², which represents ∼7% of the total surface area of a monomer.

Structure of Liganded Proteins. To gain insight into the catalytic activity of the enzyme, we grew crystals in the presence of 4CBA and AMP. It was anticipated that the enzyme would bind to both substrates and reveal an adenylate-like complex in the active site. The structure of the 4CBA-bound protein was solved by difference Fourier methods; final refinement statistics are shown in Table 2. Upon determination of the structure, there was no apparent density for the nucleotide however unambiguous electron density was observed for the 4CBA molecule (Figure 2B).

There are no large conformational changes of the protein upon binding 4CBA. The rms deviation between the unliganded and liganded structure is 0.2Å for all Cα atoms.

The 4CBA aryl substrate occupies the same binding pocket as the phenylalanine and 2,3-dihydroxybenzoate rings of PheA and DhbE, respectively. The binding pocket for 4CBA is composed of Phe184, His207, Val208, Val209, Phe249, Ala280, Ile303, Gly305, Met310, and Asn311 (Figure 4). Surprisingly, the orientation of the ring is rotated by ∼90° relative to the ring in the PheA structure. The ring of the 2,3-dihydroxybenzoate substrate of DhbE was also observed to be in the orientation similar to CBAL.

DISCUSSION

We have presented the structure of CBAL in the unliganded state as well as bound to the aryl substrate 4CBA. Analysis of the enyzme active site and overall conformation in relation to previously determined structures of related enzymes provides insights into the structure/function relationships of this family of enzymes.

Analysis of the Active Site of CBAL. Although the AMP molecule was not present in our liganded model, we compared the AMP binding site of Acs to CBAL structure. Many of the residues that interact with the AMP moiety of Acs are conserved. Asp411 of Acs, which interacts with the N6 amino group of AMP is replaced by Asn302 in CBAL; Asp500 of Acs interacts with the 2′ and 3′ hydroxyls of AMP and is retained in CBAL at position Asp385. In Acs, the completely conserved Arg515 interacts with the 3′ hydroxyl as well. The side chain of the homologous residue in CBAL, Arg400, is directed out of the active site suggesting that it rotates into position upon binding of the nucleotide. Additional differences in the AMP binding pocket exist between Acs and CBAL. In the Acs structure, the main chain region of Gly387, Glu388, and Pro389 forms a nearly planar group that stacks against the adenine ring (*22*). This region of CBAL (Gly281-Thr283) adopts a different conformation that directs the side chain of Ala282 into the adenine binding pocket suggesting that this region of CBAL will reorient upon binding the nucleotide. Gln415 of Acs interacts with the 3′-hydroxyl of AMP and is replaced in the CBAL structure by Thr306. Unless the main chain is reoriented, the threonine side chain is too short to allow any interactions to occur with the ribose hydroxyls.

The *Alcaligenes sp. AL3007* CBAL shares 57% sequence identity with the *Pseudomonas sp. CBS3* CBAL and 44% sequence identity with *Arthrobacter sp. strain U* enzyme (*47*). The residues of the CBA binding pocket (Phe184, His207, Val208, Val209, Phe249, Ala280, Ile303, Gly305, Met310, and Asn311) are conserved with the exception of Val208 and Val209, which in the Arthrobacter ligase are Asn and Leu, respectively. The large substrate activity constant ($k_{cat}/K_m = 1 \times 10^6$ M^{-1} s^{-1}) measured for 4CBA with the *Alcaligenes* enzyme argues that 4CBA is in fact the physiological substrate. Substrate specificity studies are underway, and preliminary results indicate that benzoate and phenylacetate are very poor substrates, as we might expect based on the structure of the 4CBA binding pocket.

Analysis of the protein structure bound to 4CBA identifies the residues that line the hydrophobic binding pocket for 4CBA. Comparison of the active site structures of Acs and CBAL clearly demonstrates the role of the Trp414 residue of Acs as closing the active site allowing Acs to operate most effectively with the smaller substrates (Figure 4B). Sequence alignment demonstrates a tryptophan in acetyl- and propionyl-CoA synthetases and a glycine in CBAL, PheA, DhbE, luciferase (Figure 5), and nearly all other family members. The appearance of a tryptophan at this position distinguishes the enzymes for small acyl substrates from other family members. The residues that form the 4CBA-binding pocket are similar to those residues that form the NRPS adenylation domain binding pockets, but some notable differences (Figure 5) from the pockets defined in previous studies do exist (*48*). For instance, Phe 184 of CBAL serves a similar role as Trp239 of PheA in closing the base of the substrate binding pocket but, remarkably, the two residues approach the pocket from opposite sides.

Several studies have used the previously determined NRPS adenylation domain structures to model the substrate-binding pockets of plant aryl-CoA synthetases, in particular, the 4-coumarate-CoA ligases. Schneider et al. (*49*) have proposed that the homologue of CBAL Val208 is an important specificity determining residue. This residue is an aspartic acid in PheA, where it interacts with the amino group of the phenylalanine substrate. In DhbE, this residue is Asn235 and interacts with the 2- and 3-hydroxyls of the 2,3-dihydroxybenzoate substrate. In Acs and CBAL, the homologous residue is a valine, while in the plant 4-coumarate:CoA ligases, this residue is an Ile. In these enzymes, this residue makes van der Waals contacts with the substrate.

The substrate-binding pockets of other aryl-CoA ligases have been studied through mutagenesis and the construction of hybrid enzymes. The homologues of CBAL residues Ala280 and Ile303 were mutated in a bacterial cinnamate:CoA ligase (*50*). In this study, mutating alanine residues to glycines increased the size of the binding pocket allowing the enzymes to use more effectively a variety of substituted aromatic substrates. We have identified Ile303 and Gly305, as well as Met 310 and Asn311, as being part of the binding pocket for 4CBA. A study of soybean 4-coumarate:CoA ligase demonstrated that not only are the homologues of these residues important for activity, but mutation of the intervening sequences also caused a loss in activity for a number of substrates (*51*). This suggests that the four-residue loop connecting these two active site regions is important for proper formation of the aryl substrate binding pocket.

Domain Alternation and the Adenylate-Forming Family of Enzymes. The structure of CBAL is of the conformation seen previously for other adenylate-forming enzymes that are either unliganded or bound to compounds that mimic the adenylate intermediate (*20*, *21*, *23*). We have proposed previously that the adenylate-forming family of enzymes use a domain alternation strategy in which the enzyme adopts two different conformations to carry out individual steps of a multistep reaction. The work presented here supports that hypothesis; for the thioester-forming members of this enzyme family, all crystallographic models demonstrate that the enzymes adopt conformation 1 in the unliganded state or when bound to the carboxylate substrate alone or with nucleotide. The binding of CoA is the apparent trigger to form conformation 2 that was observed in the structure of the Acs bound to the inhibitor and to CoA (*22*). The rotation of the C-terminal domain reconfigures a single active site to catalyze the two distinct half-reactions and provides an explanation for the kinetic independence of the two partial reactions observed with CBAL (*4*).

Enzymes with limited sequence homology often use a common structural and mechanistic architecture to catalyze chemically similar reactions (*52*, *53*). If this idea holds for the adenylate-forming enzymes, it implies that the three subfamilies would all utilize the conformational change of the C-terminal domain. Nevertheless, it is valuable to consider the three sub-families separately. The cumulative evidence is strongest for the acyl- and aryl-CoA synthetase subfamily, for which three crystal structures have been determined. While no single enzyme has been structurally characterized in multiple conformations, the Acs enzymes from yeast (*23*) and *S. enterica* (*22*) illustrate the two proposed conformations. Additional biochemical evidence supports the domain rotation. The active-site lysine residue from the C-terminal domain has been shown to be involved in catalysis of the adenylation but not the thioester-forming reaction through site-directed mutagenesis of this residue in

```
CBAL     1   ------------------------------------------------------------------------MQTVNEML-RRAATRAP---D
DhbE     1   ----------------------------------------MLKGFTPWPDE---LAETYRKNGC-----------WAGETFGDLL-RDRAAKYG---D
LUCF     1   ----------------------------------------MEDAKNIKKG---PAPFYPLE--------------DGTAGEQL-HKAMKRYALVPG
Acs-Se   1   MSQTHKHAIPANIADRCLINPEQYETKYKQSINDPDTFWGEQGK-ILDWITPYQKV---KNTSFAPGNVSIKWYEDG-----TLNLAANCLDRHLQENG---D
Acs-Sc  53   PISDRLQPAIATHYSPHLDGLQDYQRLHKESIEDPAKFFGSKATQFLNWSKPFDKVFIPDPKTGRPSFQNNAWFLNG-----QLNACYNCVDRHALKTP---N
PheA     1   ----------------------------------------MLNSSKSILIHAQN---KNGTHEEEQYLFAVNNTKAEYPRDKTIHQLF-EEQVSKRP---N


CBAL    18   HCALAVPA----RGLRLTHAELRARVEAVAARLH-ADGLRPQQRVAVVAPNSADVVIAILALHRLGAVPALLNPRLKSAELAELIKRGEMTAAVIAV------
DhbE    41   RIAITCG------NTHWSYRELDTRADRLAAGFQ-KLGIQQKDRVVVQLPNIKEFFEVIFALFRLGALPVFALPSHRSSEITYFCEFAEAAAYIIPD----AY
LUCF    39   TIAFTDAH----IEVNITYAEYFEMSVRLAEAMK-RYGLNTNHRIVVCSENSLQFFMPVLGALFIGVAVAPANDIYNERELLNSMNISQPTVVFVSK------
Acs-Se  92   RTAIIWEGDDTSQSKHISYRELHRDVCRFANTLL-DLGIKKGDVVAIYMPMVPEAAVAMLACARIGAVHSVIFGGFSPEAVAGRIIDSSSRLVITADEGVRAG
Acs-Sc 148   KKAIIFEGDEPGQGYSITYKELLEEVCQVAQVLTYSMGVRKGDTVAVYMPMVPEAAIITLLAISRIGAIHSVVFAGFSSNSLRDRINDGDSKVVITTDESNRGG
PheA    55   NVAIVCE------NEQLTYHELNVKANQLARIFI-EKGIGKDTLVGIMMEKSIDLFIGILAVLKAGGAYVPIDIEYPKERIQYILDDSQARMLLTQK------


CBAL   110   ----GRQVADAIFQSGS--GA-RIIFL---------------GDLVRDGEPYSYGPP------IEDPQREPAQ-PAFIFYTSGTTGLPKAAIIPQR-AAESR
DhbE   133   SGFDYRSLARQVQSKLP--TLKNIIVAGEAEEF----------LPLE----DLHA-EP------VKLPEVKSSD-VAFLQLSGGSTGLSKLIPRTHD-DYIYS
LUCF   131   ---KGLQKILNVQKKLP--IIQKIIIMDSKTDYQGF-------QSMYT--FVTSHLPPGFNEYDFVPESFDRDKTIALIMNSSGSTGLPKGVALPHR-TACVR
Acs-Se 194   RSIPLKKNVDDALKNPNVTSV-EHVIVLKRTGS-DIDWQEGRDLWWRD--LIEKASPE------HQPEAMNAED-PLFILYTSGSTGKPKGVLHTTGGYLVYA
Acs-Sc 251   KVIETKRIVDDALRE--TPGV-RHVLVYRKTNNPSVAFHAPRDLDWAT--EKKKYKTY------YPCTPVDSED-PLFLLYTSGSTGAPKGVQHSTAGYLLGA
PheA   145   ---HLVHLIHNI---QF--NG-QVEIF------------------EE----DTIKIREGT------NLHVPSKSTD-LAYVIYTSGTTGNPKGTMLEHK-GISNL


CBAL   182   VLFM--STQVGLRHGRHNVVLGIMPLYHVVGFF-AVLVAALALDGTYVVIEEF---RP----VDALQLVQQEQVTSLFATPTHL-DALAAAAAHAGSSLKLDS
DhbE   211   LKRS--VEVC-WLDH-STVYLAALPMAHNYPLSSPGVLGVLYAGGRVVLSPSP---SP----DDAFPLIEREKVTITALVPPLAMVW-MDAASS--RRDDLSS
LUCF   219   FSHARDPIFG-NQIIPDTAILSVVFPHHGFGMF-T-TLGYLICGFRVVLMYRF---EE----ELFLRSLQDYKIQSALLVPTLF-SFFAKSTL--IDKYDLSN
Acs-Se 285   ATTF--KYVFDYH--PGDIYWCTADVGWVTGHS-YLLYGPLACGATTLMFEGV---PNWPTPARMCQVVDKHQVNILYTAPTAI-RALMAEGDKAIEGTDRSS
Acs-Sc 342   LLTM--RYTFDTH--QEDVFFTAGDIGWITGHT-YVVYGPLLYGCATLVFEGT---PAYPNYSRYWDIIDEHKVTQFYVAPTAL-RLLKRAGDSYIENHSLKS
PheA   211   KVFF--ENSLNVT--EKDRIGQFASISFDASV--WEMFMALLTGASLYIILKDTINDF----VKFEQYINQKEITVITLPPTYV-VHLDPER--------ILS


CBAL   274   LRHVTF--AGATMPDAVLETVHQHLP---G-EKVNIYGTTEA-MNSLYMRQ-------P---KTGTEMAP-GFFSEVRIVR-IGGGVDEIVANG-EEGELIV-
DhbE   300   LQVIQV--GGAKFSAEAARRVKAVFG---C-TLQQVFGMAEG-LV-NYTRLD-DPEEII---VN-TQGKPMSPYDESRVWD--DHDR--DVKPG-ETGHLLT-
LUCF   309   LHEIASGGAPLS--KEVGEAVAKRFH---LPGIRQGYGLTETTSALIITPEG-DDK-------PGAVGKV-VPFFEAKVVD-LDTG--KTLGVN-QRGELCV-
Acs-Se 380   LRILGS--VGEPINPEAWEWYWKKIGKEKC-PVVDTWWQTETGGFMITPLPG-AIE--L----KAGSATRP-FFGVQPALVD-NEG---HPQEGA-TEGNLVIT
Acs-Sc 436   LRCLGS--VGEPIAAEVWEWYSEKIGKNEI-PIVDTYWQTESGSHLVTPLAGGVTP--M----KPGSASFP-FFGIDAVVLDPNTG---EELNTSHAEGVLAVK
PheA   295   IQTLIT--AGSATSPSLVNKWKE--K---V-TYINAYGPTETTICATTWVAT-KE---TIGH-SVPIGAP-IQNTQIYIVD---ENLQ-LKSVG-EAGELCI-


CBAL   356   AASDSAFVGYLNQPQATA-EKLQ--D-------GWYRTSDVAVWTPEGTVRILGRVDDMIISGGENIHPSEIERVLGTAPGVTEVVVIGLADQRWGQSVTACV
DhbE   384   R-GPYTIRGYYKAEEHNA-ASFTE-D-------GFYRTGDIVRLTRDGYIVVEGRAKDQINRGGEKVAAEEVENHLLAHPAVHDAAMVSMPDQFLGERSCVFI
LUCF   393   R-GPMIMSGYVNNPEATN-ALIDK-D-------GWLHSGDIAYWDEDEHFFIVDRLKSLIKYKGYQVAPAELESILLQHPNIFDAGVAGLPDDDAGELPAAVV
Acs-Se 468   DSWPGQARTLFGDHERFEQTYFSTFK-------NMYFSGDGARRDEDGYYWITGRVDDVLNVSGHRLGTAEIESALVAHPKIAEAAVVGIPHAIKGQAIYAYV
Acs-Sc 527   AAWPSFARTIWKNHDRYLDTYLNPYP-------GYYFTGDGAAKDKDGYIWILGRVDDVVNVSGHRLSTAEIEAAIIEDPIVAECAVVGFNDDLTGQAVAAFV
PheA   378   G-GEGLARGYWKRPELTS-QKFV--DNPFVPGEKLYKTGDQARWLSDGNIEYLGRIDNQVKIRGHRVELEEVESILLKHMYISETAVSVHKDHQEQPYLCAYF


CBAL   449   VPRLGETL----------SADAL-DTFCRSSELADFKRPK-RYFILDQLPKNALNKVLRRQLV-QQV-SS
DhbE   477   IPR-DEAP----------KAAEL-KAFLRERGLAAYKIPD-RVEFVESFPQTGVGKVSKKALR-EAI-SEKLLAGFKK
LUCF   486   VLEHGKTM----------TEKEIVD-YVRSQ-VTTAKKLRGGVVFVDEVPKGLTGKLDARKIR-E-ILIKAKKGGKSKL
Acs-Se 564   TLNHGEEPS-------PELYAEV-RNWVR-KEIGPLATPD-VLHWTDSLPKTRSGKIMRRILRKI---AAGDTSNLGDTSTLADPGVVEKLLEEKQAIAMPS
Acs-Sc 623   VLKNKSSWSTATDDELQDIKKHL-VFTVR-KDIGPFAAPK-LIILVDDLPKTRSGKIMRRILRKI---LAGESDQLGDVSTLSNPGIVRHLIDSVKL
PheA   477   VS-E-KHI----------PLEQL-RQFSSEE-LPTYMIPS-YFIQLDKMPLTSNGKIDRKQL----P-EPDL
```
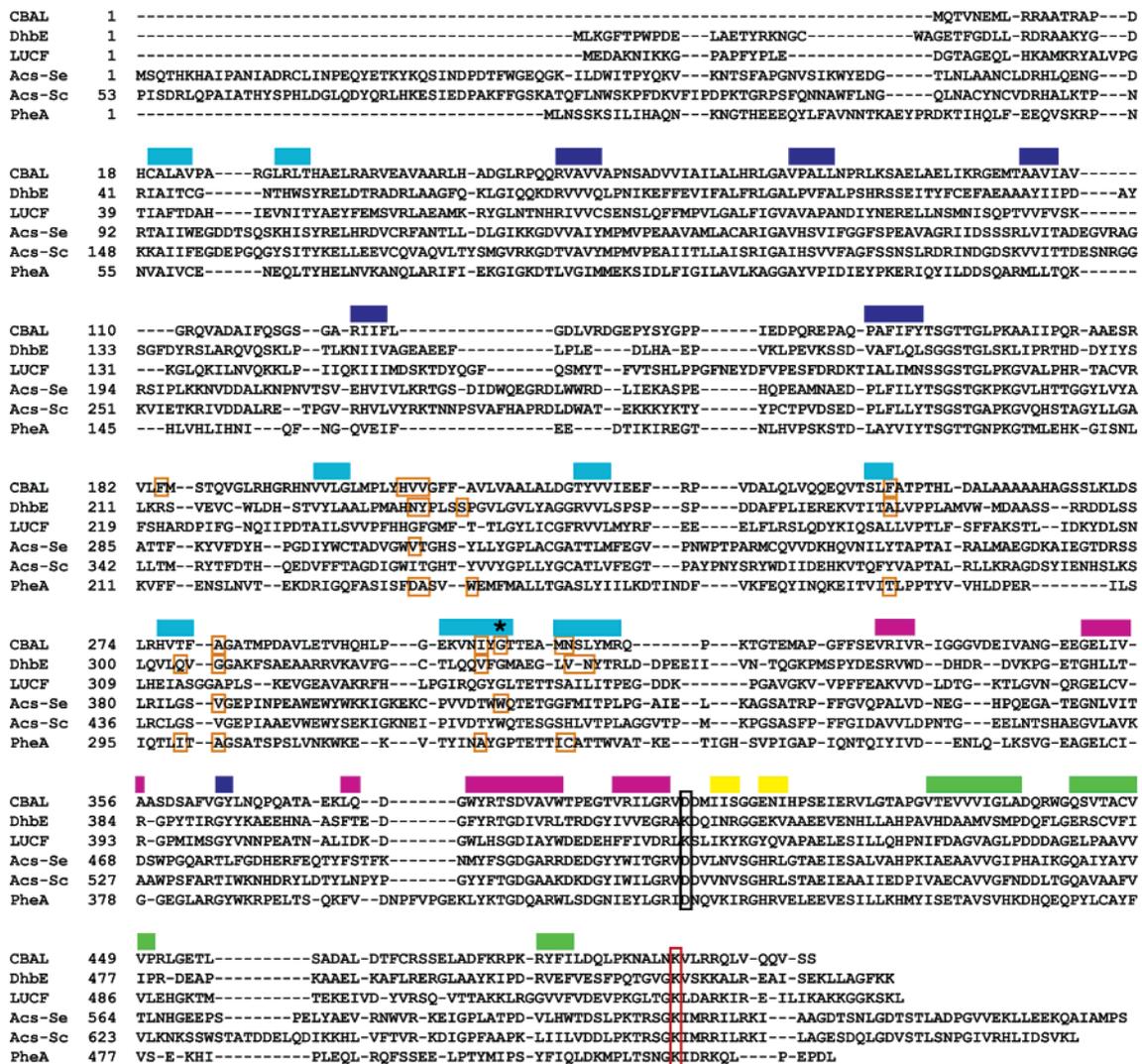
FIGURE 5: Sequence alignment of the six structurally characterized members of the adenylate-forming family. Enzymes were aligned pairwise using DALI (*62, 63*) and ALIGN (*40*) to align based on structural similarity. To align the luciferase and bacterial Acs structures, the N- and C-terminal domains were treated separately. The colored boxes above the CBAL sequence represent the locations of the strands that compose the five main sheets of the enzyme. The N-terminal sheets are colored blue, cyan, and magenta, while the C-terminal sheets are colored yellow and green, consistent with the sheets shown in Figure 3A. For the four structures that have been crystallized in the presence of the acyl substrates, the residues that form the acyl binding pocket are indicated with orange boxes. The black box indicates the position of the hinge residue that is located at the boundary between the N- and C-terminal domains. The red box denotes the conserved lysine at the C-terminal domain that interacts with substrate in the PheA structure. The asterisk indicates the position of CBAL Gly305 which is replaced by a Trp in the acetyl- and propionyl-CoA synthetases.

propionyl-CoA synthetase (*11*) or studies of acetylation of this residue in Acs (*54*).

Similarly, mutation of the C-terminal domain lysine of luciferase results in an enzyme that is unable to catalyze the adenylation reaction yet is competent to catalyze the second partial reaction (*55*). No mutations that specifically affect the second half-reaction while leaving the adenylation step unchanged have been reported for any members of this enzyme family. The X-ray crystal structure of luciferase determined in the absence of substrates (*19*) contains a C-terminal domain orientation that is intermediate in position between the two conformations observed in the other structures. In this position, the C-terminal lysine residue is close to the active site although not in the same location as was observed for the homologous residue of PheA, DhbE, or CBAL. Whether the C-terminal domain of luciferase adopts the conformations observed in liganded structures of other family members remains to be seen.

The proposed use of domain alternation by the NRPS adenylation domains is supported by changes in proteolytic susceptibility upon addition of substrate (*56*) and by the conservation of amino acid sequences on both sides of the C-terminal domain (*9*). Only a single protein conformation has been crystallized (conformation 1) and crystallization of a molecule in the proposed thioester-forming conformation may require the crystallization of a multidomain NRPS enzyme.

The use of the large domain rearrangement by the NRPS enzymes is particularly intriguing. Most NRPSs are composed of multiple catalytic domains expressed as part of a single large modular protein. If structurally rigid linkages exist between the adenylation domains and all upstream and downstream NRPS domains, then the rearrangement of the N- and C-terminal "subdomains" within the adenylation domain would cause a rotation of the upstream and downstream domains with respect to each other as well.

Other examples of enzymes that use domain alternation include methionine synthase (*24*), *E. coli* thioredoxin reductase (*57*), pyruvate phosphate dikinase (*58*, *59*), and many biotin-dependent decarboxylases (*60*). Compared to these other examples, the use of domain alternation by the adenylate-forming enzymes is unique. While the prior examples described above use the conformational change to transport a substrate or cofactor from one active site to another (*60*), the adenylate-forming enzymes appear to use this change to reconfigure a single active site, enabling the catalysis of the second chemically distinct half-reaction. This strategy is therefore a way to utilize most efficiently the common features of a single active site that are shared for the two half-reactions. The different faces of the C-terminal domain are presented sequentially to the active site for involvement in the individual half-reactions.

Quaternary Structure of Adenylate-Forming Enzymes. The CBAL enzyme is a dimer both in solution as well as in the crystalline lattice. The bacterial Acs enzyme (*22*) was shown to be a monomer by both structural work as well as gel filtration analysis. The yeast Acs1 enzyme, in contrast, is a trimer in both the crystalline lattice as well as in solution (*23*). Alignment of the yeast trimeric Acs1 with the dimeric CBAL enzyme illustrates that different faces of the enzymes are utilized in the formation of the multimers. Both enzymes, however, use the N-terminal domain to form the multimer interface, leaving the C-terminal domain free to undergo the proposed conformational change.

CONCLUSIONS

We present here the 2.0 and 2.2 Å crystal structures of CBAL both in the unliganded state and bound to 4CBA. The structure is consistent with the hypothesis that members of the adenylate-forming family of enzymes adopt two distinct conformations to catalyze the two-step adenylation and thioester-forming reaction. While the hypothesis would be most strongly supported by the determination of the structures of a single enzyme in different conformational states when bound to analogues along the reaction mechanism, the recent determination of the structure of the yeast Acs1 enzyme (*23*) suggests that members of this family will adopt the different conformations during catalysis of the two reactions. Domain alternation in the adenylate-forming enzymes can be added to the list of strategies that enzymes use to efficiently catalyze chemically difficult reactions. The enzymes appear to use one part of the active site for the shared functions in the two half-reactions and use a mobile C-terminal domain to bring to the active site important residues that are involved in the distinct steps of the adenylation and thiolation reactions.

ACKNOWLEDGMENT

REFERENCES

1. Chang, K. H., Liang, P. H., Beck, W., Scholten, J. D., and Dunaway-Mariano, D. (1992) Isolation and characterization of the three polypeptide components of 4- chlorobenzoate dehalogenase from *Pseudomonas sp. strain CBS-3*, *Biochemistry 31*, 5605−5610.

2. Dunaway-Mariano, D., and Babbitt, P. C. (1994) On the origins and functions of the enzymes of the 4-chlorobenzoate to 4-hydroxybenzoate converting pathway, *Biodegradation 5*, 259−276.

3. Layton, A. C., Sanseverino, J., Wallace, W., Corcoran, C., and Sayler, G. S. (1992) Evidence for 4-chlorobenzoic acid dehalogenation mediated by plasmids related to pSS50, *Appl. Environ. Microbiol. 58*, 399−402.

4. Chang, K. H., and Dunaway-Mariano, D. (1996) Determination of the chemical pathway for 4-chlorobenzoate:coenzyme A ligase catalysis, *Biochemistry 35*, 13478−13484.

5. Chang, K. H., Xiang, H., and Dunaway-Mariano, D. (1997) Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: a site-directed mutagenesis study with the *Pseudomonas sp. strain CBS3* 4-chlorobenzoate: coenzyme A ligase, *Biochemistry 36*, 15650−15659.

6. Babbitt, P. C., Kenyon, G. L., Martin, B. M., Charest, H., Slyvestre, M., Scholten, J. D., Chang, K. H., Liang, P. H., and Dunaway-Mariano, D. (1992) Ancestry of the 4-chlorobenzoate dehalogenase: analysis of amino acid sequence identities among families of acyl: adenyl ligases, enoyl-CoA hydratases/isomerases, and acyl-CoA thioesterases, *Biochemistry 31*, 5594−5604.

7. Gimeno, R. E., Ortegon, A. M., Patel, S., Punreddy, S., Ge, P., Sun, Y., Lodish, H. F., and Stahl, A. (2003) Characterization of a heart-specific fatty acid transport protein, *J. Biol. Chem. 278*, 16039−16044.

8. Stahl, A., Gimeno, R. E., Tartaglia, L. A., and Lodish, H. F. (2001) Fatty acid transport proteins: a current view of a growing family, *Trends Endocrinol. Metab. 12*, 266−273.

9. Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis, *Chem. Rev. 97*, 2651−2674.

10. White, H. E., Miano, J. D., and Umbreit, M. (1975) Letter: on the mechanism of firefly luciferin luminescence, *J. Am. Chem. Soc. 97*, 198−200.

11. Horswill, A. R., and Escalante-Semerena, J. C. (2002) Characterization of the propionyl-CoA synthetase (PrpE) enzyme of *Salmonella enterica*: residue Lys592 is required for propionyl-AMP synthesis, *Biochemistry 41*, 2379−2387.

12. Grayson, N. A., and Westkaemper, R. B. (1988) Stable analogues of acyl adenylates. Inhibition of acetyl- and acyl-CoA synthetase by adenosine 5′-alkyl-phosphates, *Life Sci. 43*, 437−444.

13. Walsh, C. T., Chen, H., Keating, T. A., Hubbard, B. K., Losey, H. C., Luo, L., Marshall, C. G., Miller, D. A., and Patel, H. M. (2001) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines, *Curr. Opin. Chem. Biol. 5*, 525−534.

14. Mootz, H. D., and Marahiel, M. A. (1999) Design and application of multimodular peptide synthetases, *Curr. Opin. Biotechnol. 10*, 341−348.

15. Belshaw, P. J., Walsh, C. T., and Stachelhaus, T. (1999) Amino-acyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis, *Science 284*, 486-489.

16. Gehring, A. M., Bradley, K. A., and Walsh, C. T. (1997) Enterobactin biosynthesis in *Escherichia coli*: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate, *Biochemistry 36*, 8495−8503.

17. Ehmann, D. E., Shaw-Reid, C. A., Losey, H. C., and Walsh, C. T. (2000) The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates, *Proc. Natl. Acad. Sci. U.S.A. 97*, 2509−2514.

18. May, J. J., Wendrich, T. M., and Marahiel, M. A. (2001) The dhb operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin, *J. Biol. Chem. 276*, 7209−7217.

19. Conti, E., Franks, N. P., and Brick, P. (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes, *Structure 4*, 287−298.

20. Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the nonribosomal biosynthesis of gramicidin S, *EMBO J. 16*, 4174−4183.

21. May, J. J., Kessler, N., Marahiel, M. A., and Stubbs, M. T. (2002) Crystal structure of DhbE, an archetype for aryl acid activating

domains of modular nonribosomal peptide synthetases, *Proc. Natl. Acad. Sci. U.S.A. 99*, 12120−12125.

22. Gulick, A. M., Starai, V. J., Horswill, A. R., Homick, K. M., and Escalante-Semerena, J. C. (2003) The 1.75 A crystal structure of acetyl-CoA synthetase bound to adenosine-5′-propyl-phosphate and coenzyme A, *Biochemistry 42*, 2866−2873.

23. Jogl, G., and Tong, L. (2004) Crystal structure of yeast acetyl-coenzyme A synthetase in complex with AMP, *Biochemistry 43*, 1425−1431.

24. Bandarian, V., Pattridge, K. A., Lennon, B. W., Huddler, D. P., Matthews, R. G., and Ludwig, M. L. (2002) Domain alternation switches B(12)-dependent methionine synthase to the activation conformation, *Nat. Struct. Biol. 9*, 53−56.

25. Cukovic, D., Ehlting, J., VanZiffle, J. A., and Douglas, C. J. (2001) Structure and evolution of 4-coumarate: Coenzyme A ligase (4CL) gene families, *Biol. Chem. 382*, 645−654.

26. Benning, M. M., Wesenberg, G., Liu, R., Taylor, K. L., Dunaway-Mariano, D., and Holden, H. M. (1998) The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas sp. Strain CBS-3*, *J. Biol. Chem. 273*, 33572−33579.

27. Benning, M. M., Taylor, K. L., Liu, R. Q., Yang, G., Xiang, H., Wesenberg, G., Dunaway-Mariano, D., and Holden, H. M. (1996) Structure of 4-chlorobenzoyl coenzyme A dehalogenase determined to 1.8 A resolution: an enzyme catalyst generated via adaptive mutation, *Biochemistry 35*, 8103−8109.

28. Deng, J., Davies, D. R., Wisedchaisri, G., Wu, M., Hol, W. G., and Mehlin, C. (2004) An improved protocol for rapid freezing of protein samples for long-term storage, *Acta Crystallogr., Sect. D 60*, 203−204.

29. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin, *J. Mol. Biol. 229*, 105−124.

30. Jancarik, J., and Kim, S.-H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins, *J. Appl. Crystallog. 24*, 409−411.

31. Carter, C. W., Jr., and Carter, C. W. (1979) Protein crystallization using incomplete factorial experiments, *J. Biol. Chem. 254*, 12219−12223.

32. Weeks, C. M., Blessing, R. H., Miller, R., Mungee, R., Potter, S. A., Rappleye, J., Smith, G. D., Xu, H., and Furey, W. (2002) Towards automated protein structure determination: *BnP*, the *SnB*-PHASES interface,, *Z. Kristallogr. 217*, 686−693.

33. Terwilliger, T. C. (2001) Maximum-likelihood density modification with pattern recognition of structural motifs, *Acta Crystallogr., Sect. D 57*, 1755−1762.

34. Furey, W., and Swaminathan, S. (1997) PHASES-95: A program package for processing and analysis of diffraction data from macromolecules, *Methods Enzymol. 277*, 590−620.

35. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr., Sect. D 54*, 905−921.

36. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr., Sect. D 53*, 240−255.

37. Colapietro, M., and Domenicano, A. (1982) Structural studies of benzene derivatives. X. Refinement of the structure of *p*-chlorobenzoic acid, *Acta Crystallogr., Sect. B 38*, 1953−1957.

38. Miller, R. S., Paul, I. C., and Curtin, D. Y. (1974) Reactions of molecular crystals with gases. II. X-ray structure of crystalline 4-chlorobenzoic acid and the anisotropy of its reaction with ammonia gas, *J. Am. Chem. Soc. 96*, 6334−6339.

39. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement, *Acta Crystallogr., Sect. D 57*, 122−133.

40. Cohen, G. H. (1997) ALIGN: A Program to superimpose protein coordinates, accounting for insertions and deltions, *J. Appl. Crystallography 30*, 1160−1161.

41. Kabsch, W. (1976) A solution for the best rotation to relate two sets of vectors, *Acta Crystallogr., Sect. A 32*, 922−923.

42. DeLano, W. L. (2002) *The Pymol Molecular Graphics System*, DeLano Scientific: San Carlos, CA (http://www.pymol.org).

43. CCP4. (1994) The CCP4 suite: programs for protein crystal-lography, *Acta Crystallogr., Sect. D 50*, 760−763.

44. Lee, B., and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility, *J. Mol. Biol. 55*, 379−400.

45. Saraste, M., Sibbald, P. R., and Wittinghofer, A. (1990) The P-loop−a common motif in ATP- and GTP-binding proteins, *Trends Biochem. Sci. 15*, 430−434.

46. Smith, C. A., and Rayment, I. (1996) Active site comparisons highlight structural similarities between myosin and other P-loop proteins, *Biophys. J. 70*, 1590−1602.

47. Schmitz, A., Gartemann, K. H., Fiedler, J., Grund, E., and Eichenlaub, R. (1992) Cloning and sequence analysis of genes for dehalogenation of 4-chlorobenzoate from Arthrobacter sp. strain SU, *Appl. Environ. Microbiol. 58*, 4068−4071.

48. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribo-somal peptide synthetases, *Chem. Biol. 6*, 493−505.

49. Schneider, K., Hovel, K., Witzel, K., Hamberger, B., Schomburg, D., Kombrink, E., and Stuible, H. P. (2003) The substrate specificity-determining amino acid code of 4-coumarate: CoA ligase, *Proc. Natl. Acad. Sci. U.S.A. 100*, 8601−8606.

50. Kaneko, M., Ohnishi, Y., and Horinouchi, S. (2003) Cinnamate: Coenzyme A ligase from the filamentous bacterium *Streptomyces coelicolor* A3(2), *J. Bacteriol. 185*, 20−27.

51. Lindermayr, C., Fliegmann, J., and Ebel, J. (2003) Deletion of a single amino acid residue from different 4-coumarate: CoA ligases from soybean results in the generation of new substrate specifici-ties, *J. Biol. Chem. 278*, 2781−2786.

52. Armstrong, R. N. (2000) Mechanistic diversity in a metalloenzyme superfamily, *Biochemistry 39*, 13625−13632.

53. Gerlt, J. A., and Babbitt, P. C. (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies, *Annu. Rev. Biochem. 70*, 209−246.

54. Starai, V. J., Celic, I., Cole, R. N., Boeke, J. D., and Escalante-Semerena, J. C. (2002) Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine, *Science 298*, 2390−2392.

55. Branchini, B. R., Murtiashaw, M. H., Magyar, R. A., and Anderson, S. M. (2000) The role of lysine 529, a conserved residue of the acyl-adenylate- forming enzyme superfamily, in firefly luciferase, *Biochemistry 39*, 5433−5440.

56. Dieckmann, R., Pavela-Vrancic, M., von Dohren, H., and Kleinkauf, H. (1999) Probing the domain structure and ligand-induced conformational changes by limited proteolysis of tyrocidine synthetase 1, *J. Mol. Biol. 288*, 129−140.

57. Lennon, B. W., Williams, C. H., Jr., and Ludwig, M. L. (2000) Twists in catalysis: alternating conformations of *Escherichia coli* thioredoxin reductase, *Science 289*, 1190−1194.

58. Wei, M., Ye, D., and Dunaway-Mariano, D. (2001) Investigation of the role of the domain linkers in separate site catalysis by *Clostridium symbiosum* pyruvate phosphate dikinase, *Biochemistry 40*, 13466−13473.

59. Herzberg, O., Chen, C. C., Kapadia, G., McGuire, M., Carroll, L. J., Noh, S. J., and Dunaway-Mariano, D. (1996) Swiveling-domain mechanism for enzymatic phosphotransfer between remote reac-tion sites, *Proc. Natl. Acad. Sci. U.S.A. 93*, 2652−2657.

60. Perham, R. N. (2000) Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reac-tions, *Annu. Rev. Biochem. 69*, 961−1004.

61. Terwilliger, T. C. (2000) Maximum likelihood density modifica-tion, *Acta Crystallogr., Sect. D 56*, 965−972.

62. Holm, L., and Sander, C. (1997) Dali/FSSP classification of three-dimensional protein folds, *Nucleic Acids Res. 25*, 231−234.

63. Holm, L., and Sander, C. (1995) Dali: a network tool for protein structure comparison, *Trends Biochem. Sci. 20*, 478−480.

BI049384M