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 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 aryI 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), 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 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...

‖‡ This research was supported in part by start-up funds from the John R. Oishei Foundation (A. M. G.) and the National Institutes of Health (GM28688 to D. D.-M.). Structure determination by BnP was supported by NIH/NIBIB Grant EB002057 to Charles M. Weeks (HWI). Data were collected at beamline 12C of the National Synchrotron Light Source which is supported by the Offices of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy, and from the National Center for Research Resources of the National Institutes of Health.
¶ The atomic coordinates and structure factors for both the liganded and unliganded molecules of 4-chlorobenzoyl-CoA ligase/synthetase have been deposited with the Protein Data Bank (accession codes 1TSD and 1TSH, respectively).
§ 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.

1 Abbreviations: 4CBA, 4-chlorobenzoate; CBAL, 4-chlorobenzoic acid: CoA ligase from Alcaligenes sp. AL3007; PheA, the phenylalanine-activating domain of Gramicidin Synthetase A; Acs1, acetyl-CoA synthetase from Saccharomyces cerevisiae; SeMet, selenomethionine; NRPS, nonribosomal peptide synthetase; DHB, 2,3-dihydroxybenzoate; rms, root-mean-square; SnB, The Shake-and-Bake computer program for the identification of heavy atom structures by direct methods; BnP, an automated implementation of the SnB algorithm with the heavy atom and phase refinement of PHASES.
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 aryladenylate 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 ($K_{cat} \approx 9$ s$^{-1}$, Mg$^{2+}$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$^{2+}$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.

---

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

<table>
<thead>
<tr>
<th>NRPS Aden. domain</th>
<th>Acyl-CoA Synthetase</th>
<th>Firefly Luciferase</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{NRPS Aden. domain}^a$</td>
<td>$\text{RCO}_2^- + \text{ATP}$</td>
<td>$\text{Luc}-\text{CO}_2^- + \text{ATP}$</td>
</tr>
<tr>
<td>$\text{NRPS Aden. domain}^a$</td>
<td>$\text{RCO}_2^- + \text{ATP}$</td>
<td>$\text{Luc}-\text{CO}_2^- + \text{ATP}$</td>
</tr>
<tr>
<td>$\text{RCO}_2^- + \text{ATP}$</td>
<td>$\text{RCO}^- + \text{CoA}$</td>
<td>$\text{Luc}-\text{CO}^- + \text{ATP}$</td>
</tr>
<tr>
<td>$\text{RCO}^- + \text{CoA}$</td>
<td>$\text{RCO}^- + \text{CoA}$</td>
<td>$\text{Luc}^+ \rightarrow \text{light}$</td>
</tr>
</tbody>
</table>

$^a$ PPant represents the phosphopantetheine cofactor for the NRPS carrier protein domains which follow the adenylation domain. $^b$ R = Cl–C6H4 for CBAL.
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 CaCl₂, and 50 mM CHES (pH 9.0). Crystals mounted in a quartz capillary or cryocooled at −170 °C in a stream of N₂ 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 P3₁21 or P3₂1 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 CaCl₂, 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 P3₁21 and 1.74 for P3₂1, 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 PHASERS (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...
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 crystal structures of 4CBA alone (37, 39). 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 crystal structures of 4CBA alone (37, 39). 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 crystal structures of 4CBA alone (37, 39). 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 crystal structures of 4CBA alone (37, 39). 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 crystal structures of 4CBA alone (37, 39). These data were also collected at beamline ×12C of Brookhaven National Laboratory on a crystal that diffracted to 2.0 Å.
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).

Coordinated with the experimental data, the overall structure of 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,
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 structures 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 enzyme 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, Thr283, 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 (\(k_{cat}/K_m = 1 \times 10^4 \text{ M}^{-1} \text{s}^{-1}\)) 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, Phe184 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 Met310 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 this 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 subfamilies 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
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.

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 both 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**

We would like to thank Kelly Dearing for assistance with crystal growth, Dr. Michael Malkowski for assistance with data collection, and Dr. Charles Weeks for providing helpful suggestions and access to BnP. We would also like to thank Dr. Liang Tong for access to the yeast Acs1 coordinates prior to release from the PDB.

**REFERENCES**
