The mineralization of aromatic compounds by microorganisms relies on a structurally and functionally diverse group of ring-cleaving enzymes. The recently discovered benzoate oxidation pathway in Burkholderia xenovorans LB400 encodes a novel such ring-cleaving enzyme, termed BoxC, that catalyzes the conversion of 2,3-dihydro-2,3-dihydroxybenzoyl-CoA to 3,4-dehydroadipyl-CoA without the requirement for molecular oxygen. Sequence analysis indicates that BoxC is a highly divergent member of the crotonase superfamily and nearly double the size of the average superfamily member. The structure of BoxC determined to 1.5 Å resolution reveals an intriguing structural demarcation. A highly divergent region in the C terminus probably serves as a structural scaffold for the conserved N terminus that encompasses the active site and, in conjunction with a conserved C-terminal helix, mediates dimer formation. Isothermal titration calorimetry and molecular docking simulations contribute to a detailed view of the active site, resulting in a compelling mechanistic model where a pair of conserved glutamate residues (Glu146 and Glu168) work in tandem to deprotonate the dihydroxylated ring substrate, leading to cleavage. A final deformylation step incorporating a water molecule and Cys111 as a general base completes the formation of 3,4-dehydroadipyl-CoA product. Overall, this study establishes the basis for BoxC as one of the most divergent members of the crotonase superfamily and provides the first structural insight into the mechanism of this novel class of ring-cleaving enzymes.

Aromatic compounds comprise approximately one-quarter of the earth’s biomass (1) and are the second most abundant natural product next to carbohydrates. The majority of aromatic compounds in the environment are in the form of the organic polymer lignin that plays a structural role in cross-linking cell wall polysaccharides in plants. Despite the inherent thermostability of the aromatic ring, these naturally occurring compounds are efficiently mineralized by various microorganisms. Human-made aromatic compounds, such as those used in industrial processes, however, are often recalcitrant to microbial degradation due to their chemical complexity, decreased bioavailability, and increased thermostability. Moreover, bacteria have only been exposed to these compounds for a relatively short period of time. As a result, these compounds persist in the environment, where they can increase to toxic levels and cause irreversible damage to the biosphere.

The common structural blueprint shared by natural and human-made aromatic compounds is the resonance-stabilized planar ring system. Microorganisms overcome the stability of these aromatic structures by employing specific ring-cleaving enzymes that form part of complex catabolic pathways. Until recently, two general classes of microbial processes were characterized that catalyze the degradation of aromatic compounds. These classifications, termed the aerobic and anaerobic pathways, were based primarily on the mode of initial activation and subsequent cleavage of the aromatic ring. The aerobic pathway, exemplified by the peripheral biphenyl and the central ben-cat pathway, relies on the extensive use of molecular oxygen for both the hydroxylation (activation) and cleavage of the aromatic ring (2–4). The anaerobic pathway, however, mediates a reductive dearomatization followed by a hydrolytic ring cleavage, as observed in the classical benzoate pathway (5–7). In both cases, the underlying mechanism incorporates an activation step that renders the ring susceptible to cleavage.

Recently, a third aromatic degradation pathway was identified in Burkholderia xenovorans strain LB400 (LB400) (8–10) and Azoarcus evansii (11–13). This novel pathway, termed the box (benzoate oxidation) pathway, incorporates features of both the aerobic and anaerobic pathways, resulting in a hybrid pathway. Microarray analysis of the 9.7-Mb genome of LB400 revealed two paralogous copies of the box pathway, one encoded on chromosome 1 (boxA) and the second on the megaplasmid (boxm) (9). Knock-out studies confirm that both box pathways are capable of assimilating benzoate (10) yet are differentially regulated based on available carbon source and growth phase of the organism (9). Recent structural and biochemical characterization of benzoate CoA ligase (14) and aldehyde dehydrogenase (15) from the box pathway in LB400 have provided valuable insight into the basis of substrate specificity and details describing the molecular mechanisms.

A unique feature of the hybrid box pathway is the incorporation of both CoA ligation and hydroxylation prior to ring cleavage (16), suggesting that both strategies are important for ring
Structural Characterization of BoxCC

activation. It is noteworthy that although CoA ligation is common in the activation of aromatic acids under anaerobic conditions, it has thus far been unseen in the aerobic degradation of aromatic compounds. Furthermore, investigation of the box pathway intermediates from the related A. evansii demonstrated that the thioesterified dihydrodiol intermediate was not oxidized and re aromatized as normally occurs in aerobic aromatic metabolism (11). Instead, it was shown to be directly cleaved without the requirement of molecular oxygen in a reaction that resulted in the loss of one unit of carbon and oxygen as formate (11). This critical ring cleavage step in the box pathway is catalyzed by BoxC (2,3-dihydro-2,3-dihydroxybenzoyl-CoA lyase/ hydrolyase) (11), which differs from traditional aerobic and anaerobic ring-cleaving enzymes in that oxygen is not used in catalysis, and the ring substrate is only partially reduced. Based on sequence analysis, BoxC is assigned to the crotonase superfamily. The cleavage reaction catalyzed by BoxC, however, suggests that BoxC defines a new mechanistic niche and intriguingly is one of the four outstanding crotonase superfamilies for which no structural information exists (17).

A mechanism for BoxC from A. evansii was recently proposed based on the identification of chemical species using NMR and mass spectrometry (11). In the absence of structural information of BoxC, however, the mechanistic details, including the identity of the catalytic residues, remain undefined. To investigate the detailed molecular mechanism of BoxC, we carried out a structural and biophysical analysis complemented with molecular docking. The resulting data provide a compelling mechanistic model with the identification of key catalytic residues and active site structure that stabilize proposed transition state intermediates. Furthermore, the 1.5 Å resolution structure of BoxC reveals intriguing divergent architectural features with respect to other members of the crotonase superfamily. Overall, this study provides the first structural characterization of the novel BoxC family of enzymes and is interpreted with respect to the proposed molecular mechanism and divergence within the crotonase superfamily.

**EXPERIMENTAL PROCEDURES**

**Protein Production, Purification, and Crystallization**—The chromosomally encoded boxc gene from B. xenovorans LB400 was cloned into pET-28a(+) (Novagen) in frame with an N-terminal hexahistidine tag. Sequence analysis confirmed that no mutations were introduced during cloning. Native (18) and selenomethionine BoxCC were produced and purified using nickel affinity and size exclusion chromatography. Crystals of native BoxCC were obtained using the vapor diffusion method in 25% polyethylene glycol 3350 and 100 mM Tris, pH 8.5 (18). Crystals of selenomethionine BoxCC were initially obtained by seeding with native BoxCC crystals, followed by two additional rounds of microseeding to obtain diffraction quality crystals.

**Data Collection, Structure Solution, and Refinement**—Diffraction data for native BoxCC crystals were collected as described previously (18). Diffraction data for selenomethionine BoxCC were collected on beamline X8C at the National Synchrotron Light Source (Brookhaven National Laboratories) at the optimized wavelength of 0.9794 Å for the f’ selenium edge. Data processing was carried out using Crystal Clear/d’trek (19). A total of 12 selenium sites (six from each monomer) were identified and refined using autoSHARP (20), resulting in a figure of merit of 0.297. High quality phases were obtained following density modification and 2-fold NCS averaging that enabled building and registering of ~70% of the backbone using ARP/Warp (21). The remaining structure was built manually, and solvent atoms were selected using COOT (22) and refined with REFMAC (23) to an Rcryst of 18.7% and an Rmerge of 20.9%. In total, 174,578 reflections were used in refinement selected with a σ cut-off of 2.0. All solvent atoms were inspected manually before deposition. Stereochemical analysis of the refined BoxCC structure was performed with PROCHECK (24) and SFCHECK in CCP4 (25), with the Ramachandran plot showing excellent stereochemistry with 99.8% of the residues in the most favored and additional allowed conformations and no residues modeled in disallowed orientations. Overall, 5% of the reflections were set aside for calculation of Rmerge. Data collection statistics are presented in Table 1.

**Isothermal Titration Calorimetry (ITC)**—ITC was performed using a VP™ isothermal titration calorimeter (MicroCal, Northampton, MA). All samples were characterized in 20

| TABLE 1
| Data collection and refinement statistics |
|-----------------------------------------|--------------------------------------------------|
|                                       | Native BoxCC                                      | Selenomethionine BoxCC                          |
| Data collection                        | Space group                                       | P2,2,2                                         | P2,2,2                                      |
|                                       | a, b, c (Å)                                       | 85.16, 99.85, 136.73                            | 85.0, 98.5, 136.4                          |
|                                       | Resolution (Å)                                   | 34.18-1.50                                     | 46.33-2.5                                  |
|                                       | Measured reflections                             | 1,181,949                                      | 284,672                                    |
|                                       | Unique reflections                               | 183,917                                        | 30,319                                     |
|                                       | Reduction                                        | 6.43 (4.13)                                    | 7.06 (6.86)                                |
|                                       | Completeness (%)                                 | 98.8 (88.7)                                    | 99.9 (99.9)                                |
|                                       | I/σ(l)                                          | 13.4 (3.2)                                     | 11.9 (7.2)                                 |
|                                       | Rmerge (%)                                       | 0.062 (0.358)                                  | 0.101 (0.181)                              |

| Refinement statistics                  | Resolution range (Å)                             | 33.98-1.5 (1.54-1.50)                          |
|                                       | Rcryst (%)                                       | 0.187 (0.426)                                  |
|                                       | Rmerge (%)                                       | 0.207 (0.452)                                  |
| No. of atoms                          | Protein                                          | 426                                             |
|                                       | Solvent                                          | 1312                                            |
|                                       | β-Me                                             | 16                                              |
|                                       | Glycerol                                         | 24                                              |
| B-values                              | Protein (Å²)                                     | 14.53                                           |
|                                       | Solvent (Å²)                                     | 24.44                                          |
|                                       | β-Me (Å²)                                        | 20.49                                          |
|                                       | Glycerol (Å²)                                    | 26.90                                          |
| r.m.s. deviation from ideality       | Bond lengths (Å)                                 | 0.010                                          |
|                                       | Bond angles (degrees)                            | 1.303                                          |
| Ramachandran statistics              | Most favored                                     | 91.8%                                          |
|                                       | Additional allowed                               | 8.0%                                           |
|                                       | Generously allowed                               | 0.2%                                           |
|                                       | Disallowed                                       | 0.0%                                           |

* Rcryst = \( \frac{\sum I - \langle I \rangle}{\sum I} \), where \( I \) is the intensity of unique reflection hkl and \( \langle I \rangle \) is the average over symmetry-related observation of unique reflection hkl.
* Rmerge = \( \frac{\sum |I-I'|}{\sum I} \), where \( I_1 \) and \( I_2 \) are the observed and the calculated structure factors, respectively.

αRcryst, Rmerge are computed using 5% of reflections randomly chosen and omitted from refinement.
mm Tris buffer, pH 8.5, supplemented with 150 mm NaCl. Protein and ligand solutions were filtered and degassed immediately prior to use. Titrations were performed by injecting 20-μl aliquots of ligand solution into the ITC sample cell containing 200 μM BoxCC at 22 °C. All ITC data were corrected for the heat of dilution of the titrant by subtracting the heats generated by titrating the ligand into buffer alone. The equilibrium association constant and the stoichiometry were determined by curve fitting. Two independent titration experiments were performed per ligand, and the average was taken. Thermodynamic parameters were calculated from the Gibbs free energy equation, \( \Delta G = -RT \ln K_x = \Delta H - T\Delta S \).

**Bioinformatics and Molecular Modeling**—Multiple sequence alignments and the associated neighbor joining trees were determined by ClustalW (26), using the method of Saitou and Nei (27). Buried surface area for the BoxCC dimer interface was calculated using the Protein-Protein interaction analysis server (available on the World Wide Web). Docking was performed with the program Molegro virtual docker (28), using the molecular docking algorithm Moldock score. Initially, water, glycerol, and β-mercaptoethanol (β-Me) molecules were removed from the structure coordinates of BoxCC. Prior to docking, the structure of the 2,3-dihydro-2,3-dihydroxybenzoyl-CoA was built using Molpro software (30, 31) with standard force field and optimization parameters. During energy minimization, the positions of amino acid side chains were fixed by the energy of interaction protein-ligand. Complexes were inspected and compared. Positions were also inspected and corrected active site cavity prior to each docking run, and docking figures were generated with PyMol (55).

**RESULTS AND DISCUSSION**

**Overall Structure**

BoxCC crystallized as a dimer in the asymmetric unit of the primitive orthorhombic (P2₁2₁2₁) cell. The structure was solved by single anomalous wavelength dispersion, using selenomethionine-derivatized BoxCC. The final model starts at Pro\(^{10}\) (monomer A) and Ala\(^{18}\) (monomer B) and extends from Val\(^{556}\) (Fig. 1A, left). Included in the final model are six molecules of glycerol and four molecules of β-Me. It is noteworthy that in each monomer a β-Me molecule is coordinated to the side chains of Cys\(^{90}\) and Cys\(^{111}\). No β-Me is coordinated to the remaining four cysteines, suggesting that Cys\(^{90}\) and Cys\(^{111}\) are particularly reactive.

The formation of the BoxCC dimer results in an extensive buried surface area of \( \sim 3900 \AA^2 \), consistent with our observation that BoxCC elutes as a stable dimer from a size exclusion column (18). The dimeric interface of BoxCC is formed by a network of interlocking α-helices, with α6, α7, α19, and α20 contributing the majority of the buried surface area (Fig. 1B, left). Clear electron density is observed for each interface residue, including large polar residues, such as Arg\(^{191}\) and His\(^{192}\), that bridge the two monomers through an extensive solvent network that appears to increase shape complementarity of the interface (Fig. 1B, right). Nearly 62% of the residues responsible for dimer formation are conserved in BoxC orthologs, suggesting that the dimeric structure of this novel group of ring-cleaving enzymes will be consistent. In the broader context of the crotonase superfamily, the dimeric form is rare, with most members adopting trimer, tetramer, or hexamer (dimers of trimers) forms (32–40) and, in one case, proposed to form a tri-mers of dimers (41). Only the carboxyltransferase subunits for the acetyl-CoA carboxylase from *Saccharomyces cerevisiae* (42) and the α-subunit of glutacetyl-CoA decarboxylase (GCDo) from *A. fermentans* (43), which are subunits of larger multifunctional enzymes, are reported to be dimers.
Structural Characterization of BoxCC

The 2-fold symmetry of the BoxCC dimer is mirrored in each of the monomeric subunits. The N-terminal (Fig. 1C; residues 1–250) and C-terminal (Fig. 1C; residues 251–507) domains are related by 180° rotations about the vertical and horizontal axes. A central helical bundle incorporating helix a20 formed by the terminal 49 residues of the C terminus (Fig. 1C; magenta) comprises the intramolecular interface. The N and C termini share only 18% sequence identity but adopt a conserved α/β architecture with a root mean square deviation of 1.69 Å over 131 Ca atoms (Fig. 1C, right). The N-terminal domain consists of a seven-stranded twisted β-sheet sandwiched between α-helices and an additional two-stranded β-sheet positioned perpendicular to the main β-sheet. A structural comparison indicates that the β substructures surrounded by α-helices are conserved in the crotonase superfamily (44). Interestingly, the seven-strand twisted β-sheet in the C-terminal domain differs from the N-terminal domain in that it is longer and incorporates an additional antiparallel β-strand while lacking the second perpendicular β-sheet. In addition, a short helix in the N-terminal domain is substituted for a more extended helix in the C-terminal domain, resulting in a reorganized topology and surface structure.

Structural Homology

Divergence among BoxC Orthologs—To evaluate the conserved architectural features of BoxCC in the context of its closely related orthologs, all of which exhibit greater than 57% identity, we mapped the sequence alignment results onto the Connolly surface (45) calculated for BoxCC. A striking demarcation is observed between regions of high (50–61%; magenta) and low (15–20%; gray) sequence identity. It is clear from this analysis that maximum sequence divergence is localized to the C-terminal domain with the exception of helix α18 (Fig. 2, arrow), which is composed of conserved acidic residues (Asp477, Asp480, Glu481, Glu487, and Glu488). In the context of the BoxCC dimer, this negatively charged helix is solvent-exposed, suggesting a potentially important functional role. The majority of the conserved residues map to the dimer interface, contributing 81% of the overall buried surface area. The majority of the divergent region, however, is distal to the dimer interface. We hypothesize that this divergent region serves an ancillary role by providing a structural scaffold for the N-terminal domain and the dimerization interface. In this regard, it could thus also contribute indirectly in forming a docking surface for mediating a higher order complex with BoxA/B, as suggested previously (46). A homologous docking site on the dimeric structure has also been hypothesized for the GCDα subunit (43).

A Unique Member within the Crotonase Superfamily—Members of the crotonase superfamily exhibit as little as 20% sequence identity but incorporate conserved structural hallmarks. To determine the evolutionary relationship between BoxCC and each mechanistic class of the crotonase superfamily (17), we generated phylogenetic trees using the entire sequence of BoxCC (residues 1–556) and the individual N- and C-terminal domains (Fig. 3). BoxCC is approximately double the size of most members of the crotonase superfamily with the exception of the dioxygenase (47) and the carboxyltransferase subunit of biotin-dependent carboxylases (42, 43, 48). The carboxyltransferase subunits, however, have been omitted from the alignment, since they represent subunits of larger multifunctional enzymes. The N-terminal region (residues 50–250) of BoxCC shows highest similarity with the enoyl-CoA hydratases/isomerases, whereas the C-terminal region (residues 360–475) shares homology with 1,4-dihydroxynaphthyl-CoA synthase and an enoyl-CoA hydratase. Residues 260–350 show no significant identity to any particular class, suggesting that this region is structurally and/or functionally specific to BoxCC. It is interesting to note, however, that this region lies in the most divergent stretch within the BoxCC orthologs (Fig. 2), consistent with a role as a support scaffold. A similar scenario was recently observed in dihydroxyphenylglyoxylate synthase, where the initial one-third of the sequence (N-terminal) shares no sequence homology to any known proteins (47). With the active site localized in the C termini of dihydroxyphenylglyoxylate synthase, this novel region has not been ascribed any specific role. Overall, the phylogram clearly demonstrates that the proteins analyzed, BoxCC, is the most divergent member of the crotonase superfamily (Fig. 3A). Furthermore, the C-terminal domain of BoxCC is significantly more divergent than the N-terminal domain (Figs. 3, B and C), displaying a similar pattern to that observed for the BoxCC orthologs (Fig. 2).

Identifying and Mapping the Active Site

A Conserved Structural Scaffold—In the absence of the BoxCC co-structure, we used structural overlays with members of the crotonase superfamily for which active sites have been structurally characterized to define the location of the active site (Fig. 4, A–H, black arrows). Despite the mechanistically divergent nature of these enzymes, it is clear that members of the crotonase superfamily share a common catalytic scaffold...
Structural Characterization of BoxCC

(49). Based on these overlays, the active site in BoxCC maps to the N-terminal domain and is completely encompassed by a single monomer. The strict localization of the active site to a single monomer is rare in the crotonase superfamily (33, 50), with most members presenting an active site that spans a multimeric interface (34, 35, 38, 40, 51–53). The calculated volume of the predicted active site is ∼300 Å³ and forms an 18-Å-deep tunnel, consistent with the ability to coordinate the extended diphenolic dihydrodiol substrate. An open pocket positioned at the bottom of the active site tunnel appears sufficiently large to accommodate the linear aliphatic chain that results from cleavage of the dihydroxyxylated ring.

To validate the location and structure of the active site and identify potential catalytic residues, we used a molecular docking approach with the native substrate. Previous NMR studies by Gescher et al. (11) indicated that the dihydrodiol of the native BoxCC substrate adopts the cis conformation. We therefore used the cis-isomer of the dihydrodiol, including both of the possible diastereoisomers, 2S,3R and 2R,3S, in our docking scenarios. Energy minimizations confirmed that our predicted active site location resulted in the highest score with no steric clashes. A detailed analysis indicates that the lower portion of the active site (Fig. 4) is completely hydrophobic and defined by residues Ile49, Leu99 (helix α2), Phe110 (α3), Leu172 (α5), Leu174 (loop region connecting α5 and α6), and Phe528 (α20). This structural configuration provides a rationale for why BoxCC is selective for the cis-isomer 2R,3S, such that neither the hydroxyl is directed toward the hydrophobic region. In our model, helix α4 contributes structural integrity to the upper portion of the substrate binding tunnel and encodes one residue (Glu144) for the putative oxyanion hole. It is noteworthy that the structurally homologous helix α4 is referred to as the “active site helix” (35, 38), where the dipole and hydrogen bonding interactions are implicated in polarizing the thioester carbonyl of the aromatic ring substrate in 4-chlorobenzoyl-CoA dehalogenase (34).

A similar binding mode for CoA is observed in members of the crotonase superfamily. The specific residues, however, are unique to the individual homologues. In our docked model, CoA adopts a hook shape binding mode with the pantethenic acid moiety much less solvent-exposed than observed with most members of the crotonase superfamily. The key residues in BoxCC that appear to be responsible for coordinating the CoA are Lys51, Ser92, and Ser165, each of which is highly conserved within the BoxCC orthologs. Stabilizing hydrogen bonds are formed between the backbone carbonyl of Lys51 and the γO of Ser92 and the pantothenate moiety and between the γO of Ser165 and the adenine ring.

Conserved Catalytic Residues—To define catalytically important residues in the active site of BoxCC, we compared our docked model with the homologous enoyl-CoA hydratase (ECH) (35) and dienoyl-CoA isomerase (DCI) (38) (Fig. 4I). ECH is a classical member of the crotonase superfamily, and sequence comparisons indicate conservation of key catalytic residues; whereas DCI catalyzes a similar reaction to that proposed as the initial reaction in BoxCC. Interestingly, a similar arrangement of hydrophobic to hydrophilic residues is observed in ECH and DCI, where the pockets are largely hydrophobic with strategically positioned polar catalytic residues to catalyze stereo-specific reactions (35, 38). The critical conserved elements in the active site of BoxCC, however, are a pair of acidic residues that participate in acid-based catalysis in related crotonase superfamily members (35, 38). Glu146 in BoxCC adopts a conserved spatial orientation with Glu144 in ECH and Asp176 in DCI, whereas Glu168 in BoxCC superimposes with Glu164 in ECH and Glu191 in DCI (Fig. 4I). The proximity of Cys111 to the modeled substrate provides a powerful nucleophile that may be involved in catalysis and also rationalizes why we were unable to obtain a co-crystal structure with BoxCC. As shown in Fig. 1A, Cys111 is covalently modified by the...
required crystallization additive, β-Me, thereby preventing proper coordination with the substrate analog.

Ligand Binding Studies

ITC—Both the substrate analog benzoyl-CoA and CoA alone were used to complement the structural studies and validate the molecular docking solution. We were unable to use the native substrate (2,3-dihydro-2,3-dihydroxybenzoyl-CoA/benzoyl-CoA dihydrodiol), since it is commercially unavailable, and the chemical synthesis requires a complex biotransformation step to produce the cis form of the dihydrodiol.

Benzoyl-CoA bound to BoxCC with a $K_{d}$ of 116.4 μM ± 7 μM (Fig. 5). The stoichiometry was determined to be 0.90 ± 0.0, consistent with the presence of one active site per monomer as predicted from our structural overlays. In the only other study of a BoxC family member, the $K_{m}$ for the enzymatically synthesized native substrate was determined to be 17 ± 2 μM (11). Overall, the binding of benzoyl-CoA to BoxCC is enthalpically driven with an accompanying favorable change in entropy. The combination of a relatively small enthalpy change with a favorable entropic contribution suggests that the hydrophobic tail of benzoyl-CoA is the primary source of interactions with BoxCC, consistent with our molecular docking solution. It should be noted that the thermodynamic profile of the native substrate, which incorporates a dihydroxylated benzoyl ring (Fig. 5B, inset), will display an increased enthalpy due to the additional hydroxyl groups available for coordination and will bind with a lower $K_{d}$. We also determined that CoA by itself interacts with BoxCC, with nearly 25-fold lower affinity (~3 mM; data not shown). These results suggest a limited role for CoA in substrate recognition and are consistent with previous studies of a BoxCC ortholog, where acetocetyl-CoA, which is a potential

FIGURE 4. Structural overlays of the BoxCC monomer (gray) with members of the crotonase superfamily for which active sites have been localized. A, 4-chlorobenzoyl-CoA dehalogenase (Protein Data Bank code 1NZY) colored in blue; B, ECH (Protein Data Bank code 2DUB) colored in pink; C, 1,4-dihydroxy-2-naphthoyl-CoA synthase (Protein Data Bank code 2UZF) colored in teal; D, DCl (Protein Data Bank code 1DCI) colored in wheat; E, dihydroxyphenyglyoxylate synthase (Protein Data Bank code 2PG8) colored in pale green; F, carboxymethylproline synthase (Protein Data Bank code 2A81) colored in purple; G, hydroxylcinnamoyl-CoA hydratase-lyase (Protein Data Bank code 2VSS) colored in deep salmon; H, methylmalonyl-CoA decarboxylase (Protein Data Bank code 1EF8) colored in cyan. The root mean square deviation values are 1.62 Å with 4-chlorobenzoyl-CoA dehalogenase, 1.35 Å with ECH, 1.8 Å with 1,4-dihydroxy-2-naphthoyl-CoA synthase, 2.03 Å with DCl, 1.82 Å with dihydroxyphenyglyoxylate synthase, 1.5 Å with carboxymethylproline synthase, 1.41 Å with hydroxylcinnamoyl-CoA hydratase-lyase, and 1.38 Å with methylmalonyl-CoA decarboxylase. I, molecular docking solution with 2,3-dihydro-2,3-dihydroxybenzoyl-CoA. Left, side view; right, end-on view of the docked substrate. In the end-on view, the CoA portion of the substrate that would normally be directed into the page has been removed to simplify the figure, resulting in an unobstructed view of the proposed catalytic residues. Residues from BoxCC are shown in gray, whereas the conserved basic residues from ECH and DCl are shown in pink and wheat, respectively.
inhibitor of ECH, did not inhibit enzyme activity (11). Additionally, the crotonase superfamily member 2-ketocyclohexane-carboxyl-CoA hydrolase was shown to be unreactive with free 2-ketocyclohexane-carboxyl-CoA hydrolase, acetoacetyl-CoA, cyclohex-1-enecarboxyl-CoA, and 2-hydroxycyclohexane carboxyl-CoA (54), indicating that, although CoA is important for catalysis, it plays only a minor role in mediating enzyme substrate binding.

Proposed Catalytic Mechanism

It has been proposed that BoxC catalyzes both an isomerization (38) and hydrolytic deformylation (54) reaction in converting the dihydroxylated ring into the linear aliphatic chain (11). Intriguingly, both reactions are catalyzed by members of the crotonase superfamily and are CoA-dependent.

By incorporating NMR studies carried out by Fuchs and co-workers (11) with our high resolution crystal structure and modeling data, we propose a catalytic mechanism with putative functions ascribed to specific active site residues (Fig. 6). In the initial catalytic step of BoxC, the dihydroxylated ring of the substrate is deprotonated at either the C-2 or C-3 position. Hydrogen bonding interactions involving these hydroxyl groups will play an important role in biasing the initial attack. We propose that the first step involves deprotonation of the hydroxyl at C-2 by the conserved Glu^{146} (Fig. 6, step I). A nearby arginine, Arg^{118}, probably facilitates this step by lowering the pK_a of Glu^{146} to yield a stronger base. Alternatively, it is conceivable that deprotonation is initiated at the C-3 hydroxyl by Cys^{111}. In this scenario, Cys^{111} probably proceeds through an activated water that would be required to bridge the ∼3.5 Å distance between the thiol group and the C-3 hydroxyl of the substrate. By comparison, the OE1 group of Glu^{146} is positioned ∼2.7 Å from the C-2 hydroxyl, enabling direct deprotonation of the substrate. The initial deprotonation by Glu^{146} results in an oxanion intermediate on the carbonyl oxygen of the thioester bond that is stabilized by an oxanion hole formed by the backbone amino groups of Ala^{94} and Gly^{143}. It is noteworthy that both of these residues share structural equivalents with many members of the superfamily where the oxanion hole is a required mechanistic feature (17).

In our model, the deprotonation at the C-2 hydroxyl is followed by a second deprotonation at the C-3 hydroxyl (step II). Although it is conceivable that step II is catalyzed by the second conserved glutamate, Glu^{168}, we propose that a proton shuttle between the pair of glutamates (Glu^{146} and Glu^{168}) resets Glu^{146}, allowing it to act as a base in the second deprotonation event. Following each deprotonation event, Glu^{168} delivers the proton back to the substrate to enable restructuring of the double bonds and resetting the glutamate proton shuttle. In the event the initial deprotonation is mediated by Cys^{111} at the C-3
hydroxyl, the second deprotonation at the C-2 hydroxyl would probably be carried out by Glu\textsuperscript{146}.

Following the deprotonation events in steps I and II, the resulting aldehyde groups of Compound A will exist in equilibrium with water, as shown in step III. For the sake of clarity, we have only shown one of the equilibrium products (Compound B), which ultimately undergoes a third and final deprotonation, resulting in release of the formyl group (HCOOH) (step IV).

The scenario depicted in Fig. 6 involves Cys\textsuperscript{111} as mediating the final deprotonation event. As discussed previously, the distance of Cys\textsuperscript{111} to the modeled substrate suggests that it functions through a catalytic water, which, in the activated state, abstracts the proton from one of the C-2 hydroxyls. The flexible side chain of a nearby lysine (Lys\textsuperscript{107}) may lower the p\textsubscript{Ka} of Cys\textsuperscript{111}, facilitating its role as a general base. The incorporation of an activated water molecule in our proposed catalytic mechanism is consistent with the chemical requirement to yield formic acid, as suggested previously (11). If the initial deprotonation occurs at the C-3 via Cys\textsuperscript{111}, then this third deprotonation step is likely to be mediated by Glu\textsuperscript{168}. In a similar fashion to step I, the oxyanion intermediate formed in step IV will be stabilized by the oxyanion hole. Upon the addition of a proton, Compound C decomposes (step V) to Compound D (3,4-dehydrodipyl-CoA), as indicated from previous NMR experiments (11).

The pair of conserved glutamates (Glu\textsuperscript{146} and Glu\textsuperscript{168}) along with Cys\textsuperscript{111} probably represent the key catalytic residues in the active site of BoxC\textsubscript{C}. The incorporation of an active site cysteine in the crotonase superfamily is rare, since it has only been described for carboxymethylproline synthase, although recent studies have assigned the catalytic role to a histidine (52). BoxC\textsubscript{C} also offers an array of polar residues, including Tyr\textsuperscript{55}, Thr\textsuperscript{114}, Thr\textsuperscript{178}, and Gin\textsuperscript{524}, that may prove essential in stabilizing and coordinating reaction intermediates. Site-directed mutagenesis combined with biochemical characterization will ultimately be required to explicitly define the roles of the individual active site residues.

**Conclusion**

The 1.5 Å resolution crystal structure of the novel ring-cleaving enzyme BoxC\textsubscript{C} reveals an intriguing structural divergence and establishes it as a unique member of the crotonase superfamily. By complementing the high resolution structural data with ITC and molecular docking, we are able to propose catalytic roles for specific active site residues. These data extend the initial work of Fuchs and co-workers (11) in proposing a detailed molecular mechanism for BoxC and of Schofield and co-workers (17) in identifying BoxC as a promising target for structural and mechanistic elucidation in the broader context of the crotonase superfamily.

**Acknowledgments**—We are grateful to the National Synchrotron Light Source staff at beamline X8C and to A. Boraston and L. Fickoblean for assistance with the calorimetry.

**REFERENCES**

1. Gibson, J. S., and Harwood, C. (2002) *Annu. Rev. Microbiol.* **56**, 345–369
2. Bugg, T. D. (2003) *Tetrahedron* **59**, 7075–7101
3. Harwood, C. S., and Parales, R. E. (1996) *Annu. Rev. Microbiol.* **50**, 553–590
4. Stanier, R. Y., and Ornston, L. N. (1973) *Adv. Microb. Physiol.* **9**, 89–151
5. Evans, W. C. (1977) *Nature* **270**, 17–22
6. Evans, W. C., and Fuchs, G. (1988) *Annu. Rev. Microbiol.* **42**, 289–317
7. Harwood, C. S., and Gibson, J. (1986) *J. Bacteriol.* **165**, 504–509
8. Denef, V. J., Klappenbach, J. A., Patrauchan, M. A., Florizone, C., Rodriguez, J. L., Tsai, T. V., Verstraete, W., Eltis, L. D., and Tiedje, J. M. (2006) *Appl. Environ. Microbiol.* **72**, 585–595
9. Denef, V. J., Park, J., Tsai, T. V., Rouillard, J. M., Zhang, H., Wibbenmeyer, J. A., Verstraete, W., Gulari, E., Hashsham, S. A., and Tiedje, J. M. (2004) *Appl. Environ. Microbiol.* **70**, 4961–4970
10. Denef, V. J., Patrauchan, M. A., Florizone, C., Park, J., Tsai, T. V., Verstraete, W., Eltis, L. D., and Tiedje, J. M. (2005) *J. Bacteriol.* **187**, 7996–8005
11. Gescher, J., Eisenreich, W., Wörth, J., Bacher, A., and Fuchs, G. (2005) *Mol. Microbiol.* **56**, 1586–1600
12. Gescher, J., Ismail, W., Olgeschläger, E., Eisenreich, W., Wörth, J., and Fuchs, G. (2006) *J. Bacteriol.* **188**, 2919–2927
13. Gescher, J., Zaar, A., Mohamed, M., Schägger, H., and Fuchs, G. (2002) *J. Bacteriol.* **184**, 6301–6315
14. Bains, J., and Boulanger, M. J. (2007) *J. Mol. Biol.* **373**, 965–977
15. Bains, J., and Boulanger, M. J. (2008) *J. Mol. Biol.* **379**, 597–608
