Three-Dimensional Structure of the L-Threonine-*O*-3-phosphate Decarboxylase (CobD) Enzyme from *Salmonella enterica*[†,‡]

Cheom-Gil Cheong,[§] Cary B. Bauer,[§] Kevin R. Brushaber,[||] Jorge C. Escalante-Semerena,*[,||] and Ivan Rayment*[,§]

Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, and Department of Bacteriology, University of Wisconsin, Madison, Wisconsin 53706

Received December 3, 2001; Revised Manuscript Received February 11, 2002

ABSTRACT: The three-dimensional structure of the pyridoxal 5′-phosphate (PLP)-dependent L-threonine-*O*-3-phosphate decarboxylase (CobD) from *Salmonella enterica* is described here. This enzyme is responsible for synthesizing (*R*)-1-amino-2-propanol phosphate which is the precursor for the linkage between the nucleotide loop and the corrin ring in cobalamin. The molecule is a molecular dimer where each subunit consists of a large and small domain. Overall the protein is very similar to the members of the family of aspartate aminotransferases. Indeed, the arrangement of the ligands surrounding the cofactor and putative substrate binding site are remarkably close to that observed in histidinol phosphate aminotransferase, which suggests that this latter enzyme might have been its progenitor. The only significant differences in structure occur at the N-terminus, which is approximately 12 residues shorter in CobD and does not form the same type of interdomain interaction common to other aminotransferases. CobD is unusual since within the aspartate aminotransferase subfamily of PLP-dependent enzymes the chemical transformations are substantially conserved, where the only exceptions are 1-aminocyclopropane-1-carboxylate synthase and CobD. Although there are a large number of PLP-dependent amino acid decarboxylases, these are generally larger and structurally distinct from the members of the aspartate aminotransferase subfamily of enzymes. The structure of CobD suggests that the chemical fate of the external aldimine can be redirected by modifications at the N-terminus of the protein. This study provides insight into the evolutionary history of the cobalamin biosynthetic pathway and raises the question of why most PLP-dependent decarboxylases are considerably larger enzymes.

The availability of numerous complete sequences for a wide range of organisms coupled with an increasing number of protein structures provides an opportunity to examine the footprints left by the evolution of complete biosynthetic pathways. This is possible because even though the sequences may diverge beyond recognition the signature of the protein fold remains intact. Thus, with sufficient structures and sequences, it should be possible to retrace the passage of time back to the primordial protein folds. Such studies are of particular interest in those cases where the biosynthetic pathway leads to a molecule that arose early in the development of life, since these should provide a retrospective on the protein folds that were available at that time. One such product of a particularly complex biosynthetic pathway is cobalamin which is one of nature's most intricate cofactors.

Cobalamin is an essential cofactor for many living organisms, although it is only synthesized de novo by bacteria and archaea (Figure 1). The existence of similar genes in both bacteria and archaea for many of the cobalamin biosynthetic enzymes suggests that this pathway evolved prior to the divergence of these domains. All other organisms that require cobalamin acquire it as a nutrient. Surprisingly, even though this cofactor is essential for survival, it is only found in relatively few enzymes, the best known of which include ribonucleotide reductase, methionine synthase, and methylmalonyl CoA mutase. Clearly, cobalamin has been retained during the evolution of life because of the biosynthetic advantage provided by the labile cobalt−carbon bond that lies at the center of this cofactor. Much of what is known about the biosynthetic pathway for cobalamin has been derived from the study of *Paracoccus denitrificans*, *Paracoccus freudenreichii*, and *Salmonella enterica* (*1*−*5*).

The complete biosynthetic pathway for cobalamin utilizes a shared pathway up to uro′genIII that is common to all tetrapyrroles, including hemes and chlorophylls (*1*). Thereafter, at least 24 additional dedicated genes are required. Thus, a substantial part of the genome for simple organisms, which typically contain ∼4000 genes, is devoted to synthesis of this cofactor. This pathway is typically divided into four sections: (i) synthesis of the corrin ring, (ii) attachment of the upper 5′-deoxyadenosine ligand to the cobalt ion, (iii) synthesis of the lower 5,6-dimethylbenzimidazole ligand, and

[†] This research was supported in part by NIH Grants GM58281 to I.R. and GM40313 to J.C.E.-S. Use of the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Basic Energy Research, under Contract W-31-109-ENG-38.

[‡] The X-ray coordinates have been deposited in the Protein Data Bank as entry 1KUS.

* To whom correspondence should be addressed. I.R.: Department of Biochemistry, 433 Babcock Dr., Madison, WI 53706; phone, (608) 262-0437; fax, (608) 262-1319; e-mail, Ivan_Rayment@biochem.wisc.edu. J.C.E.-S.: Department of Bacteriology, Fred Hall, Linden Drive, Madison, WI 53706; e-mail, escalante@bact.wisc.edu.

[§] Department of Biochemistry.

[||] Department of Bacteriology.

10.1021/bi012111w CCC: $22.00
Published on Web 03/22/2002

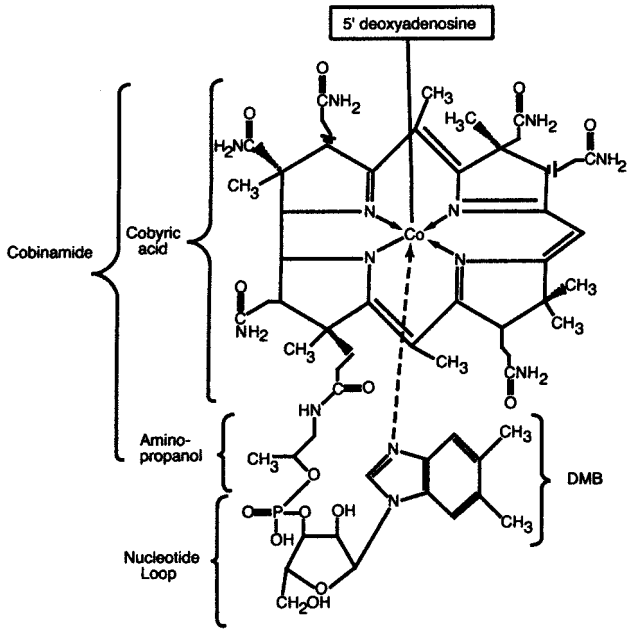

FIGURE 1: Chemical structure of 5′-deoxyadenosylcobalamin. In *S. enterica*, the lower ligand is 5,6-dimethylbenzimidazole (*36*).

(iv) assembly of the nucleotide loop. A schematic for the final stages of cobalamin biosynthesis in *S. enterica* is shown in Figure 2.

The final step in corrin ring biosynthesis is the attachment of an (*R*)-1-amino-2-propanol phosphate group that serves as a linker between the nucleotide loop and the corrin ring (Figure 1). This component is synthesized in *S. enterica* by decarboxylation of L-threonine phosphate by the *cobD* gene product which has a molecular mass of 40.8 kDa and is composed of 364 amino acid residues (*6*). Interestingly, this enzyme was originally identified as a member of the pyridoxal 5′-phosphate-dependent family of amino acid aminotransferases by sequence comparison (*7*). Indeed, the sequence of this enzyme is significantly similar with the sequences of both histidinol-phosphate aminotransferase and aspartate aminotransferase. The sequence of the latter is the most extensively studied PLP-dependent enzyme and the one whose structure was the first known for this class of enzymes (*8*, *9*).

Pyridoxal 5′-phosphate-dependent enzymes are abundant and catalyze a wide range of reactions that include racemization, amino group transfer, β-elimination in deaminases, and decarboxylation to name a few; thus, it is not surprising to find a PLP-dependent enzyme involved in a decarboxylation reaction in the cobalamin biosynthetic pathway. PLP-dependent enzymes have been classified into four groups on the basis of sequence comparison and are defined as the α-family, the β-family, the D-alanine aminotransferase family, and the alanine racemase family (*10*). The α-family and β-family have also been denoted as the aspartate aminotransferase and tryptophan synthase β-family because of their prototype enzymes. The folds for these enzymes are unrelated (*11*), yet the underlying chemistry is similar in all cases. All these enzymes utilize the PLP group to form a transient Schiff base between the aldehyde moiety and amino group of the substrate which serves to activate the substrate. This is known as the external aldimine. There are many fates for the Schiff base which include removal of the α-hydrogen to form an enolate anion, decarboxylation of a β-oxoacid, and aldol cleavage. The loss of the α-hydrogen leads to a quinonoid−carbanionic intermediate which can form the basis for racemization, cyclization, β-elimination, and transamination reactions, the latter of which plays a key role in amino acid metabolism.

The α-family constitutes the largest group of PLP-dependent enzymes where the name implies that the chemical transformations occur at the α-carbon of the substrate. It is a diverse group that includes a very large number of aminotransferases together with representative decarboxylases and lyases. These enzymes all contain a characteristic PLP binding domain, though their overall size varies considerably and their level of sequence similarity is exceedingly low. Careful sequence analysis has further divided the α-family into at least eight subgroups which include three subfamilies of aminotransferases, three subfamilies of amino acid decarboxylases, the CoA-dependent acyltransferases, and the γ-subfamily, members of which catalyze reactions at the γ-carbon even though they evolutionarily belong to the α-family (*10*, *12*). Within the α-family, the size of the proteins varies considerably, even though structural studies have shown that they share a common PLP binding domain. For example, the prototypical aspartate aminotransferase of the α-family typically contains approximately 400 amino acid residues, whereas the amino acid decarboxylases vary from 400 to 730 for ornithine decarboxylase from *Lactobacillus*.

Although there are a remarkable number of reactions catalyzed by the α-family of enzymes, within each of the aminotransferase subfamilies all of the proteins function as aminotransferases across the entire kingdom, albeit with a wide range of substrate specificity. The exceptions to this observation are 1-aminocyclopropane-1-carboxylate synthase (*13*) and CobD. Likewise, most of the amino acid decarboxylases lie within a restricted area of the evolutionary tree of the α-family (*10*). From an evolutionary point of view, this implies that most of the divergence in chemical function occurred very early in the phylogenetic evolution of these proteins, which is remarkable since all PLP-dependent enzymes share common mechanistic features.

A fundamental structural mechanism for all PLP-dependent enzymes has been deduced and confirmed that provides an explanation for the specificity of the step that occurs after the formation of the Schiff base (*14*, *15*). This predicts that the bond that is broken lies above or below the plane of the coenzyme−imine π-system. This arrangement minimizes the energy of the transition state for bond breaking since it allows maximum $\sigma-\pi$ overlap between the bond to be broken and the π-system of the pyridine ring−imine system. This proposal was confirmed by structural studies of aspartate aminotransferase (*16*) and has been shown to be true for all structures of PLP-dependent enzymes that contain analogues of the substrate or coenzyme−substrate adducts within their active sites. Thus, as a first principle, the product of a PLP-catalyzed reaction is dictated by the orientation of the groups attached to the C4 position of the external aldimine.

Many PLP-dependent reactions occur in solution in the presence of PLP and its substrate, so it is generally accepted that a major function of PLP-dependent enzymes is to enhance one pathway for the decomposition of the external aldimine and suppress the side reactions. Consequently, evolution of a new chemical consequence, for example,

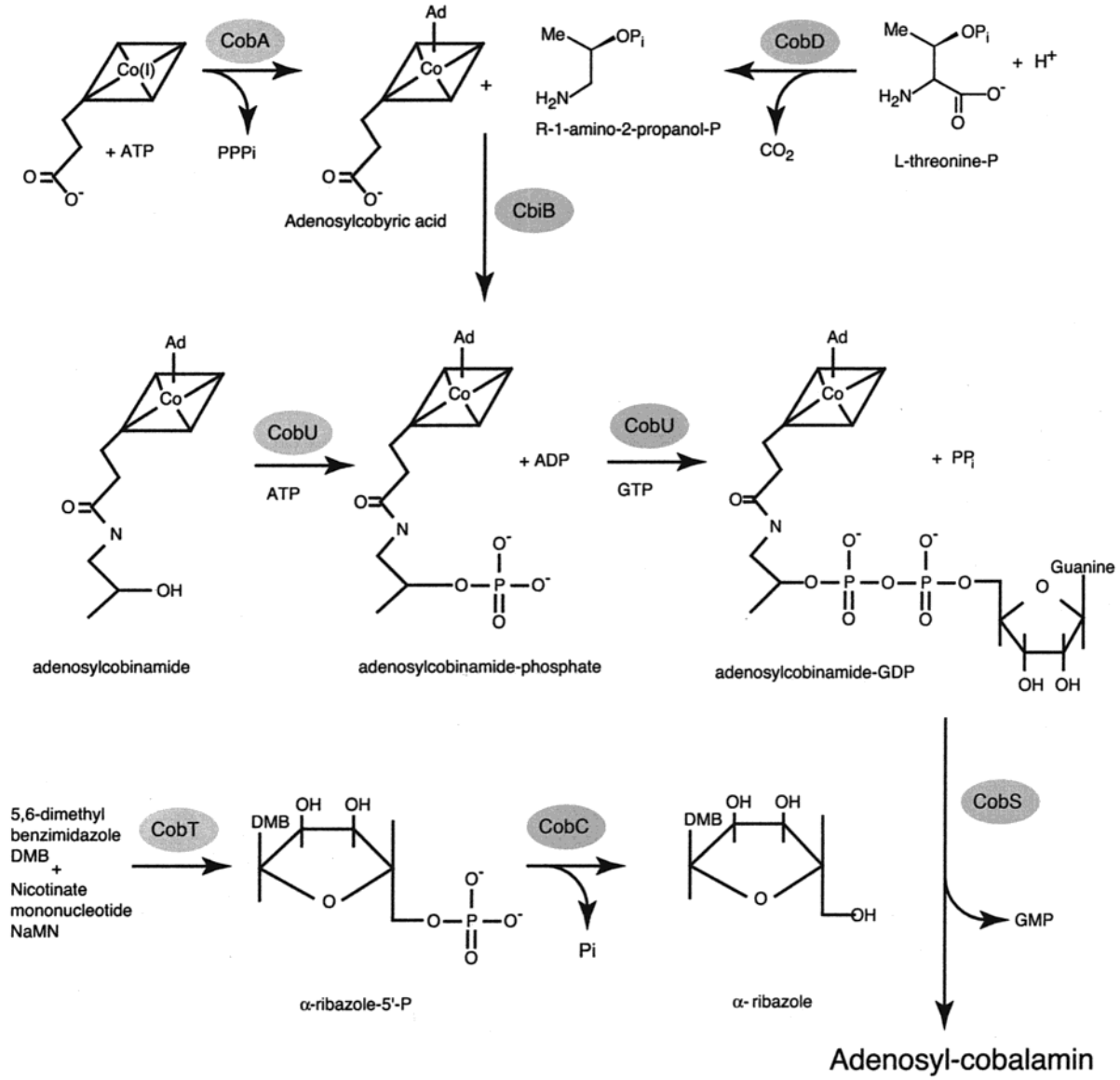

FIGURE 2: Schematic chemical representation of chemical steps involved in nucleotide loop biosynthesis and assembly in adenosylcobalamin biosynthesis in *S. enterica* together with the names of the genes responsible for these transformations. Adapted with permission from Thompson et al. (*37*).

decarboxylation instead of amino transfer, should simply require enhancement of one reaction over another. Fundamentally, this should be accomplished by small changes in the orientation and coordination of the substrate in the active site. Ironically, the observation of substantially uniform chemical consequences within each of the sequence-related subfamilies of PLP-dependent enzymes suggests that evolution of an altered orientation of the α-carbon is indeed a rare event (*10*). This is in contrast to other enzyme superfamilies where evolution of function has clearly occurred well after the phylogenetic expansion and still continues today (*17*).

L-Threonine-*O*-3-phosphate decarboxylase (CobD from *S. enterica*), as one of the outliers in the aminotransferase sequence family, poses several interesting mechanistic and evolutionary questions. First, what components of the structure allow this enzyme to function as a decarboxylase when almost all other members of this family are aminotransferases? Second, if a member of the α-family of aminotransferases can evolve to become an amino acid decarboxylase, why is this motif not used more frequently for this purpose? This latter question is of particular interest because in general the amino acid decarboxylases are twice the size of the aminotransferases. To address these questions, the structure of L-threonine-*O*-3-phosphate decarboxylase complexed with PLP to 1.8 Å resolution is reported here. This shows that the structure is remarkably similar to aspartate aminotransferase and histidinol aminotransferase and provides an example of divergent evolution within the restricted class of aminotransferases.

MATERIALS AND METHODS

Purification and Overproduction of the L-Threonine-O-3-phosphate Decarboxylase (CobD) Enzyme. The construction of the *cobD* overexpression plasmid pCOBD6 (*cobD*$^+$ *bla*$^+$ pT7-7 overexpression vector) has been described previously (*6*). Cells of strain JE4094 (*Escherichia coli* strain BL21/λDE) carrying plasmid pCOBD6 were grown in Luria-Bertani (LB) broth containing pyridoxine (0.1 mM), ampicillin (100 μg/mL), and glucose (22 mM) at 30 °C with

shaking until they reached a cell density A_{600} of ~0.6. Cultures were shifted to 15 °C, and IPTG was added to a final concentration of 0.4 mM to induce the synthesis of the T7 RNA polymerase enzyme. Cultures were maintained at 15 °C for 16 h to allow synthesis of the CobD protein. Cells were harvested by centrifugation (12000*g* at 4 °C) using a KOMPspin KA-21.50 rotor (Composite Rotor, Inc., Mountain View, CA) in a J2-21 Beckman refrigerated centrifuge (Beckman, Palo Alto, CA). Cells (3.5 g, wet weight) were resuspended in 17 mL of 50 mM sodium phosphate buffer (pH 6.8) containing DTT[1] (1 mM; buffer A), protease inhibitor phenylmethanesulfonyl fluoride (1 mM), and EDTA (1 mM). Cells were broken using a French pressure cell press operating at 8×10^4 kPa; three passes ensured a level of cell breakage of greater than 95%. Cell-free extracts were obtained after centrifugation at 42000*g* for 30 min. Approximately 30 mg of protein was obtained per liter of cell-free extract. Purification of the enzyme was monitored using SDS−PAGE (*18*) and the activity assay described elsewhere (*6*). Finely ground Ultrapure ammonium sulfate (ICN Biochemicals, Aurora, OH) was used to precipitate proteins in the cell-free extract at 4 °C. Ammonium sulfate was added to 10% saturation; precipitated material was removed by centrifugation at 12000*g* for 15 min at 4 °C.

Hydrophobic Interaction Chromatography. The supernatant from the previous step (145 mg of total protein) was loaded onto a Phenyl-Sepharose CL-4B (Sigma) column (2.5 cm × 10 cm, 50 mL) equilibrated with buffer A at a flow rate of 50 mL/h. After the sample had been loaded, the column was washed with 50 mL of buffer A. The column was developed with a 200 mL linear gradient of buffer A containing increasing amounts of ethylene glycol (from 0 to 8 M, from 0 to 50% v/v). The column was washed with an additional 50 mL of buffer A containing 8 M ethylene glycol to ensure complete elution of the CobD protein. Fractions containing the CobD protein were combined (60 mL) and dialyzed for 24 h at 4 °C against buffer A (2 L). During this period of time, buffer A was changed twice to ensure removal of ethylene glycol from the sample. The sample contained 31 mg of protein.

Anion Exchange Chromatography. The sample from the previous step (31 mg of protein) was applied onto a 25-mL, fast-flow Toyopearl TSK-DEAE 650 M column (Suppelco, Inc., Bellefonte, PA) (5 cm × 2.5 cm) equilibrated with buffer A. The column was developed at a rate of 250 mL/h. The CobD protein did not interact with the resin. Fractions containing the CobD protein were combined and concentrated using a Centriprep 10 concentrator (Amicon, Inc., Beverly, MA). After this purification step, the CobD protein was judged by densitometry to be greater than 95% homogeneous. This protocol yielded an average of 10 mg of CobD protein per liter of culture.

CobD Crystallization and Data Collection. Prior to crystallization, CobD was dialyzed overnight against 50 mM Hepes (pH 7.0), 150 mM NaCl, and 2 mM DTT at 4 °C. The protein was concentrated to 10 mg/mL and immediately frozen dropwise in liquid nitrogen and stored at −80 °C for future use. Crystals of CobD (both wild-type and selenomethionine) were grown from hanging drops at 4 °C from 1.5 M ammonium phosphate, 20 mM KCl, and 5% glycerol in 50 mM Bis Tris propane (pH 6.5). No additional PLP was added to the crystallization medium. The final protein concentration in each droplet was approximately 5 mg/mL, and the final volume of each drop was approximately 12 μL. Small crystals appeared with seeding overnight and achieved maximum dimensions 0.2 mm × 0.3 mm × 0.4 mm over a period of 1 week.

Prior to data collection, crystals of selenomethionine-substituted CobD were transferred to a synthetic mother liquor which contained 1.25 M ammonium phosphate and 450 mM KCl in 50 mM Bis Tris propane (pH 6.5) and allowed to stand at 4 °C overnight. Crystals were then transferred to a cryoprotectant which consisted of 1.25 M ammonium phosphate, 450 mM KCl, and 25% ethylene glycol in 50 mM Bis Tris propane (pH 6.5). The crystals were transferred successively to droplets containing cryoprotectant/synthetic mother liquor ratios of 30, 50, and 100% and allowed to stand for 2 min at each stage. A single crystal was picked up in a loop of surgical suture and flash-frozen in a stream of cold liquid nitrogen gas at −160 °C. CobD crystallizes in orthorhombic space group *I*222 with the following unit cell dimensions: $a = 67.96$ Å, $b = 101.55$ Å, and $c = 117.23$ Å.

Multiple anomalous dispersion (MAD) data were collected at four wavelengths on a single frozen crystal at beamline 19-ID of the Structural Biology Center at the Advanced Photon Source (Argonne, IL). For the peak and inflection point wavelengths, 145 images were recorded on a 3 × 3 tiled CCD detector using 1° oscillations and 7 s exposure times at a crystal-to-detector distance of 200 mm and a temperature of −160 °C. The remote wavelength data were collected in a similar fashion, where only 180 images were collected at each wavelength. The diffraction data were processed and scaled using the HKL 2000 software suite (*19*). The Friedel differences in the reference data set (peak) were locally scaled to remove systematic errors. The other three data sets were then scaled to the reference data set. Data collection and processing statistics are presented in Table 1.

CobD Structure Determination. The positions of five of the six selenium atoms were located using direct methods from Δ*F* data calculated from the peak wavelength data set using the program SHELXS-97. Phases were then calculated from the MAD data sets using the program CNS. Density modification (solvent flipping) was performed using CNS (*20*). This yielded an electron density map that showed continuous electron density over a large region of the unit cell. Residues 8−362 were built into the electron density with the program TURBO-FRODO (*21*). The missing residues at the N- and C-termini could not be located. The structure was refined first with TNT (*22*) and then CNS. Observation of difference electron density maps revealed strong density in the region of the protein believed to contain the PLP binding motif, specifically interacting with the protein through a covalent linkage through residue Lys^{216} (Figure 3). Refinement statistics are presented in Table 2. The final *R*-factor was 20.1% for data between 1.8 and 30 Å resolution with an R_{free} of 23.5%. Analysis of the backbone

[1] Abbreviations: DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; rms, root-mean-square; Bis Tris propane, 1,3-bis[tris-(hydroxymethyl)methylamino]propane; AdoCbl, adenosylcobalamin; AdoCbi, adenosylcobinamide; HO-CBL, hydroxycobalamin; DMB, 5,6-dimethylbenzimidazole; MES, 2-(*N*-morpholino)ethanesulfonic acid; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis.

Table 1: Overall Data Collection Statistics for MAD Data

	remote 1 (λ = 0.94645 Å)	peak (λ = 0.97915 Å)	edge (λ = 0.97926 Å)	remote 2 (λ = 1.0205 Å)	native
resolution (Å)	2.0	2.0	2.0	2.0	1.8
no. of unique reflections	27775	27798	27729	27547	37844
redundancy	6.1	3.7	3.6	5.8	6.7
completeness (%)[a]	99.7 (99.7)	99.7 (98.6)	99.6 (97.7)	99.0 (92.3)	99.6 (100)
average I/σ	44.3	55.5	53.3	42.8	37.0
R_{merge}[b]	0.046	0.042	0.035	0.041	0.069 (0.243)

[a] The numbers in parentheses represent completeness in highest-resolution shell (native data, 1.86−1.80 Å). [b] $R_{merge} = (\Sigma|I_{hkl} - I|)/(\Sigma I_{hkl})$, where the average intensity I is taken over all symmetry equivalent measurements and I_{hkl} is the measured intensity for a given reflection.

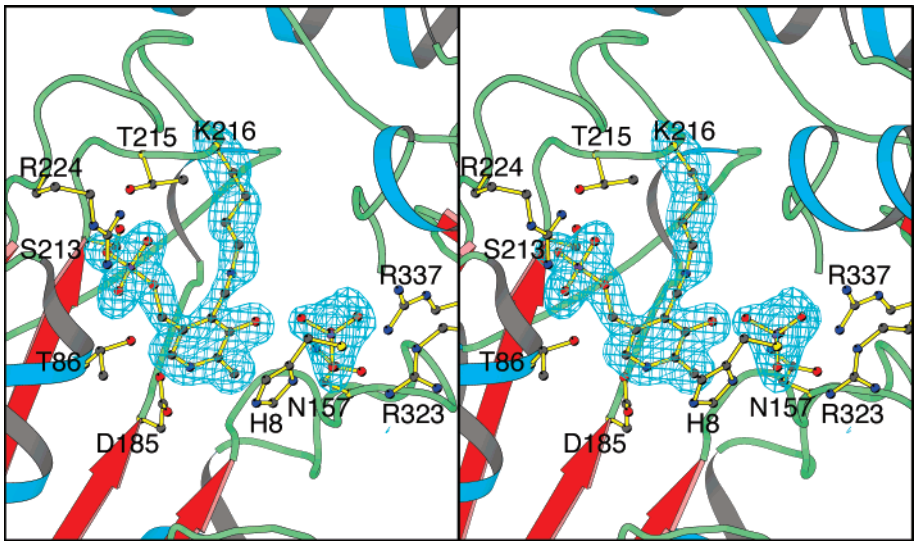

FIGURE 3: Stereoview of the representative electron density corresponding to the PLP bound to Lys[216]. The map was calculated with coefficients of the form $2F_o - F_c$. This figure was prepared with the programs Molscript and Bobscript (*38*, *39*).

Table 2: Refinement Statistics

resolution limits (Å)	30.0−1.8
final *R*-factor	20.1
R_{free}	23.5
no. of reflections (working set)	35944
no. of reflections (test set)	1899
no. of protein atoms	2738
no. of solvent molecules	219
other molecules, ions	1 PLP, 1 phosphate, 2 ethylene glycols
average *B* value	
main chain atoms	23.3
all protein atoms	25.6
solvent atoms	31.3
weighted rms deviations from ideality	
bond lengths (Å)	0.015
bond angles (deg)	1.69
planarity (trigonal) (Å)	0.003
planarity (others) (Å)	0.014
torsional angles (deg)	14.3
cis peptide bonds	Pro[155], Pro[106], Pro[158]

dihedral angles with PROCHECK (*23*) revealed that 91.1% of the residues conformed to the most favorable regions and the other 8.9% conformed to other additionally allowed regions. The three-dimensional coordinates and structure factors have been deposited in the Protein Data Bank as entry 1KUS.

RESULTS AND DISCUSSION

Structure of CobD. The final model for CobD complexes with PLP contains 355 amino acid residues out of a total of 364, including 129 water molecules, one molecule of PLP, one phosphate ion, and two molecules of ethylene glycol. The locations of the PLP and phosphate ion serve to identify the location of the active site. The polypeptide chain extends continuously from His[8] to Pro[362] without breaks. Seven residues at the N-terminus and two at the C-terminus are disordered. A section of the electron density as represented by the PLP cofactor is shown in Figure 3. A ribbon representation of the crystallographic subunit of CobD is shown in Figure 4. This reveals that the molecule consists of a large domain and a small domain. The large central domain extends from Ala[31] to Ser[263], whereas the small domain is composed of residues His[8]−Ala[31] and Leu[273]−Pro[362]. The domains are connected by a section of random coil at the N-terminal boundary, whereas a long α-helix, extending between Trp[265] and Leu[284], forms the connection between the C-terminal regions.

The small domain consists of two layers of secondary structure that is dominated by a highly curved arrangement of three β-strands from the C-terminal section of the polypeptide chain and two stretches of random coil provided by the N-terminal component (Figure 4). The polypeptide chain within this feature is mostly antiparallel. The second layer of secondary structural elements is formed by three α-helices contributed by the C-terminal section of the polypeptide chain which lie within the curved enclosure provided by the sheet. The other surface of this domain abuts the central domain.

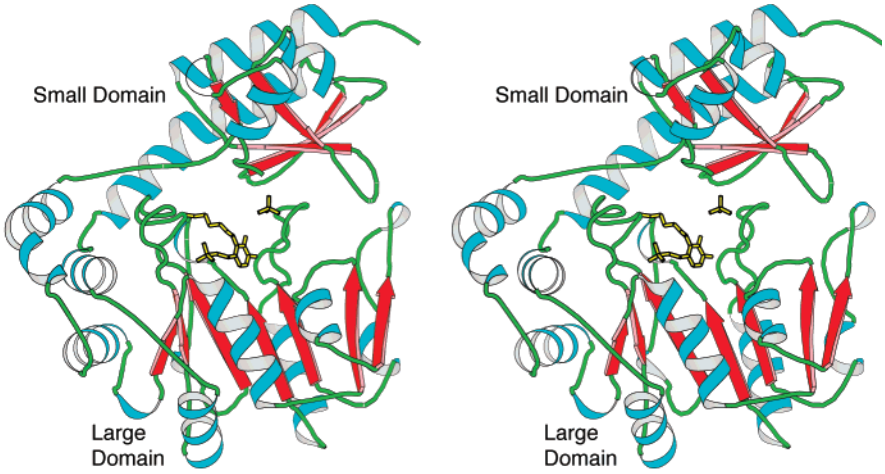

FIGURE 4: Stereoview ribbon representation of one subunit of CobD. This reveals that the protein consists of two domains. This figure was prepared with the programs Molscript and Bobscript (*38*, *39*).

FIGURE 5: Ribbon representation of the two crystallographically related subunits of CobD viewed down the 2-fold axis of symmetry where the active site is revealed by the location of PLP. One of the subunits is blue and green and the other red and orange. The active site lies at the subunit interface which is a feature common to the α-family of aminotransferases. This figure was prepared with the programs Molscript and Bobscript (*38*, *39*).

The central domain consists of a large mostly parallel *β*-sheet bounded on both sides by α-helices with strand order 3245671 where the seventh strand in the sheet lies antiparallel to the rest. This sheet is also highly curved where three helices lie on the re-entrant face of the sheet and five on the outer face (Figure 4). This domain has topology identical to that of the PLP binding domain of aspartate aminotransferase (*9*).

Quaternary Structure of CobD. CobD forms a tight dimer across a crystallographic 2-fold axis (Figure 5) where the interface buries 1764 Å^2 of surface area per subunit or 14.6% of the surface area of the monomer. This quaternary structure is common to all members of the PLP-dependent enzymes that fall into the aminotransferase fold. The interface is primarily formed by contacts between the large domains, although the N-terminal residues belonging to the small domain extend into the interface. This extension serves to cap the active site as discussed later. Significantly, the dimer interface allows both subunits to contribute to each active site.

Comparison of the Fold with Other PLP-Dependent Enzymes. The folds for PLP-dependent enzymes have been grouped into four fold families (folds I−IV), of which the family that contains aspartate aminotransferase is the largest. This fold family, also known as the α-family, has been further divided into groups that are segregated according to the chemistry performed and evolutionary relationships (*10*). A search of the protein database (*24*, *25*) with the program DALI (*26*) reveals that CobD is most similar to histidinol-phosphate aminotransferase [PDB entry 1GEW (*27*)], followed by aspartate aminotransferase from *Thermus thermophilus* [1BJW (*28*)], tyrosine aminotransferase from *Trypano-*

Table 3: Structural Similarities between Representative Members of the Aspartate Aminotransferase Family[a]

		Z score				
protein	CobD	1GEW	1BJW	1BW0	1B8G	1ORD
L-threonine-*O*-3-phosphate decarboxylase (CobD)	**364**	36.4	31.7	30.2	29.0	13.3
histidinol-phosphate aminotransferase from *E. coli* (1GEW)	1.41 (253) 2.2 (311)	**356**	34.3	32.2	31.5	17.0
aspartate aminotransferase from *T. thermophilus* (1BJW)	1.35 (220) 2.6 (304)	1.55 (216) 2.6 (321)	**382**	**46.8**	42.0	14.1
tyrosine aminotransferase from *Tr. cruzi* (1BW0)	1.52 (210) 2.9 (309)	1.51 (188) 2.9 (323)	1.33 (333) 1.9 (373)	416	**38.3**	17.7
1-aminocyclopropane-1-carboxylate synthase from *M. domestica* (1B8G)	1.56 (206) 3.0 (306)	1.66 (120) 2.9 (323)	1.45 (288) 2.4 (372)	1.42 (262) 2.5 (372)	429	**14.4**
ornithine decarboxylase from *Lactobacillus* sp. (1ORD)	1.34 (119) 3.7 (271)	1.65 (220) 3.7 (288)	1.50 (137) 4.6 (296)	1.36 (143) 3.6 (259)	1.51 (119) 5.0 (320)	731
	rms differences between the cores and overall structures					

[a] The upper-right diagonal elements give the *Z* scores between respective coordinate files as determined with the program DALI (*26*), whereas the lower-left diagonal elements give the rms differences between structurally equivalent α-carbons. The first pair of values refer to the rms difference and number of residues in the core domain and were determined with the program Align (*35*). In this case, the number equivalent residues indicates the degree of similarity. In contrast, the second pair of values reflect the topological equivalences generated by DALI. Here the increase in the rms value and the decrease in the number of equivalences are indicative of divergence. The diagonal elements give the number of amino acids in each protein.

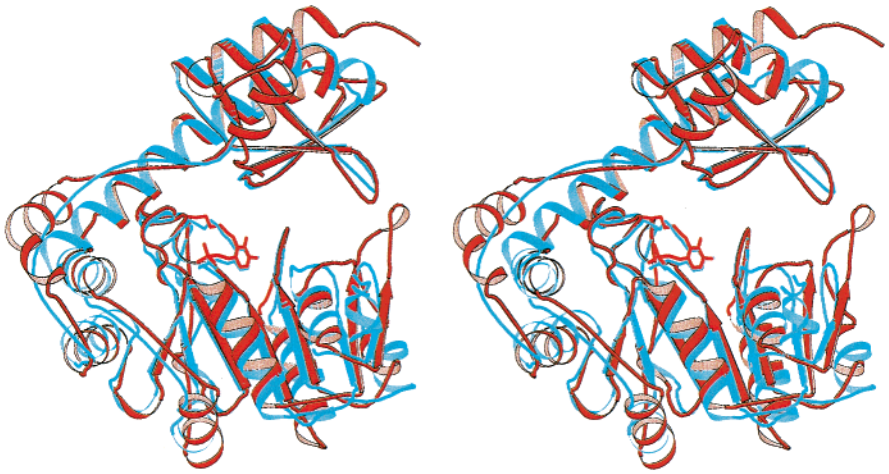

FIGURE 6: Stereo comparison of L-threonine decarboxylase with histidinol-phosphate aminotransferase [PDB entry 1GEW (*27*)]. The structures were overlapped with the program Align (*35*). This figure was prepared with the programs Molscript and Bobscript (*38*, *39*).

soma cruzi [1BW0 (*29*)], and 1-aminocyclopropane-1-carboxylate synthase from apple, *Malus domestica* [1B8G (*13*)] (Table 3). The *Z* scores for these alignments are 36.4, 31.7, 30.2, and 29.0, respectively. Significantly, the *Z* scores for the alignment of 2,2-dialkylglycine decarboxylase (*30*) and ornithine decarboxylase (*31*, *32*) with CobD are 17.8 and 13.3, respectively, which are almost half of that for the aminotransferases, indicating the closer structural similarity of CobD to the aminotransferases rather than the amino acid PLP-dependent decarboxylases. This trend is reflected in the rms differences between CobD and this group of proteins, where the closer the similarity the greater the number of residues in the common core and the smaller the rms difference for the molecules as a whole. CobD shares the greatest degree of structural similarity with histidinol-phosphate aminotransferase (Figure 6) for which 85% of the α-carbons differ by an rms of 2.2 Å. Examination of similarities between aspartate aminotransferase and tyrosine aminotransferase reveals that these aminotransferases are more similar to each other than to the remaining members of this group. Conversely, the cross-wise similarity between CobD, histidinol-phosphate aminotransferase, aspartate aminotransferase, and 1-aminocyclopropane-1-carboxylate synthase suggests that as a group they are most similar to the aspartate aminotransferase.

Active Site of CobD. The position of the active site in CobD was inferred by the location of PLP and a phosphate ion, and also by the comparison of CobD with other PLP-dependent enzymes. From the electron density shown in Figure 3, it is clear that PLP is covalently bound to Lys^{216} in the form of a Schiff base. As seen in all other members of the α-family of the aspartate aminotransferases, the active site is built from components of both subunits of the dimer. The bulk of the PLP binding site is provided by the C-terminal ends of the large β-sheet that forms the center of the large domain. The cofactor lies across the ends of the strands and is buttressed, in part, on both ends by loops that connect the strands to the outer layers of α-helices. In addition, two loops from the 2-fold related subunit (Tyr^{56}, Pro^{245}, and Trp^{246}) provide residues that close off the active site pocket. Significantly, these provide a hydrophobic plug that, together with His^{8}, excludes water from the active site.

The phosphate moiety of PLP forms an extensive network of hydrogen bonds with the binding pocket (Figure 7). Except

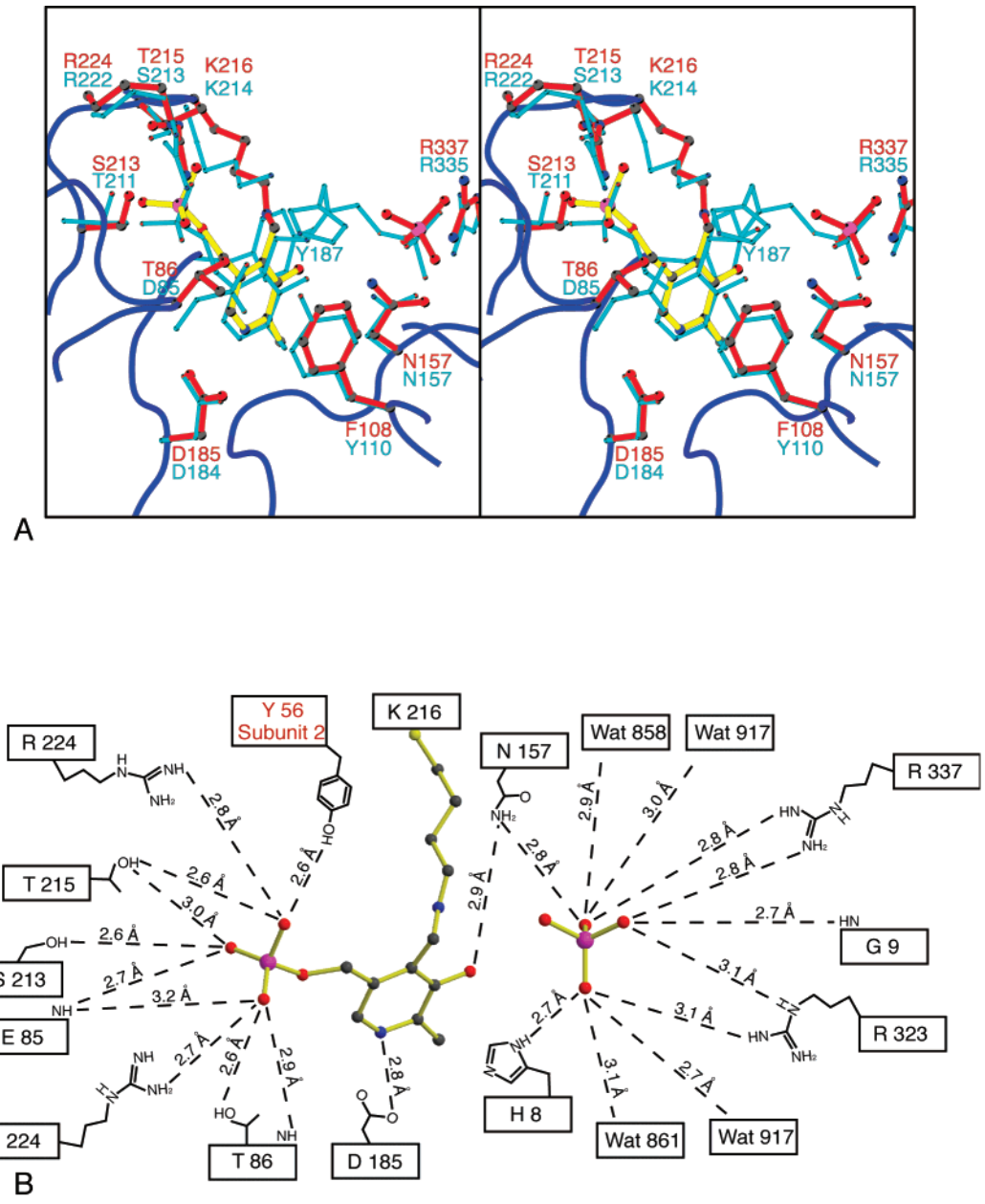

FIGURE 7: Stereo overlap of PLP in CobD and histidinol-phosphate (A) aminotransferase together with a ligand coordination diagram of the binding interactions with PLP in CobD (B). In panel A, CobD is depicted as a blue coil for the backbone atoms with red side chains, whereas histidinol-phosphate aminotransferase is colored in cyan. The amino acids in CobD and histidinol-phosphate aminotransferase are identified with red and cyan labels, respectively. The histidinol-phosphate aminotransferase shows the ketimine intermediate formed by the reaction of histidinol phosphate with PLP [PDB entry 1GEX (*27*)]. His^{8}, which interacts with the phosphate ion in CobD, has been omitted for clarity. This figure was prepared with the programs Molscript and Bobscript (*38*, *39*).

for the bridging oxygen, the hydrogen bonding potential of the phosphoryl group is completely satisfied. Most of these interactions arise from the residues and connecting loops at the ends of the β-strands at the edge of the large sheet (strands 1, 6, and 7 in the sheet, given a strand order of 3245671). In particular, the main chain amide hydrogen of Glu^{85} and side chain of Thr^{86} from the first strand contribute a cluster of interactions that coordinate two of the phosphoryl oxygens. Strand 7 contributes an important ionic interaction from Arg^{224} to two of the phosphoryl oxygens and in combination with the amide hydrogen of Glu^{85} contributes to neutralizing the charge on the phosphoryl group. Additional stabilizing hydrogen bonding interactions are provided by the side chains of Ser^{213} and Thr^{215} together with a hydrogen bond from the hydroxyl group of residue $Tyr^{56'}$ of the 2-fold related subunit.

With the exception of hydrogen bonds to the side chain of Asp^{185} and the amido moiety of Asn^{157}, there are comparatively few hydrogen bonds between the protein and the remainder of the pyridoxal moiety. In contrast, there are a significant number of hydrophobic interactions with the aromatic ring which include Phe^{108}, Ala^{187}, and Phe^{188}. The last interaction is quite close to O3 of the pyridoxal ring and serves to restrict the orientation of the cofactor by repulsion and increase the basicity of this oxygen. Together, the interactions between the phosphoryl and pyridoxal moieties serve to anchor the cofactor in the active site pocket during its catalytic cycle.

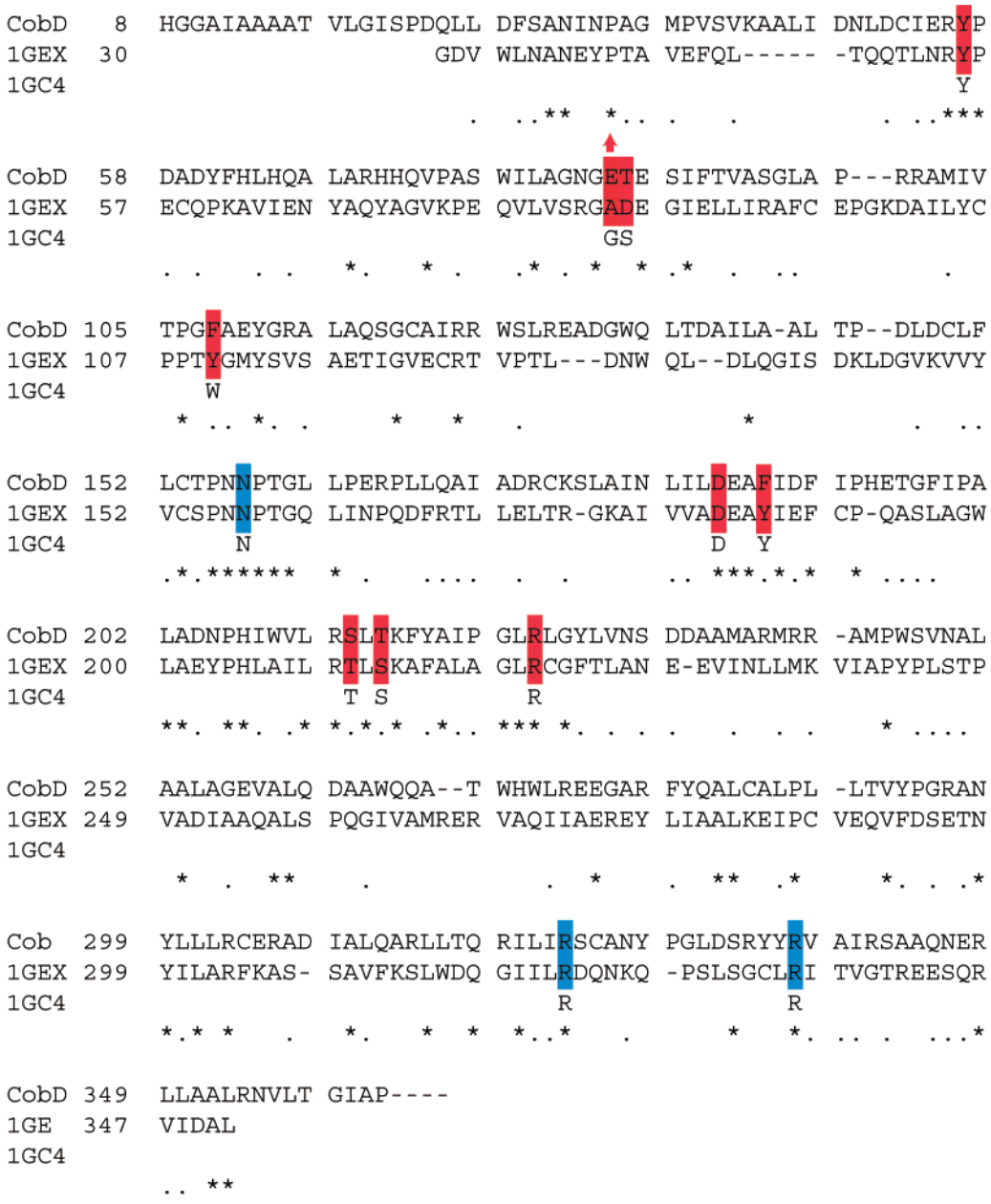

FIGURE 8: Sequence alignment of CobD and histidinol-phosphate aminotransferase (PDB entry 1GEX) based on a structural alignment of the proteins derived with the Swisspdb viewer (*40*). The residues marked in red and blue denote amino acids that interact with the PLP moiety or phosphate ion, whereas the vertical arrows denote those that interact via their main chain atoms. The third line of each set also includes the identity of the structural equivalent in aspartate aminotransferase. *T. thermophilus* complexed with PLP and aspartate (1GC4). The fourth line indicates the location of identical and similar residues in CobD and 1GEX with asterisks and periods, respectively.

The phosphate ion observed in the active site most likely lies at the location of the phosphoryl moiety of L-threonine phosphate since it binds at the same place as the equivalent group in histidinol phosphate in histidinol-phosphate aminotransferase (Figure 7). Comparison of the structures of L-threonine phosphate and histidinol phosphate indicates that the atoms between the amino group and the phosphate are structurally equivalent in both substrates. Thus, it is anticipated that the amino group of L-threonine phosphate would be able to interact with the carbonyl group of PLP if the phosphoryl group were located in the current position of the phosphate ion.

The residues that bind the PLP cofactor and the phosphate ion in the active site are remarkably similar to those that coordinate PLP and the phosphate moiety of histidinol phosphate in histidinol-phosphate aminotransferase as shown in Figures 6−8. Indeed, of the 12 residues that form their coordination sphere, six are identical, one utilizes an equivalent main chain interaction, and two involve either serine or threonine. The only significant differences are Thr^{86} (Asp^{85}) and Phe^{188} (Tyr^{187}), where the latter residue contributes an additional hydrogen bond to the coordination of the cofactor. In contrast, the overall level of identity between these two proteins is ~16%, as revealed by the structural alignment (Figure 8), whereas the level of sequence similarity is 44%. The overall structural and sequence similarity of these proteins suggests that they share a common ancestor. The same residues that coordinate PLP and the phosphate anion are also conserved in aspartate aminotransferase (Figure 8). This explains why some orthologs of CobD

have been annotated as amino acid aminotransferases in sequence databases.

Conformational State of CobD. PLP-dependent enzymes in the aspartate aminotransferase α-family exhibit a conformational change on substrate binding, where the small domain closes over the active site pocket. The extent of this change differs among members of the family depending on whether the domain movements involve the entire domain or primarily the N-terminal segment of the protein (*33*). In most members of the aspartate aminotransferase α-family whose structures are known, the N-terminus of one subunit starts as a short helical segment that lies against the 2-fold related subunit. Thereafter, the polypeptide chain extends across the active site pocket until it joins a second α-helix that lies against the *β*-sheet of the small domain. The conformational changes are associated with the position of the connecting loop and the second α-helix. In histidinol-phosphate aminotransferase, the conformational change manifests itself as a disorder−order transition in the connecting loop between the N-terminal α-helix and the small domain (*27*). Specifically in the unliganded structure, residues 18−32 are disordered, whereas when histidinol phosphate is bound, only residues 22−30 remain unseen and the additional residues appear to be adjacent to the substrate.

CobD utilizes a shorter interdomain N-terminal extension than either histidinol-phosphate aminotransferase or tyrosine aminotransferase by approximately 14 amino acid residues. As noted earlier, in CobD the seven residues at the N-terminus are disordered in the absence of substrate and the last residue seen in the electron density map, His^8, interacts with the inorganic phosphate ion bound in the active site. Clearly, there are insufficient residues in CobD to form an N-terminal extension that can interact in the same manner as the aminotransferases, although it is conceivable that the missing residues might contribute to the coordination of L-threonine phosphate. Even so, it would appear that the structure of CobD described here is nearer to the anticipated closed form of the enzyme than the apo state.

Structural Basis for Decarboxylation. The remarkable structural similarity in the active sites of CobD and histidinol-phosphate aminotransferase (Figure 7A) raises the question of how the progenitor of CobD, which was presumably an aminotransferase, was altered to become a decarboxylase. As mentioned earlier, the fate of the Schiff base formed with the amino group of the substrate depends on the conformation of the external aldimine in the active site and in particular on the orientation of the groups attached to the C4 position (*14*, *15*). It has been shown that the bond that is broken lies above or below the plane of the coenzyme−imine *π*-system such that alteration of the chemical outcome of a PLP-dependent reaction should simply require a reorientation of the substrate in the active site. In the absence of a ternary complex, modeling of L-threonine phosphate onto the ketimine histidinol phosphate intermediate (data not shown) suggests that the structural determinants responsible for the orientation of the substrate in the active site reside in the N-terminal extension of the small domain. This section of the protein differs most significantly in sequence and structure from the other members of the aspartate aminotransferase family. It is also the section of the protein that experiences a significant structural change during the catalytic cycle for this class of proteins (*33*). Thus, it seems likely that changes in the N-terminus of the progenitor of CobD were responsible for the alteration in chemical outcome of the PLP-dependent reaction.

CONCLUSIONS

The structural study of CobD presented here provides compelling evidence that this PLP-dependent decarboxylase evolved from a member of the aspartate aminotransferase family and perhaps that its progenitor was a histidinol-phosphate aminotransferase. The arrangement of the ligands surrounding the cofactor and putative substrate binding site is strikingly similar to that observed in histidinol-phosphate aminotransferase (*27*). This expands the known chemical fate for the external aldimine in the aspartate aminotransferase subfamily of PLP-dependent enzymes to include decarboxylation and complements the only other exception of 1-aminocyclopropane-1-carboxylate synthase (*13*).

The small differences between CobD and representative members of the aspartate aminotransferase family reinforce the question of why this subfamily has not been used more commonly to generate decarboxylases through enzyme evolution, rather than evolving distinct families of PLP-dependent enzymes devoted specifically to this task. This question is relevant since site-directed mutagenesis has been used to efficiently convert aspartate aminotransferase into an L-aspartate *β*-decarboxylase (*34*). Although this issue cannot be readily addressed, perhaps the close structural similarity between the amino and phosphoryl groups of L-threonine phosphate and histidinol phosphate accounts for the utilization of an aspartate aminotransferase instead of a member of the amino acid decarboxylase family in creating an enzyme that can decarboxylate this substrate. None of the normal amino acids is structurally equivalent to L-threonine phosphate.

A major reason for studying the structure and function of CobD was to determine how this enzyme, which is closely related to the aspartate aminotransferases, is able to facilitate a decarboxylation reaction. The present structure suggests that the N-terminal extension of the small domain is involved in coordinating the substrate and might be responsible for the alteration of the chemical fate of the external aldimine, though the groups responsible for coordinating the carboxylate moiety cannot be discerned with confidence. Certainly, it would be easier to evolve the N-terminal section of the protein rather than components of the core domain; however, the question of why this strategy has not been utilized elsewhere to create amino acid decarboxylases remains. A full understanding of how this enzyme controls the conformation of the external aldimine in the active site must await a structural study of this complex. Efforts to attain this structure are in progress.

ACKNOWLEDGMENT

We gratefully acknowledge the superb technical assistance from the following staff members at the Structural Biology Center (Advanced Photon Source, Argonne, IL): Dr. Frank Rotella, Dr. Norma E. C. Duke, Dr. Ruslan Sanishvili, Dr. Jack Lazarz, and Andrzej Joachimiak. The thoughtful comments of the reviewers are also appreciated.

REFERENCES

1. Scott, A. I. (1993) *Angew. Chem., Int. Ed. Engl. 32*, 1223−1243.

2. Blanche, F., Cameron, B., Crouzet, J., Debussche, L., Thibaut, D., Vuilhorgne, M., Leeper, F. J., and Battersby, A. R. (1995) *Angew. Chem., Int. Ed. Engl. 34*, 383−411.
3. O'Toole, G. A. (1994) Ph.D. Dissertation, University of Wisconsin, Madison, WI.
4. Rondon, M. R., Trzebiatowski, J. R., and Escalante-Semerena, J. C. (1997) *Prog. Nucleic Acid Res. Mol. Biol. 56*, 347−384.
5. Raux, E., Schubert, H. L., and Warren, M. J. (2000) *Cell. Mol. Life Sci. 57*, 1880−1893.
6. Brushaber, K. R., O'Toole, G. A., and Escalante-Semerena, J. C. (1998) *J. Biol. Chem. 273*, 2684−2691.
7. Mehta, P. K., and Christen, P. (1993) *Eur. J. Biochem. 211*, 373−376.
8. Christen, P., and Metzler, D. E. (1985) *Transaminases*, John Wiley and Sons, New York.
9. Ford, G. C., Eichele, G., and Jansonius, J. N. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 2559−2563.
10. Mehta, P. K., and Christen, P. (2000) *Adv. Enzymol. Relat. Areas Mol. Biol. 74*, 129−184.
11. Jansonius, J. N. (1998) *Curr. Opin. Struct. Biol. 8*, 759−769.
12. Mehta, P. K., Hale, T. I., and Christen, P. (1993) *Eur. J. Biochem. 214*, 549−561.
13. Capitani, G., Hohenester, E., Feng, L., Storici, P., Kirsch, J. F., and Jansonius, J. N. (1999) *J. Mol. Biol. 294*, 745−756.
14. Dunathan, H. C. (1966) *Proc. Natl. Acad. Sci. U.S.A. 55*, 712−716.
15. Dunathan, H. C. (1971) *Adv. Enzymol. Relat. Areas Mol. Biol. 35*, 79−134.
16. Kirsch, J. F., Eichele, G., Ford, G. C., Vincent, M. G., Jansonius, J. N., Gehring, H., and Christen, P. (1984) *J. Mol. Biol. 174*, 497−525.
17. Gerlt, J. A., and Babbitt, P. C. (2001) *Annu. Rev. Biochem. 70*, 209−246.
18. Laemmli, U. K. (1970) *Nature 227*, 680−685.
19. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C. W. J., Sweet, R. M., Abelson, J. N., and Simon, M. I., Eds.) pp 307−326, Academic Press, New York.
20. 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) *Acta Crystallogr. D54*, 905−921.
21. Roussel, A., and Cambillau, C. (1991) in *Silicon Graphics Geometry Partners Directory*, Silicon Graphics, Mountain View, CA.
22. Tronrud, D. E. (1997) *Methods Enzymol. 277*, 306−319.
23. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
24. Berman, H. M., Bhat, T. N., Bourne, P. E., Feng, Z., Gilliland, G., Weissig, H., and Westbrook, J. (2000) *Nat. Struct. Biol. 7*, 957−959.
25. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
26. Holm, L., and Sander, C. (1999) *Nucleic Acids Res. 27*, 244−247.
27. Haruyama, K., Nakai, T., Miyahara, I., Hirotsu, K., Mizuguchi, H., Hayashi, H., and Kagamiyama, H. (2001) *Biochemistry 40*, 4633−4644.
28. Nakai, T., Okada, K., Akutsu, S., Miyahara, I., Kawaguchi, S., Kato, R., Kuramitsu, S., and Hirotsu, K. (1999) *Biochemistry 38*, 2413−2424.
29. Blankenfeldt, W., Nowicki, C., Montemartini-Kalisz, M., Kalisz, H. M., and Hecht, H. J. (1999) *Protein Sci. 8*, 2406−2417.
30. Malashkevich, V. N., Strop, P., Keller, J. W., Jansonius, J. N., and Toney, M. D. (1999) *J. Mol. Biol. 294*, 193−200.
31. Momany, C., Ernst, S., Ghosh, R., Chang, N. L., and Hackert, M. L. (1995) *J. Mol. Biol. 252*, 643−655.
32. Kern, A. D., Oliveira, M. A., Coffino, P., and Hackert, M. L. (1999) *Struct. Folding Des. 7*, 567−581.
33. Rhee, S., Silva, M. M., Hyde, C. C., Rogers, P. H., Metzler, C. M., Metzler, D. E., and Arnone, A. (1997) *J. Biol. Chem. 272*, 17293−17302.
34. Graber, R., Kasper, P., Malashkevich, V. N., Strop, P., Gehring, H., Jansonius, J. N., and Christen, P. (1999) *J. Biol. Chem. 274*, 31203−31208.
35. Cohen, G. H. (1997) *J. Appl. Crystallogr. 30*, 1160−1161.
36. Johnson, M. G., and Escalante-Semerena, J. C. (1992) *J. Biol. Chem. 267*, 13302−13305.
37. Thompson, T. B., Thomas, M. G., Escalante-Semerena, J. C., and Rayment, I. (1999) *Biochemistry 38*, 12995−13005.
38. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
39. Esnouf, R. M. (1999) *Acta Crystallogr. D55*, 938−940.
40. Schwede, T., Diemand, A., Guex, N., and Peitsch, M. C. (2000) *Res. Microbiol. 151*, 107−112.

BI012111W