# Crystal Structure of the Biphenyl-Cleaving Extradiol Dioxygenase from a PCB-Degrading Pseudomonad

Seungil Han, Lindsay D. Eltis, Kenneth N. Timmis, Steven W. Muchmore,* Jeffrey T. Bolin†

Polychlorinated biphenyls (PCBs) typify a class of stable aromatic pollutants that are targeted by bioremediation strategies. In the aerobic degradation of biphenyl by bacteria, the key step of ring cleavage is catalyzed by an Fe(II)-dependent extradiol dioxygenase. The crystal structure of 2,3-dihydroxybiphenyl 1,2-dioxygenase from a PCB-degrading strain of *Pseudomonas cepacia* has been determined at 1.9 angstrom resolution. The monomer comprises amino- and carboxyl-terminal domains. Structural homology between and within the domains reveals evolutionary relationships within the extradiol dioxygenase family. The iron atom has five ligands in square pyramidal geometry: one glutamate and two histidine side chains, and two water molecules.

Microbial pathways that degrade aromatic compounds play a key role in detoxifying and mineralizing significant industrial pollutants and thus are targeted by strategies for bioremediation (*1*). A critical point in these pathways is the opening of the chemically stable aromatic ring. In typical aerobic pathways, the ring is first activated by hydroxylation on adjacent carbons to form a catechol-like compound. Ring cleavage, catalyzed by a dioxygenase, is then effected between the hydroxylated carbon atoms (intradiol- or *ortho*-cleavage) or adjacent to one of these carbons (extradiol- or *meta*-cleavage) (*2*). *Meta*-cleavage reactions catalyzed by extradiol dioxygenases are particularly significant for alkyl- and poly-aromatics as well as their halogenated analogs. Extradiol dioxygenases exploit Fe(II) in a mononuclear nonheme, nonsulfur environment and are established subjects for pathway adaptation and enzyme engineering (*3*). Although Fe(III)-dependent intradiol dioxygenases have been crystallized and structurally characterized (*4*), diffraction quality crystals of extradiol dioxygenases have proved difficult to obtain.

In the biphenyl/PCB-degrading pathway, the ring-cleavage step converts 2,3-dihydroxybiphenyl to 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid and is catalyzed by an extradiol dioxygenase, 2,3-dihydroxybiphenyl 1,2-dioxygenase (DHBD; E.C. 1.13.11.39). The crystal structure of DHBD from *Pseudomonas cepacia* LB400 (*Ps.* LB400), a strain exhibiting an ability to transform a broad range of PCB congeners (*5*), has been determined at 1.9 Å resolution under anaerobic conditions (*6*, *7*). The diffraction data and crystallographic procedures are summarized in Table 1. Initial phases at 2.8 Å resolution were obtained by the multiple isomorphous replacement (MIR) method and then improved and extended to 2.3 Å resolution by an iterative solvent-flattening procedure. The resulting electron density maps were of very high quality, and an atomic model was constructed for residues 2 through 288; its *R* factor was 34% at 2.5 Å resolution. The model was refined to $R = 16\%$ with the use of data between 7 and 1.9 Å resolution with $|F|/\sigma \geq 1$. No unusual combinations of $\phi$ and $\psi$ are present in the final model, and the agreement of bond lengths and angles with expected values is quite good (Table 1). A Luzzati plot (*8*) indicates that the average coordinate error is no larger than 0.2 Å.

DHBD is an homo-octamer of the 33-kD subunit encoded by the *bph*C gene (*9*). The octamer is a nearly isometric, hollow particle with 422 ($D_4$) symmetry and a maximum dimension of 95 Å (*10*). An $NH_2$-terminal methionine is excised so that each monomer comprises one chain of 297 residues and one Fe atom, which is Fe(II) in the active enzyme (*11*). The COOH-terminal 10 residues appear to be disordered and are not included in the final model.

Surprisingly, the monomer has $NH_2$- and COOH-terminal domains of very similar structure related by local pseudo-twofold symmetry (Fig. 1). The first 134 residues may be assigned to the N-domain, which abuts but does not penetrate the slightly larger C-domain, residues 135 to 298. Cα atoms from 93 residues of each domain can be superimposed with a root mean square (rms) Cα-Cα distance of 1.3 Å. The three-dimensional structures of the domains are thus remarkably similar despite only 17% sequence identity.

Each domain includes an eight-stranded mixed β sheet partially surrounding a large, funnel-shaped space lying entirely within

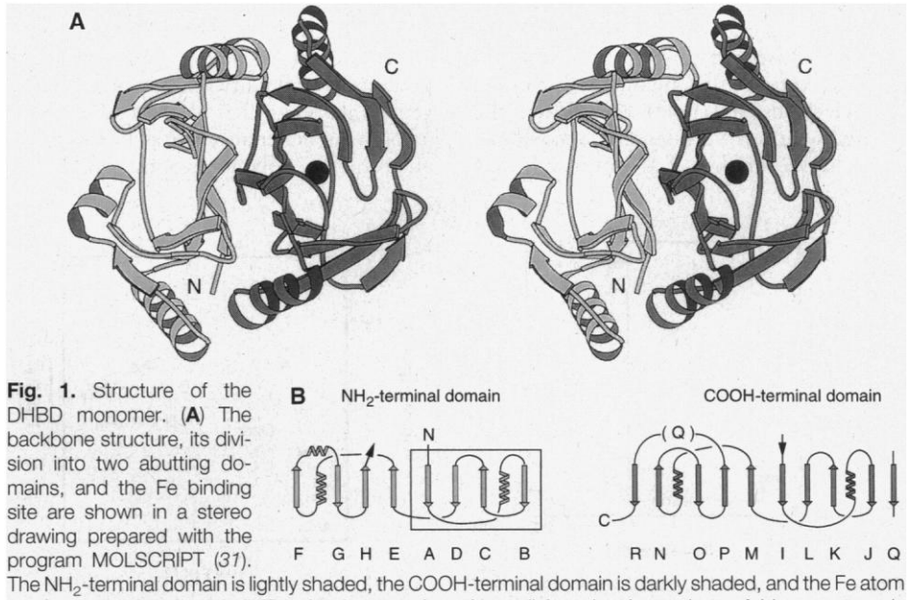


**Fig. 1.** Structure of the DHBD monomer. (**A**) The backbone structure, its division into two abutting domains, and the Fe binding site are shown in a stereo drawing prepared with the program MOLSCRIPT (*31*). The $NH_2$-terminal domain is lightly shaded, the COOH-terminal domain is darkly shaded, and the Fe atom is represented by a dark ball. The view is approximately parallel to a local pseudo-twofold symmetry axis valid for major portions of the two domains. N, N-domain on the protein terminus; C, C-domain on the protein terminus. (**B**) Schematic drawing illustrating the similar folding patterns of the two domains. The β strands are labeled in sequential order. βQ of the C-domain is shown twice: once as an isolated arrow to indicate its hydrogen bonding with βJ, and once by the symbol (Q) to indicate its location within the connection between βP and βR. The box in the N-domain surrounds a βαββ unit that occurs four times in the two domains.

S. Han, S. W. Muchmore, J. T. Bolin, Department of Biological Sciences, Purdue University, West Lafayette, IN 47907–1392, USA.
L. D. Eltis, Department of Biochemistry, Université Laval, Ste. Foy, Quebec G1K 7P4, Canada.
K. N. Timmis, Bereich Mikrobiologie, Gesellschaft für Biotechnologische Forschung mbH, Mascheroder Weg 1, D-38124 Braunschweig, Germany.

*Present address: Abbott Laboratories, Department 42T Building AP9A, 100 Abbott Park Road, Abbott Park, IL 60064–3500, USA.
†To whom correspondence should be addressed.

the domain. The active-site Fe binds deep within this space in the C-domain. Corresponding strands from the β sheets of the two domains as well as peripheral helices form the domain-domain interface, which includes a core of predominantly hydrophobic side chains.

Within each domain are two copies of a substructure comprising four β strands and one α helix in sequential order βαβββ. Thus, the monomer contains four copies of this unit within two domains, suggesting genetic duplication at two points in the evolutionary history of the enzyme. A search of a current version of the Families of Structurally Similar Proteins (FSSP) database of structurally aligned protein-fold families with the DALI algorithm identified two occurrences of topologically similar βαβββ motifs but no occurrences of structures similar to the complete core of the DHBD domains (*12*).

**Table 1.** Summary of *Ps.* LB400 DHBD structure determination. Diffraction measurements were made at room temperature with Cu-Kα radiation. Data from native crystal 1 and the derivatives were measured with San Diego Multiwire Systems area detectors and evaluated with program XDS (*32*). Data from native crystal 2 were measured with an R-axis II system (Molecular Structures) and evaluated with the HKL package (*33*). Programs from the CCP4 package (*34*) were used for various calculations. Heavy-atom parameter refinements and phase calculations were performed to 2.8 Å resolution with MLPHARE (*35*). A locally developed software package (*36*) was used in three passes to improve and extend the phases by solvent flattening. In both the first and second passes, the limiting resolution of the maps was extended in 41 steps from 5.0 to 2.8 Å. Five cycles of solvent-flattening were used for each step. Reflections added at each step were assigned a combined phase on the basis of MIR and back-transform phase probability distributions for five cycles. The final phases and maps from the first two passes were used to improve the mask used for solvent-flattening and the heavy-atom model. In the final pass, phase improvement progressed from 5.0 to 2.3 Å resolution in 60 steps. From 5.0 to 2.8 Å, the phase assignment scheme described above was used; from 2.8 to 2.3 Å, all reflections were assigned the back-transform phase in every cycle. Atomic models were constructed and modified using O (*37*). PROCHECK (*38*) was used to assess stereochemical properties. X-PLOR (*39*) was used for computational refinement. The model was initially refined against 2.2 Å resolution data from native crystal 1 and finally refined against 1.9 Å data from native crystal 2. During the final refinement cycle, Fe-ligand distances were restrained by the default X-PLOR bond energy function with an energy constant of 10 kcal $mol^{-1}$ $Å^{-2}$ and a target distance of 2.1 Å. Protein coordinates have been deposited with the Protein Data Bank (accession number 1HAN).

Diffraction data statistics

| Crystal* | Resolution (Å) | Total observations† | | Unique reflections | | $R_{merge}$‡ (%) | |
|---|---|---|---|---|---|---|---|
| | | All data | Last shell | All data | Last shell | All data | Last shell |
| Native1 | 2.2 | 130,446 | 3,951 | 19,380 (89%) | 1,998 (65%) | 6.2 | 17.0 |
| Native2 | 1.9 | 287,451 | 13,817 | 33,261 (99%) | 3,303 (99%) | 6.0 | 22.5 |
| PIP | 2.8 | 31,222 | 1,663 | 10,031 (94%) | 969 (64%) | 7.7 | 25.4 |
| EMTS | 2.8 | 31,214 | 1,507 | 8,765 (82%) | 759 (50%) | 6.2 | 17.9 |
| $K_2PtCl_4$ | 2.8 | 32,261 | 2,397 | 9,830 (92%) | 1,226 (81%) | 6.4 | 18.9 |
| $Me_3Pb$ | 2.8 | 28,276 | 1,364 | 9,892 (92%) | 879 (58%) | 6.7 | 17.1 |

Phasing statistics

| Derivative crystal | Resolution (Å) | Number of sites | Isomorphous *R* factor§ | | Phasing power‖ | |
|---|---|---|---|---|---|---|
| | | | All data | Last shell | All data | Last shell |
| PIP | 2.8 | 2 | 0.16 | 0.24 | 0.7 | 0.5 |
| EMTS | 2.8 | 1 | 0.13 | 0.19 | 0.8 | 0.7 |
| $K_2PtCl_4$ | 2.8 | 2 | 0.11 | 0.16 | 1.2 | 0.8 |
| $Me_3Pb$ | 2.8 | 5 | 0.13 | 0.19 | 0.8 | 0.6 |

Refinement statistics

| | | | | |
|---|---|---|---|---|
| Number of reflections (7 to 1.9 Å) | Working set: | 29,672 | Test set: | 1867 |
| Number of non-H atoms | Protein + Fe: | 2,208 | Solvent: | 144 |
| *R* factors | Standard: | 0.162 | Free: | 0.190 |
| Average *B* factors ($Å^2$) | Main chain: | 19.6 | Side chain: | 23.6 |
| Stereochemical deviations (rms) | Bond lengths: | 0.010 Å | Bond angles: | 1.45° |

*The derivative crystals are as follows: PIP, di-μ-iodo bis(ethylenediamine) diplatinum (II) nitrate; EMTS, sodium ethylmercurithiosalicylate; $Me_3Pb$, trimethyllead acetate. †"All data" encompasses from 30 Å to the limiting resolution; "Last shell" refers to the highest resolution data. The width of the shell varies, but is typically in the range 0.1 to 0.4 Å. ‡$R_{merge} = \Sigma_h\Sigma_i | I_{hi} - \langle I_{hi}\rangle | / \Sigma_h\Sigma_i | I_{hi} |$, where *h* specifies unique reflection indices, *i* indicates symmetry equivalent observations of *h*, and $\langle I_{hi}\rangle$ is the mean value. §The isomorphous *R* factor = $\Sigma_h||F_{PH}| - |F_P||/\Sigma_h|F_P|$, where $|F_{PH}|$ and $|F_P|$ are the measured structure factor amplitudes of the derivative and native structures. ‖The phasing power is the ratio of the structure factor amplitude for the added atoms in a derivative to the estimated error in the phasing model (closure error).

An important distinction between the domains is the presence in the C-domain of two β strands, βQ and βR, which follow the common core structure and are hydrogen bonded to opposite edges of the core sheet. As a result, the outer wall of the C-domain's central funnel appears to be more effectively closed. Closure is not achieved by continuous, backbone-to-backbone hydrogen bonding as in a true β barrel. Rather, a loop between βN and βO packs into one surface of the β sheet such that it makes contact with side chains from βJ, βK, and βQ. These interactions, as well as others involving side chains from βL, βN, βO, and nearby connecting segments, complete the outer wall of the funnel.

Examination of the sequences of extradiol dioxygenases with monomers of relative molecular mass ≈33 kD suggests that a two-domain structure is a common feature of many Fe-dependent enzymes that use the *meta*-cleavage reaction to open a variety of aromatic rings (*13*). Moreover, it appears that metal binding and catalytic activity are not likely to be present in the N-domains of any of the two-domain enzymes characterized to date.

The presence of the domain structure also leads to a prediction that a class of smaller (21 kD) DHB dioxygenases represented by the BphC2 enzyme of *Rhodococcus globerulus* strain P6 (*14*) includes homologs that incorporate only one copy of the domain. Although an alignment of the sequences of the BphC2 enzyme and the C-domain (Fig. 2) yields only 17% sequence identity (*15*), the identities include all of the ligands, several residues that form hydrogen bonds with them, several residues potentially involved in substrate binding, and glycine and proline residues at points where the polypeptide chain dramatically changes course.

Thus, it appears that during the course of the evolution of extradiol dioxygenases, the second of two duplication events was followed by divergence of one- and two-domain enzymes, and that further adaptation of the two-domain enzymes resulted in the loss of a second active site within the N-domain. It is also noteworthy that residues of BphC2 corresponding to C-domain residues at the largely hydrophobic domain-domain interface are similarly hydrophobic. This observation suggests that BphC2 may be constructed from dimers analogous to the two-domain enzymes, and that an ancestral protein was itself dimeric.

The active site Fe atom was identified as the strongest feature in initial electron density maps; Bijvoet difference maps confirmed the assignment. As noted, the Fe is bound within a funnel-shaped space surrounded by the C-domain's β strands and loops. The funnel is 20 Å long and is open

at both ends. As defined by side chain atoms, one opening is 10 Å wide, whereas the other is only 6 Å wide. Thus, the Fe is probably accessible to catecholic substrates only through the wide opening, which faces the outside of the octamer. However, water or dioxygen may approach the Fe through the smaller opening, which opens to the large internal channel (*10*).

The Fe atom is bound midway between the openings and somewhat off the funnel's axis such that it is closest to β strands I and M along the wall that participates in the domain-domain interface. Three side chains from this wall bind directly to the Fe, whereas interactions with side chains from the opposite wall are bridged by water molecules. Five ligands bind the Fe atom: the side chains of three conserved residues, $His^{146}$, $His^{210}$, and $Glu^{260}$; and two water molecules, numbered 3001 and 3012. Figure 3 demonstrates the quality of the electron density at the Fe site as well as difference density that confirms the identification of the Fe atom and water ligands. The geometry at the Fe site is remarkably consistent with square-pyramidal coordination, with $His^{146}$ as the axial ligand (see Table 2).

Spectroscopic studies of extradiol dioxygenases are generally, if not exclusively, in agreement with the above description. On the basis of several spectroscopic methods, Mabrouk *et al.* (*16*) concluded that the Fe(II) site of catechol 2,3-dioxygenase from *Pseudomonas putida* has five-coordinate, square-pyramidal coordination in the resting state. Recently, x-ray absorption spectroscopy has been used by two groups to probe the Fe site in the same enzyme. An analysis of pre-edge and extended x-ray absorption fine structure (EXAFS) spectral features by Shu *et al.* (*17*) supports the square-pyramidal model. In contrast, Bertini *et al.* (*18*) argued that pre-edge features were consistent with a higher symmetry site, implying six ligands in octahedral coordination. Evidence for at least two and probably three exogenous ligand-binding sites accessible to displaceable water molecules, substrates, and the $O_2$ analog NO has been provided by electron paramagnetic resonance studies of enzyme-NO and enzyme-NO-substrate complexes (*19*).

The only conflict among the spectroscopic and crystallographic results concerns the possible presence of a sixth ligand opposite the axial ligand, $His^{146}$. In the crystal structure, the conserved, hydrogen-bonded side chains of $His^{241}$ and $Tyr^{250}$ approach and block the sixth site (Fig. 4), but neither is close enough to act as a ligand. $Tyr^{250}$-$O^{\eta}$ is near the octahedral axis, but 3.8 Å from the Fe and 1.8 Å from the predicted sixth site; $His^{241}$-$N^{\varepsilon 2}$ is 4.8 Å from the Fe. Another potential ligand is the second carboxylate oxygen of $Glu^{260}$, $O^{\varepsilon 2}$. However, hydrogen bonds position it 3.6 Å away from the Fe and 3.4 Å from the sixth site. The possibility that a water molecule might serve as the sixth ligand has been carefully considered and rejected (*20*).

Table 2 lists ligand-Fe distances and ligand-Fe-ligand angles. Compared to typical distances of 2.1 to 2.2 Å observed in crystal structures of appropriate model complexes (*21*), the distances in the refined DHBD structure, which average 2.2 Å, are reasonable (*22*). Likewise, the distances are compatible with the EXAFS analyses

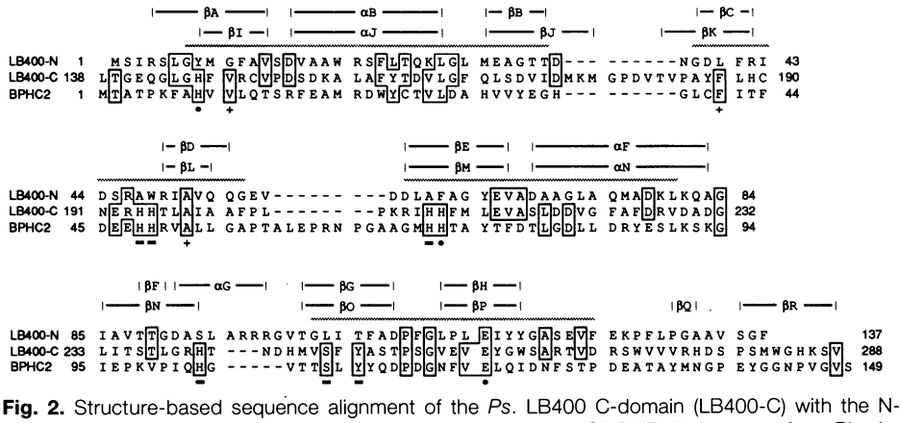


**Fig. 2.** Structure-based sequence alignment of the *Ps.* LB400 C-domain (LB400-C) with the N-domain (LB400-N) and with an $NH_2$-terminal segment of the BphC2 DHB-dioxygenase from *Rhodococcus globerulus* strain P6 (BPCH2). Identical residues are boxed. The extent of secondary structure elements in the N- and C-domains is indicated in the top two lines. Symbols beneath the sequences indicate the Fe ligands (●), conserved residues involved in hydrogen bonds with the ligands (−), or other conserved residues in the substrate binding site (+). Hatched lines above the sequences indicate segments of the N- and C-domains for which Cα atoms may be superimposed to within 3.2 Å. Gaps (dashes) were introduced to account for differences in polypeptide chain length.

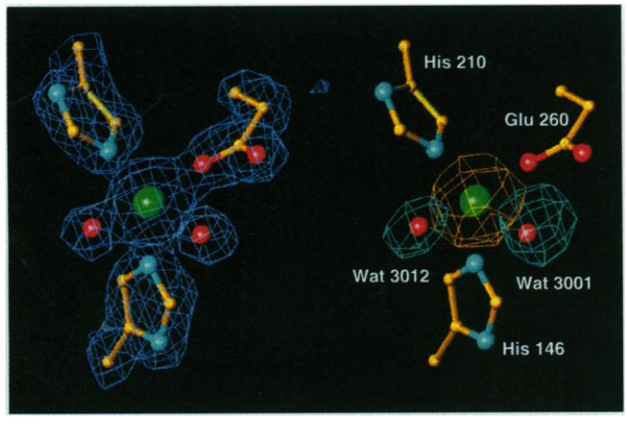


**Fig. 3.** Electron density maps in the vicinity of the active-site Fe atom. On the left is the final $2F_o - F_c$ map contoured at three times the rms density of the map. On the right, the orange density (20 rms) is from a Bijvoet anomalous difference map used to confirm the identity of the Fe atom, and the turquois density (3 rms) is from an $F_o - F_c$ map calculated just before the water ligands were added to the model. The Fe atom is represented by a large green sphere.

**Table 2.** Coordination geometry at the active site Fe.

| Metal-ligand distances (Å) | |
|---|---|
| Defining atoms* | Distance† |
| Fe-H146-$N^{\varepsilon 2}$ | 2.15 |
| Fe-H210-$N^{\varepsilon 2}$ | 2.25 |
| Fe-E260-$O^{\varepsilon 1}$ | 1.96 |
| Fe-Wat3001-O | 2.41 |
| Fe-Wat3012-O | 2.12 |
| Fe to central plane | 0.36‡ |

| Ligand-metal-ligand angles (degrees) | | |
|---|---|---|
| Defining atoms | Observed | Expected§ |
| H210-$N^{\varepsilon 2}$-Fe-E260-$O^{\varepsilon 1}$ | 86 | (88) |
| E260-$O^{\varepsilon 1}$-Fe-Wat3001-O | 82 | (88) |
| Wat3001-O-Fe-Wat3012-O | 91 | (88) |
| Wat3012-O-Fe-H210-$N^{\varepsilon 2}$ | 95 | (88) |
| H146-$N^{\varepsilon 2}$-Fe-central plane | 81 | (90) |
| H146-$N^{\varepsilon 2}$-Fe-H210-$N^{\varepsilon 2}$ | 102 | (100) |
| H146-$N^{\varepsilon 2}$-Fe-E260-$O^{\varepsilon 1}$ | 110 | (100) |
| H146-$N^{\varepsilon 2}$-Fe-Wat3001-O | 96 | (100) |
| H146-$N^{\varepsilon 2}$-Fe-Wat3012-O | 92 | (100) |
| H210-$N^{\varepsilon 2}$-Fe-Wat3001-O | 158 | (160) |
| E260-$O^{\varepsilon 1}$-Fe-Wat3012-O | 161 | (160) |

*H, histidine; E, glutamate; Wat, water. †The precision is ~0.1 Å. ‡The central plane is the least-squares plane of H210-$N^{\varepsilon 2}$, E260-$O^{\varepsilon 1}$, Wat3001-O, and Wat3012-O; all four atoms are within 0.02 Å of the plane. §Expected values are based on Fe-ligand distances of 2.1 Å and an angle of 10° between the equatorial ligand-to-Fe vectors and the plane of the ligands.

cited above, which fit single shells of N or O atoms at 2.09 Å (*17*) and 2.05 Å (*18*).

In addition to the $His^{241}$-$Tyr^{250}$ pair described above, several other residues in the second shell contribute to the environment of the Fe atom. It is of particular interest that the positions of all ligands appear to be stabilized by supporting hydrogen bonds (Fig. 4). Many of these bonds involve side chains that are conserved or conservatively substituted among extradiol dioxygenases. Additional features of the extended active site include two His-His pairs linked by $N^{\delta1}$-$N^{\delta1}$ hydrogen bonds, disruptions to the β structure near the Fe atom, and the apparent binding of a *t*-butanol molecule (*23*). Although $N^{\delta1}$-$N^{\delta1}$ hydrogen bonds are not necessarily rare (*24*), the residues involved in this case, $His^{194}$ with $His^{195}$ and $His^{209}$ with ligand $His^{210}$, are conserved and adjacent in sequence. Moreover, they are also involved in a β bulge or related structure that places the second residue near the Fe atom. This circumstance is indeed rare, as histidine is very infrequently found to be the second residue in a β bulge (*25*). Both bulges are also found in the N-domain, which may support the hypothesis that Fe binding was lost by the N-domain during evolution, rather than gained by the C-domain.

Despite the structural similarity of the two domains, Fe binding is not observed within the N-domain and is unlikely because side chains that might serve as ligands are not available. Moreover, the space in the N-domain where the Fe and substrates would bind is filled by large side chains and a salt bridge (*26*).

Previous studies of extradiol dioxygenases indicate that catecholic substrates bind to Fe via both hydroxyl groups before dioxygen binds (*16*, *19*, *27*). Water ligands are displaced when substrate binds, but the coordination geometry is believed to be similar to that of the free enzyme. By extension of these results to DHBD, substrate binding was modeled under the following assumptions: (i) the enzyme:DHB complex is five-coordinate with square-pyramidal geometry; and (ii) protein ligands are not displaced, so that the substrate's O2 and O3 atoms either replace the water ligands or bind in the vacant sixth site. These assumptions allow six binding modes, of which four were eliminated on the basis of steric conflicts with the protein.

In both of the remaining two modes, O3 of the substrate replaces water-3001. In mode A, O2 replaces water-3012, whereas in mode B, O2 binds in the vacant sixth site. In both modes, the substrate's dihydroxylated ring is tucked into a pocket defined by the Fe atom; the side chains of several conserved residues, $Phe^{187}$, $His^{195}$, $His^{241}$, and $Tyr^{250}$; and the side chains of variable residues $Asn^{243}$, $Asp^{244}$, and $Met^{246}$. The interactions of the distal ring differ between the two modes because it occupies different portions of a larger binding pocket that extends toward the wide opening of the active site funnel and encompasses the *t*-butanol binding site (*28*). A clear difference between the two modes is the torsion angle between the rings, which is ~30° in mode A and ~80° in mode B.

For both modes, the presence of free space near edges of both rings suggests *Ps.* LB400 DHBD may be able to act on various chloro-substituted substrates, although this has yet to be established. Mode B may be favored in that substituents are more readily tolerated at either edge of the distal ring. In addition, a larger torsion angle is sterically favorable for *ortho*-substituted compounds.

The active site structure of DHBD is consistent with the initial steps of current proposals (*17*) for the general mechanism of extradiol dioxygenases. A key consideration is the presence of an active site basic residue (or residues) to facilitate deprotonation of one or both of the substrate's hydroxyl groups. $His^{195}$ is the only conserved active site base likely to fulfill this role. In the free enzyme, $His^{195}$-$N^{\varepsilon2}$ forms a hydrogen bond with one of the Fe atom's water ligands, water-3012, whereas $His^{195}$-$N^{\delta1}$ is involved in a hydrogen bond with $N^{\delta1}$ of conserved $His^{194}$. $His^{194}$-$N^{\varepsilon2}$ in turn forms a hydrogen bond with the carboxylate group of $Asp^{171}$. It is thus probable that $His^{194}$ and $His^{195}$ are both deprotonated and that $His^{195}$ is available as a base to assist in substrate deprotonation and binding. $His^{195}$ may also play a role in the binding and activation of $O_2$ (*29*).

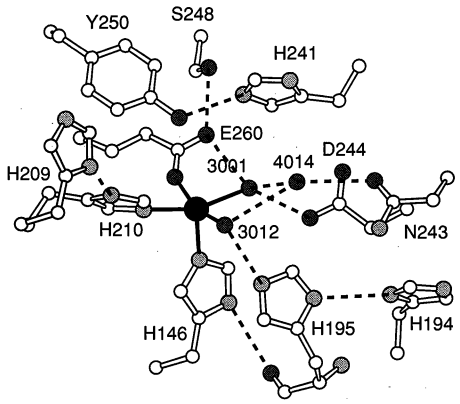


**Fig. 4.** View of the Fe binding site. Protein side chains and water molecules in the vicinity of the Fe atom are illustrated. The Fe atom is shown as a large, dark ball. N and O atoms are gray, with O darker. C atoms are white. Bonds between ligands and the Fe are solid black, and hydrogen bonds are dashed. Prepared with the program MOLSCRIPT (*31*).

## REFERENCES AND NOTES


1. K. N. Timmis, R. J. Steffan, R. Unterman, *Annu. Rev. Microbiol.* **48**, 525 (1994).
2. S. Harayama, M. Kok, E. L. Neidle, *ibid.* **46**, 565 (1992).
3. J. L. Ramos, A. Wasserfallen, K. Rose, K. N. Timmis, *Science* **235**, 593 (1987); A. Wasserfallen, M. Rekik, S. Harayama, *Bio/Technology* **9**, 296 (1991); P. Cerdan, A. Wasserfallen, M. Rekik, K. N. Timmis, S. Harayama, *J. Bacteriol.* **176**, 6074 (1994).
4. D. H. Ohlendorf, J. D. Lipscomb, P. C. Weber, *Nature* **336**, 403 (1988); D. H. Ohlendorf, A. M. Orville, J. D. Lipscomb, *J. Mol. Biol.* **244**, 586 (1994).
5. L. H. Bopp, *J. Ind. Microbiol.* **1**, 23 (1986); D. A. Abramowitz, *Crit. Rev. Biotechnol.* **10**, 241 (1990).
6. We purified DHBD by adapting established procedures (*11*) to anaerobic conditions so as to enhance its long-term stability. Crystals were grown in a glove box at 5° to 10°C by vapor diffusion under a $N_2$ atmosphere containing less than 2 parts per million $O_2$. The crystallization solution contained 18 to 22% polyethylene glycol 4000 and 10 to 15% *t*-butanol and was buffered at pH 7.5. DHBD crystals are colorless and enzyme recovered from dissolved crystals is active. Immediately before diffraction experiments crystals were mounted and sealed in capillaries within the glove box. The diffraction pattern indicates space group *I*422 with $a$ = 122.6 Å and $c$ = 111.4 Å. Each DHBD octamer occupies a site of 422 symmetry so that there is one monomer (297 amino acids) in the asymmetric unit and a solvent content of 61%. See Table 1 for additional information on crystallographic procedures.
7. Portions of this work were presented by J. T. Bolin at the 2nd European Bioinorganic Chemistry Conference (EUROBIC II), 30 August 1994, on the basis of a model refined at 2.2 Å resolution. A preliminary report on the structure of the inactive, Fe(III) form of a homologous enzyme at 2.6 Å resolution has been presented by K. Sugiyama *et al.*, [*Proc. Jpn Acad. Ser. B* **71**, 32 (1995)].
8. V. Luzzati, *Acta Crystallogr.* **5**, 802 (1952).
9. B. Hofer, L. D. Eltis, D. N. Dowling, K. N. Timmis, *Gene* **130**, 47 (1993).
10. The crystal structure agrees with previously reported electron micrographs, which indicate that the octamer is assembled from two planar, tetrameric layers stacked in a staggered conformation (*11*). A wide channel traverses the octamer along its fourfold axis. In the middle and at its ends the channel is 40 Å wide, but it narrows to 20 Å midway between the center and the ends.
11. L. D. Eltis, B. Hofmann, H.-J. Hecht, H. Lünsdorf, K. N. Timmis, *J. Biol. Chem.* **268**, 2727 (1993).
12. The FSSP database [L. Holm, C. Ouzounis, C. Sander, G. Tuparev, G. Vriend, *Protein Sci.* **1**, 1691 (1992); L. Holm and C. Sander, *Nucleic Acids Res.* **22**, 3600 (1994)] is a compilation of representative three-dimensional protein structures based on an automated analysis of the Protein Data Bank [F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977)] coordinate entries. The DALI algorithm is described by L. Holm and C. Sander [*J. Mol. Biol.* **233**, 123 (1993)].
13. J. T. Bolin, L. D. Eltis, S. Han, unpublished results. A two-domain structure is also predicted for a recently characterized Mn-dependent extradiol dioxygenase [Y. R. Boldt, M. J. Sadowsky, L. B. M. Ellis, L. Que Jr., L. P. Wackett, *J. Bacteriol.* **177**, 1225 (1995)].
14. J. A. Asturias, L. D. Eltis, M. Prucha, K. N. Timmis, *J. Biol. Chem.* **269**, 7807 (1994).
15. The alignment of the BphC2 and DHBD COOH-terminal domain sequences is an extrapolation of the structure-based alignment of the two domains of DHBD. All gaps in the alignment correspond to segments between elements of secondary structure.
16. P. A. Mabrouk, A. M. Orville, J. D. Lipscomb, E. I. Solomon, *J. Am. Chem. Soc.* **113**, 4053 (1991).
17. L. Shu *et al.*, *Biochemistry* **34**, 6649 (1995).
18. I. Bertini, F. Briganti, S. Mangani, H. F. Nolting, A. Scozzafava, *ibid.* **33**, 10777 (1994).
19. D. M. Arciero, A. M. Orville, J. D. Lipscomb, *J. Biol. Chem.* **260**, 14035 (1985); D. M. Arciero and J. D. Lipscomb, *ibid.*, p. 2170.
20. When the water ligands were added to the model, they were represented in ($F_o$ – $F_c$) maps by spherical

features with maxima higher than five times the rms deviation from the mean density (Fig. 4). The only feature that might have represented a sixth ligand had a maximum density similar to the rms deviation. In the final ($F_o - F_c$) map, a bean-shaped feature at three times the rms deviation was observed 2.4 Å from the Fe and along the line between $Tyr^{250}$-$O^{\eta}$ and water-3012. However, when a water molecule was added at this site, refinement resulted in displacement to 2.6 Å away from the Fe and a *B* factor of 64 $Å^2$; in contrast, the *B* factors for the ligands and Fe atom are 11 to 14 $Å^2$. Thus, the addition of a water molecule as a sixth ligand is not warranted.
21. R. C. Scarrow *et al.*, *Biochemistry* **33**, 15023 (1994).
22. To assess the reliability of the Fe-ligand distances, we investigated the influence of refinement restraints, the consequences of variations in the resolution and source of the diffraction data, and tested for bias in the final model by systematically changing Fe-ligand distances or randomly perturbing the entire model. In brief, we found that the use of weak Fe-ligand bond distance restraints (for example, 10 kcal $mol^{-1}$ $Å^{-2}$) produced consistent results for all ligands given a variety of starting models and diffraction data sets. None of the perturbations caused variations as large as 0.1 Å for any distance. The final structure is also entirely consistent with omit-refine maps (*30*) obtained by omitting all atoms within 6 Å of the Fe.
23. A *t*-butanol molecule was assigned to a large-density feature 6 Å from the Fe. In the view shown in Fig. 4, the *t*-butanol is located along the line-of-sight to $Tyr^{250}$-$O^{\eta}$ and approximately at the depth of the side chain of $His^{209}$; it is not illustrated because it obscures the Fe. The *t*-butanol blocks the most direct route to the Fe via the large opening of the funnel. This is consistent with the observation that *t*-butanol, in concentrations similar to that in the crystallization solution, inhibits and stabilizes the enzyme.
24. J. Singh and J. M. Thornton, *Atlas of Protein Side-Chain Interactions* (Oxford Univ. Press, New York, 1992).
25. A. W. E. Chan, E. G. Hutchinson, D. Harris, J. M. Thornton, *Protein Sci.* **2**, 1574 (1993).
26. C-domain ligands $His^{146}$, $His^{195}$, and $His^{210}$ correspond to N-domain residues $Tyr^{8}$, $Trp^{48}$, and $Phe^{62}$. $Glu^{115}$, the residue corresponding to ligand $Glu^{260}$, forms a salt bridge with $Arg^{98}$, which extends from across the funnel. Thus, neither the ligands nor the free space necessary for Fe and substrate binding are available. However, a second Fe is found at the surface of the C-domain for crystals that were transferred to a stabilizing solution containing 1 mM $(NH_4)_2Fe(SO_4)_2$ before diffraction measurements. The additional Fe is bound only by $His^{189}$ and, according to its anomalous difference density, is present at partial (~50%) occupancy. Anomalous difference density is observed only for crystals treated with excess Fe.
27. K. Hori, T. Hashimoto, M. Nozaki, *J. Biochem. (Tokyo)* **74**, 375 (1973).
28. Strongly conserved residues, including $Val^{148}$, $Phe^{187}$, $Ala^{198}$, $His^{209}$, and $Tyr^{250}$, border on this larger pocket, consistent with its assignment as the distal ring binding site.
29. Binding to the Fe(II) center may activate dioxygen by converting it to an Fe-bound, superoxide-like species (*17*), which attacks the substrate at C1, the point of attachment for the distal ring. For either binding mode, conserved residues are available to promote the formation of bound superoxide by stabilizing a partial negative charge on its distal oxygen atom. For mode A, dioxygen would likely bind in the vacant sixth site of the free enzyme. This position is near the hydrogen bonded $His^{241}$-$Tyr^{250}$ pair and especially close to $O^{\eta}$ of $Tyr^{250}$, which might stabilize the superoxide by an interaction involving the partial positive charge associated with the $O^{\eta}$ proton. In the case of mode B, dioxygen would likely bind in the water-3012 site near $His^{195}$. If $His^{195}$ is protonated on substrate binding, it would provide a positive charge to promote formation of the superoxide.
30. A. Hodel, S.-H. Kim, A. T. Brünger, *Acta Crystallogr. Sect. A* **48**, 851 (1992).
31. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
32. W. Kabsch, *ibid.* **21**, 916 (1988).
33. Z. Otwinowski, in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*, L. Sawyer, N. Issacs, S. Bailey, Eds. [Science and Engineering Research Council (SERC) Daresbury Laboratory, Warrington, UK, 1991)], pp. 56–62; W. Minor, *XDISPLAYF* (Purdue University, West Lafayette, IN, 1993).
34. *CCP4: A Suite of Programs for Protein Crystallography* (SERC Daresbury Laboratory, Warrington WA4 4AD, UK, 1979).
35. Z. Otwinowski, in *Proceedings of the CCP4 Study Weekend: Isomorphous Replacement and Anomalous Scattering*, W. Wolf, P. R. Evans, A. G. W. Leslie, Eds. (SERC Daresbury Laboratory, Warrington, UK, 1991), pp. 80–86.
36. J. T. Bolin, J. L. Smith, S. W. Muchmore, paper presented at the Annual Meeting of the American Crystallographic Association, Albuquerque, NM, 26 May 1993.
37. T. A. Jones, J.-Y. Zou, S. W. Cowan, M. Kjelgaard, *Acta Crystallogr. Sect. A* **47**, 110 (1991).
38. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).
39. A. T. Brünger, *X-PLOR, Version 3.1* (Yale Univ. Press, New Haven, CT, 1992).
40. We acknowledge the contributions of S. Harayama, B. Hofmann, and H.-J. Hecht to preliminary crystallization and diffraction studies. The assistance of R. K. Thauer and R. Hedderich in the initial anaerobic purification of DHBD is also gratefully acknowledged. Research in Braunschweig was supported in part by the German Ministry of Research and Technology. K.N.T. also thanks the Fonds der Chemischen Industrie for generous support. J.T.B. thanks L. Holm and G. Vriend for assistance with a preliminary evaluation of structural homology, and the Lucille P. Markey Foundation and the National Institutes of Health (NIH) (GM 52831) for support. Shared diffraction and computing facilities at Purdue University have been supported by grants from NIH and the National Science Foundation.

31 May 1995; accepted 23 August 1995

# Guidelines for Protein Design: The Energetics of β Sheet Side Chain Interactions

Catherine K. Smith and Lynne Regan*

**To determine the interaction energy between cross-strand pairs of side chains on an antiparallel β sheet, pairwise amino acid substitutions were made on the solvent-exposed face of the B1 domain of streptococcal protein G. The measured interaction energies were substantial (1.8 kilocalories per mole) and comparable to the magnitude of the β sheet propensities. The experimental results paralleled the statistical frequency with which the residue pairs are found in β sheets of known structure.**

A complete understanding of the construction of α helices and β sheets is essential for the manipulation of the structure and properties of proteins. A quantitative energetic description of many factors that determine the stability of α-helical secondary structure is available (*1–4*). Our understanding of the interactions that determine β sheet stability is much less advanced. Different amino acids do have measurably different intrinsic propensities for forming β sheets (*5*, *6*), and these intrinsic β sheet–forming propensities differ between central and edge strand positions in a β sheet (*7*). Statistical surveys of proteins of known structure reveal that a nonrandom pairwise distribution of amino acids occurs in adjacent positions across two β strands (*8–10*). This observation implies that in addition to the underlying differences in β sheet propensity, side chain interactions contribute to the local structure and stability of β sheets. Our study here measures the extent to which these side chain interactions contribute to β sheet stability.

A variant of the immunoglobulin G binding domain of streptococcal protein G (referred to as β1 and illustrated in Fig. 1A) was originally used to quantify the β sheet–forming propensities of the amino acids (*6*). Here, β1 was used to measure the energetics of side chain interactions between pairs of amino acids located in adjacent β strands. Pairwise amino acid substitutions were made on the solvent-exposed face of β1 in an antiparallel β sheet. The investigated site, positions 44 and 53, was selected to provide maximal solvent exposure for the studied side chains in the most homogeneous sheet environment available. Because the backbone atoms of amino acids 44 and 53 are hydrogen-bonded to one another, the site of mutation occurs in a hydrogen-bonded position (*8*), also called a narrow H-bond ring (*11*). This type of site has statistical pairwise preferences distinct from those of a non–H-bonded pair such as 53-6 (*8*). Pairwise mutants are referred to as β1XY, where X is the amino acid at position 44 and Y is the residue at position 53.

The side chain interactions between pairs of the best β sheet–forming residues (Trp, Phe, Tyr, Thr, Ile, and Val) and between pairs of charged residues were examined (*5–7*). The charged residues (Arg, Lys, and Glu) are among the intrinsically poorest β sheet formers, yet these residues would be useful for engineering solubility

Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520, USA.

*To whom correspondence should be addressed.