

# Evolution of Enzymatic Activity in the Enolase Superfamily: Structure of *o*-Succinylbenzoate Synthase from *Escherichia coli* in Complex with $\text{Mg}^{2+}$ and *o*-Succinylbenzoate<sup>†,‡</sup>

Thomas B. Thompson,<sup>§</sup> James B. Garrett,<sup>‡</sup> Erika A. Taylor,<sup>‡</sup> R. Meganathan,<sup>||</sup> John A. Gerlt,<sup>\*,‡</sup> and Ivan Rayment<sup>\*,§</sup>

Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53705, Departments of Biochemistry and Chemistry, University of Illinois, Urbana, Illinois 61801, and Department of Biological Sciences, Northern Illinois University, Illinois 60115

Received April 14, 2000

**ABSTRACT:** The X-ray structures of the ligand free (apo) and the  $\text{Mg}^{2+}$ •*o*-succinylbenzoate (OSB) product complex of *o*-succinylbenzoate synthase (OSBS) from *Escherichia coli* have been solved to 1.65 and 1.77 Å resolution, respectively. The structure of apo OSBS was solved by multiple isomorphous replacement in space group  $P2_12_12_1$ ; the structure of the complex with  $\text{Mg}^{2+}$ •OSB was solved by molecular replacement in space group  $P2_12_12$ . The two domain fold found for OSBS is similar to those found for other members of the enolase superfamily: a mixed  $\alpha/\beta$  capping domain formed from segments at the N- and C-termini of the polypeptide and a larger  $(\beta/\alpha)_7\beta$  barrel domain. Two regions of disorder were found in the structure of apo OSBS: (i) the loop between the first two  $\beta$ -strands in the  $\alpha/\beta$  domain; and (ii) the first sheet–helix pair in the barrel domain. These regions are ordered in the product complex with  $\text{Mg}^{2+}$ •OSB. As expected, the  $\text{Mg}^{2+}$ •OSB pair is bound at the C-terminal end of the barrel domain. The electron density for the phenyl succinate component of the product is well-defined; however, the 1-carboxylate appears to adopt multiple conformations. The metal is octahedrally coordinated by Asp<sup>161</sup>, Glu<sup>190</sup>, and Asp<sup>213</sup>, two water molecules, and one oxygen of the benzoate carboxylate group of OSB. The loop between the first two  $\beta$ -strands in the  $\alpha/\beta$  motif interacts with the aromatic ring of OSB. Lys<sup>133</sup> and Lys<sup>235</sup> are positioned to function as acid/base catalysts in the dehydration reaction. Few hydrogen bonding or electrostatic interactions are involved in the binding of OSB to the active site; instead, most of the interactions between OSB and the protein are either indirect via water molecules or via hydrophobic interactions. As a result, evolution of both the shape and the volume of the active site should be subject to few structural constraints. This would provide a structural strategy for the evolution of new catalytic activities in homologues of OSBS and a likely explanation for how the OSBS from *Amycolaptosis* also can catalyze the racemization of N-acylamino acids [Palmer, D. R., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., and Gerlt, J. A. (1999) *Biochemistry* 38, 4252–4258].

Members of the enolase superfamily share a similar structural scaffold but catalyze a remarkable range of chemical reactions including racemization,  $\beta$ -elimination of water or ammonia, and cycloisomerization (1). These

enzymes feature two domains, a mixed  $\alpha/\beta$  domain formed from the N- and C-termini and a larger  $(\beta/\alpha)_7\beta$  barrel domain. The former domain caps the barrel domain at the C-terminal ends of the  $\beta$ -strands; variability in both the sequences and the structures of these capping domains accounts for much of the differing substrate specificities among members of the superfamily. The structure of the latter is modified from that of other members of the triosephosphate isomerase (TIM)<sup>1</sup> superfamily that contain  $(\beta/\alpha)_8$  barrel motifs (3). Conserved functional groups required for formation of enolic intermediates via abstraction of the  $\alpha$ -proton of the carboxylate anion substrates are located at the C-terminal ends of the  $\beta$ -strands in the barrel domain (1). In all members of the superfamily, the binding site for

<sup>†</sup> This research was supported in part by NIH Grants AR35186 to I.R., GM-40570 to J.A.G., GM-52594 to J.A.G. and I.R., and GM 50262 to R.M. T.B.T. was supported by the NIH Biophysics Training Grant GM08293. 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 No. W-31-109-ENG-38.

<sup>‡</sup> The X-ray coordinates have been deposited in the Protein Data Bank with file names 1FHU and 1FHV.

<sup>\*</sup> To whom correspondence should be addressed at the Department of Biochemistry, 1710 University Avenue, Madison, WI 53705. Phone: (608) 262-0529. Fax: (608) 262-1319. E-mail: ivan\_rayment@biochem.wisc.edu and Department of Biochemistry, 600 South Mathews Avenue, University of Illinois, Urbana, IL 61801. Phone: (217) 244-7414. Fax: (217) 265-0385. E-mail: j-gerlt@uiuc.edu.

<sup>§</sup> University of Wisconsin.

<sup>‡</sup> University of Illinois.

<sup>||</sup> Northern Illinois University.

<sup>1</sup> Abbreviations: MIR, multiple isomorphous replacement; MR, mandelate racemase; MLE, muconate lactonizing enzyme; NAAAR, N-acylamino acid racemase; OSB, *o*-succinylbenzoate; OSBS, *o*-succinylbenzoate synthase; rms, root-mean-square; SHCHC, 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate; TIM, triose phosphate isomerase.

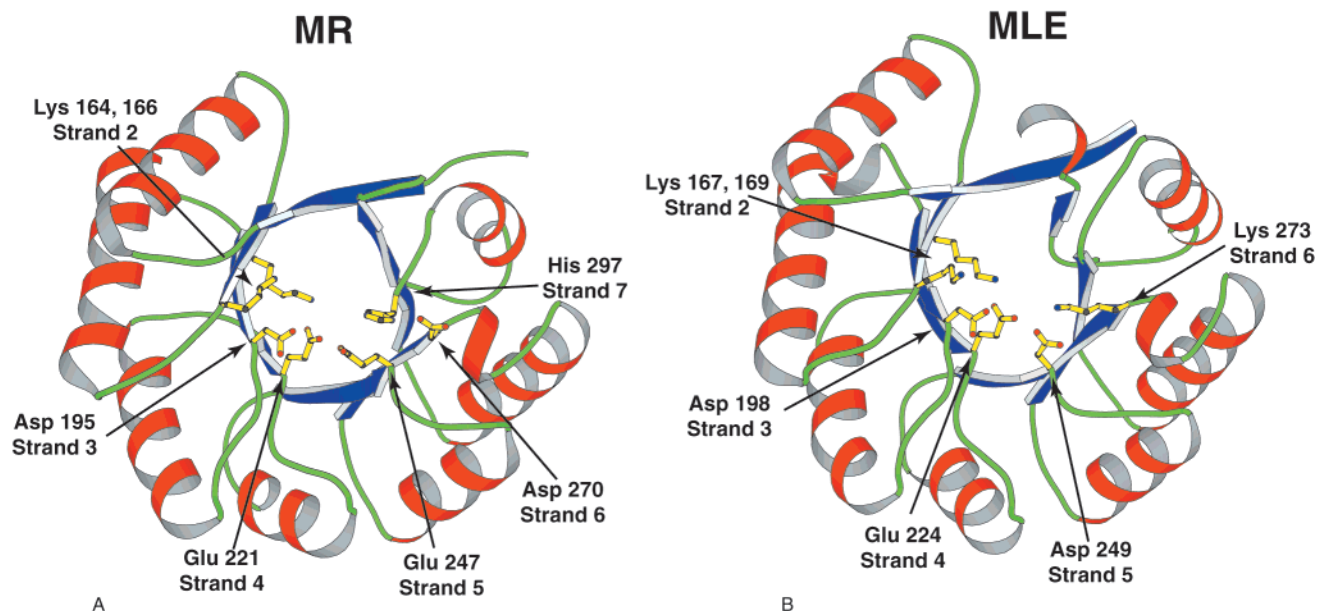


FIGURE 1: Ribbon representation of the barrel domains of (A) mandelate racemase (MR) and (B) muconate lactonizing enzyme (MLE) revealing the location of the key functional groups and their associated strands. The figure was prepared with the program Molscript (41) from the PDB files 1MDR and 1MUC (7, 14).

a divalent metal ion is formed by three, usually acidic, amino acids (4–6) located at the ends of the third, fourth, and fifth  $\beta$ -strands of the barrel domain. The catalytically essential metal ion assists in abstraction of the  $\alpha$ -proton of a carboxylate anion substrate by electrostatic stabilization of the enolic intermediate.

The members of the enolase superfamily can be organized into three subfamilies based on the identities of the base/acid residues utilized to abstract/donate a proton of the carboxylate anion substrate (1). These are designated the mandelate racemase (MR), muconate lactonizing enzyme (MLE), and enolase subfamilies based on the sequences and structures of these prototypic enzymes. The reactions catalyzed by members of the MR subfamily can utilize two acid/base catalysts to abstract the  $\alpha$ -proton of the substrate, e.g., Lys<sup>166</sup> and the hydrogen-bonded Asp<sup>270</sup>–His<sup>297</sup> dyad for MR; these are found at the C-terminal ends of the second (Lys<sup>166</sup>), sixth (Asp<sup>270</sup>), and seventh (His<sup>297</sup>)  $\beta$ -strands of the barrel domain (Figure 1A). The members of the MLE subfamily also can utilize two acid/base residues, e.g., Lys<sup>169</sup> and Lys<sup>273</sup> for MLE; these are found at the C-terminal ends of the second (Lys<sup>169</sup>) and sixth (Lys<sup>273</sup>)  $\beta$ -strands of the barrel domain (Asp<sup>270</sup> in MR and Lys<sup>273</sup> in MLE are sequentially and structurally homologous) (Figure 1B). In these positions, acid/base catalysts are located on opposite sides of the active site, allowing chemistry on either face of the enolic intermediate; for example, this design allows the 1,1-proton-transfer reaction catalyzed by MR. In contrast, enolases utilize a single residue (Lys<sup>345</sup> in yeast enolase) to abstract the  $\alpha$ -proton; this residue is found at the C-terminal end of the sixth  $\beta$ -strand of the barrel domain (homologous to Lys<sup>273</sup> in MLE and Asp<sup>270</sup> in MR).

The wealth of crystallographic information available for members of the enolase superfamily provided the framework for a careful analysis of the relationships between the structures and the sequences of the superfamily (1). The initial analysis was based on the structures of MR, MLE, and enolase (4, 5, 7, 8) and provided great insight into the

molecular architecture that underlies the enolase superfamily. Since the original study, several additional structures have been determined (9–11). Interestingly, until the present study, all structurally characterized members of the enolase superfamily possessed oligomeric quaternary structures: MR, MLE, and D-galactonate dehydratase are octamers (10); D-glucarate dehydratase is a tetramer; and yeast enolase is a dimer.

Structural studies of complexes with substrates, products, and analogues have played a major role in understanding the relationship between structure and function for this superfamily (4, 5, 7–9). Almost all of the structures contain a catalytically essential divalent metal ion (usually Mg<sup>2+</sup>) coordinated to the functional groups of three amino acids located individually at the ends of the third, fourth, and fifth  $\beta$ -strands in the barrel domain. The ligands are carboxylate groups, except in the case of D-glucarate dehydratases where an asparagine at the end of the fifth  $\beta$ -strand is a ligand. The metal ion is an integral part of the substrate-binding site, where the substrate is typically a bidentate ligand to the metal ion via either both carboxylate oxygens or one carboxylate oxygen and the  $\alpha$ -hydroxyl group. Comparisons of structures with and without metal ions reveal that metal ion binding induces little conformational change in the protein (7, 12); however, in each enzyme that has been studied in the absence and presence of an active site ligand (substrate or inhibitor), a loop in the capping domain becomes ordered when the ligand binds in the active site (8, 13–15). The ordering of this loop permits specific interactions with the substrate and shields the active site from the solvent.

Our premise is that functional and structural characterization of diverse members of the enolase superfamily will provide insights into how new functions have evolved from earlier templates as well as a guide for protein engineering and directed evolution. “New” members of the superfamily identified in sequence databases present the opportunity to expand the repertoire of reactions that can be catalyzed by

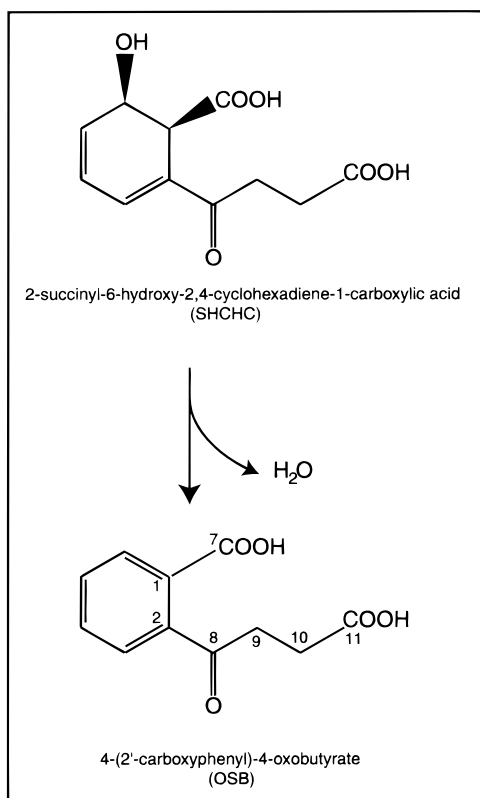


FIGURE 2: Reaction catalyzed by *o*-succinylbenzoate synthase. Selected carbon atoms of the product are labeled to indicate the numbering convention used in the discussion.

members of the superfamily. However, the sequence similarity relating different members of the superfamily typically is so low (<25%) that neither the substrates nor the reactions can be assigned directly for "new" members identified in genome sequencing projects. We expect that an understanding of the range of reactions that can be catalyzed will assist predictions of functions and likely will complement the comprehensive structure determination strategy that is the paradigm of structural genomics. The functions of several members of the enolase superfamily have been assigned by this method, including *o*-succinylbenzoate synthases (OSBS) from several bacterial species (1).

OSBS occurs in the menaquinone biosynthetic pathway in eubacteria, including *Escherichia coli* (16, 17). OSBS converts 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) to 4-(2'-carboxyphenyl)-4-oxobutyrates (*o*-succinylbenzoate or OSB) via  $\beta$ -elimination of water as shown in Figure 2. The reaction likely proceeds by a mechanism similar to those employed by other members of the enolase superfamily with initial abstraction of the  $\alpha$ -proton of the carboxylic acid and stabilization of the resulting enolic intermediate (1, 18). Alignment of the sequence of OSBS from *E. coli* with other members of the superfamily reveals the conserved presence of two lysine residues, Lys<sup>133</sup> and Lys<sup>235</sup>, either of which could be responsible for abstraction of the  $\alpha$ -proton. Accordingly, OSBS is classified as a member of the MLE subgroup. The sequence of *E. coli* OSBS also contains three totally conserved acidic residues, Asp<sup>161</sup>, Glu<sup>190</sup>, and Asp<sup>213</sup> that are predicted to be ligands for the required Mg<sup>2+</sup>.

In this paper, we report the structures of both the apo and the Mg<sup>2+</sup>•OSB product complex of OSBS from *E. coli*. The

identities of the acid/base catalysts and ligands for the divalent metal ion previously predicted by sequence alignments are confirmed by these structures. They also provide the second structure for a member of the MLE subfamily and the first with an active site ligand. The divergence in sequence that characterizes OSBSs from different eubacteria is analyzed in terms of the structural elements that form the active site of OSBS from *E. coli*. We expect that the structure for this OSBS, which does not catalyze racemization of N-acylamino acids, will be valuable in understanding and explaining the structural origins of the catalytic promiscuity displayed by the OSBS/NAAAR from *Amycolaptosis* (2).

## EXPERIMENTAL PROCEDURES

**Isolation of OSBS from *Escherichia coli*.** The gene encoding OSBS was amplified from *E. coli* strain MG1655 chromosomal DNA by PCR using *Pfu* DNA polymerase and inserted into a modified pET15b expression vector that encodes an N-terminal His-tag followed by a thrombin cleavage site. The gene was sequenced to verify that no mutations had occurred during the PCR amplification. The sequence of the recombinant OSBS utilized here differs from the sequence encoded by the genome by the addition of Gly-Ser-His sequence at the N-terminus.<sup>2</sup> The resulting plasmid was transformed into *E. coli* strain BL-21(DE3). The transformed cells were grown at 37 °C in LB medium for 18 h without induction by IPTG.

The cells were resuspended in 200 mL of binding buffer (20 mM Tris-HCl, pH 7.9, containing 5 mM imidazole and 0.5 M NaCl) and then disrupted by sonication with a Fisher Sonic Dismembrator model 550. The debris was removed by centrifugation, and the supernatant was loaded onto a Chelating Sepharose Fast-Flow charged with Ni<sup>2+</sup> and equilibrated with binding buffer. Extraneous proteins were removed from the column with 5 vol of 20 mM Tris-HCl, pH 7.9, containing 200 mM imidazole and 0.5 M NaCl. OSBS was then eluted with 20 mM Tris-HCl, pH 7.9, containing 1.0 M imidazole and 0.5 M NaCl. Fractions containing OSBS were identified by absorbance and SDS-PAGE, combined, and concentrated to 50 mL.

The His-tagged OSBS was dialyzed twice against 4 L of phosphate-buffered saline. The His-tag was cleaved using 1 U thrombin/mg OSBS. The reaction proceeded for 72 h at 25 °C and was monitored by SDS-PAGE. After the reaction reached completion, the solution was concentrated, dialyzed against 50 mM Tris-HCl, pH 8.0, containing 5 mM MgCl<sub>2</sub>, and filtered. An aliquot (100 mg of protein) was loaded onto a Resource-Q anion-exchange column and eluted with a 500-mL linear NaCl gradient (0 to 1 M NaCl in 50 mM Tris-HCl, pH 8.0). Fractions containing homogeneous OSBS were pooled, concentrated, and dialyzed against 10 mM Tris-HCl, pH 8.0, containing 5 mM MgCl<sub>2</sub>. The mass of the resulting OSBS (theoretical, 35 757 g/mol) was confirmed by electrospray mass spectroscopy (experimental, 35 755 ± 2 g/mol).

**Synthesis and Purification of OSB.** Crude OSB was prepared from phthalic anhydride and succinic anhydride according to a published procedure (19). An aqueous solution

<sup>2</sup> The sequence for *E. coli* OSBS used here was derived from g2507062, accession number P29208. This differs from the original sequence, g126970 by the inclusion of two extra residues after Ala174.



Table 1: Data Collection and MIR Statistics

data set	native <sup>a</sup>	LuCl <sub>3</sub> <sup>b</sup>	Au(CN) <sub>2</sub> <sup>b</sup>	uranyl acetate <sup>c</sup>	Me <sub>3</sub> HgCl <sup>d</sup>
conc (mM)		2.0	1.0	1.0	0.5
lgth of soak (days)		0.75	7	10	0.75
no. of crystals	1	1	1	1	1
resolution (Å)	2.2	2.3	2.2	2.5	2.4
average <i>I</i> /average $\sigma$	13.5 (3.0)	12.5 (2.2)	13.8 (3.5)	7.9 (2.4)	13.3 (3.1)
uniq reflect	15 567	13 461	14 836	10 623	12 242
redundancy	2.34	2.75	2.69	4.26	2.74
anomalous (%)		78	79	66	60
complete (%)	91.5 (70.1)	90.9 (69.3)	87.3 (61.9)	90.6 (63.9)	92.3 (73.6)
<i>R</i> <sub>merge</sub>	4.1 (13.3)	5.7 (19.1)	5.9 (15.5)	10.3 (23.1)	6.4 (16.1)
<i>R</i> <sub>iso</sub>		25.7	22.9	28.1	33.7
sites		3	2	3	4
phasing power <sup>e</sup>		0.81 (0.55)	1.07 (1.01)	1.10 (0.87)	1.01 (0.84)
figure of merit <sup>e</sup>	0.71 (0.62)				

<sup>a</sup> The values in parentheses are from resolution shells of 2.26–2.20 Å. <sup>b</sup> The values in parentheses are from resolution shells of 2.37–2.30 Å. <sup>c</sup> The values in parentheses are from resolution shells of 2.47–2.40 Å. <sup>d</sup> The values in parentheses are from resolution shells of 2.57–2.50 Å. <sup>e</sup> The values for the phasing power and figure of merit were calculated to 2.6 Å for all derivatives.

of the resulting oil was neutralized to pH 8 and purified by anion-exchange chromatography on a column of Dowex 1 × 8-400 using a linear gradient of 0 to 1 M LiCl. OSB was precipitated as the bislithium salt by addition of 10:1 acetone/methanol.

**Crystallization and X-ray Data Collection of Apo OSBS.** Crystals of apo OSBS were grown by microbatch from 27% methyl ether poly(ethylene glycol) 5000, 267 mM NaCl, and 50 mM MES, pH 6.75, at room temperature with a protein concentration of 15–20 mg/mL. Microbatch experiments were nucleated by streak-seeding and produced large single crystals within 14 to 21 days. Two crystal morphologies were observed under these conditions; orthorhombic rods with pointed ends and plates that grew in clusters. Only crystals with the rodlike morphology were utilized in this study. These crystals achieved a size of 1.0 × 0.5 × 0.3 mm with diffraction maxima observed beyond 2.2 Å resolution. The crystals belong to the space group *P*2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> with unit cell dimensions of *a* = 56.1 Å, *b* = 70.0 Å, *c* = 80.4 Å. The crystal lattice contained one molecule in the asymmetric unit with a solvent content of 45%. Crystals were transferred to a synthetic mother liquor of 30% methyl ether poly(ethylene glycol) 5000, 400 mM NaCl, and 50 mM MES at pH 6.75 for all noncryogenic X-ray data collection and heavy atom preparation. Crystals were stable for more than 2 weeks in the synthetic mother liquor. For the high-resolution structure, OSBS crystals were transferred in a stepwise manner to a final cryoprotectant solution consisting of 32% methyl ether poly(ethylene glycol) 5000, 400 mM NaCl, 10% ethylene glycol, and 50 mM MES at pH 6.75. The crystal remained in the final solution for 15 s prior to freezing. The thicker crystals survived the transfer much better than the thin ones and maintained a low mosaic spread (0.25°). Upon freezing the crystals, the unit cell dimensions changed to *a* = 55.4 Å, *b* = 69.0 Å, *c* = 79.3 Å.

X-ray data were collected to 2.2 Å resolution at –3 °C with a Siemens HiStar area detector at a crystal to detector distance of 12 cm. Cu K $\alpha$  radiation was generated by a Rigaku RU2000 X-ray generator operated at 50 kV and 80 mA and equipped with Siemens Göbel focusing optics. Diffraction data were collected in 0.15° wedges and were processed with XDS (20) and internally scaled with XSCALIBRE (21). Data collection statistics for the native and heavy atom derivatives are shown in Table 1.

Table 2: High Resolution Data Collection Statistics for the *o*-Succinyl Benzoate Co-Crystal Form and the Apo Crystal Form

data set	<i>o</i> -succinyl benzoate <sup>a</sup>	native <sup>b</sup>
conc (mM)	2 mM Mg <sup>2+</sup> •OSB	
space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
unit cell (Å)	<i>a</i> = 71.8, <i>b</i> = 83.0, <i>c</i> = 57.4	<i>a</i> = 55.4, <i>b</i> = 69.0, <i>c</i> = 79.3
no. of crystals	1	1
resolution (Å)	1.77	1.65
average <i>I</i> /average $\sigma$	16.31(12.3)	14.1 (11.0)
uniq reflections	33 959	36 924
redundancy	9.4	11.4
complete (%)	99.8 (99.1)	99.0 (98.4)
<i>R</i> <sub>merge</sub>	4.5 (11.2)	6.9 (29.4)

<sup>a</sup> The values in parentheses are from resolution shells of 1.83–1.77 Å. <sup>b</sup> The values in parentheses are from resolution shells of 1.71–1.65 Å.

High-resolution data to 1.45 Å resolution were collected at the Advanced Photon Source on 19-ID beamline of the Structural Biology Center at Argonne National Laboratory. Two diffraction data scans with exposure times of 3 and 8 s and crystal-to-detector distances of 150 and 120 mm, respectively, were collected with a frame size of 1° and a wavelength of  $\lambda$  = 0.9537 Å. Data were processed with HKL2000 and internally scaled with SCALEPACK (22, 23). Data were truncated to 1.65-Å resolution based on the quality of the scaling. Data collection statistics are shown in Table 2.

**OSBS Structure Determination.** The three-dimensional structure of apo OSBS was determined by the technique of multiple isomorphous replacement (MIR). All attempts to produce a molecular replacement structure with MLE I as the search model failed, probably due to structural differences in the variable capping domains. Four heavy atom derivatives were prepared by soaking crystals in synthetic mother liquor containing 2.0 mM lutetium chloride, 1.0 mM gold cyanide, 1.0 mM uranyl acetate, or 0.5 mM methyl mercury chloride. Heavy atom binding sites were located by appropriate difference Patterson maps and refined with the program HEAVY (24, 25). The correct hand of the heavy atom constellation was chosen from the anomalous difference Fouriers. Difference maps were generated with solvent flattened phases from a single derivative. These maps were examined to confirm heavy atom sites in the other three

Table 3: Refinement Statistics<sup>a</sup>

	<i>o</i> -succinyl benzoate	apo
resolution limits (Å)	30–1.77	30.0–1.65
final <i>R</i> -factor <sup>b</sup>	18.5	19.5
final <i>R</i> -free <sup>b</sup>	22.4	26.5
no. of reflections used	33 508	36 924
no. of protein atoms	2492	2292
no. of solvent molecules	335	351
average <i>B</i> value	15.1	16.0
mainchain atoms		
all protein atoms	20.1	19.0
solvent atoms	30.4	31.0
Mg <sup>2+</sup> atom	17.5	
OSB molecule	36.5	
weighted rms deviations		
from ideality		
bond lengths (Å)	0.014	0.013
bond angles (deg)	2.03	2.730
planarity (trigonal) (Å)	0.009	0.007
planarity (others) (Å)	0.021	0.021
torsional angles (deg) <sup>c</sup>	16.03	16.11

<sup>a</sup> TNT refinement. <sup>b</sup>  $R = \sum ||F_o| - k|F_c|| / \sum |F_o|$ . *R*-free was calculated with 5% of the omitted *F*<sub>o</sub>. *R*-free for the OSB structure was calculated to 1.9 Å. <sup>c</sup> No restraints were placed on torsional angles during refinement.

derivatives and to locate any minor sites that may have been missed in the original solution. Table 1 shows the heavy atom data collection and refinement statistics for the four heavy atoms. The program SOLVE was utilized to generate initial phases giving a figure of merit of 0.71 (26). The resulting phases were modified with the solvent flattening program DM, which increased the figure of merit to 0.91 (27, 28).

An electron density map to 2.4-Å resolution was calculated with the density-modified phases. Continuous regions of electron density were lined with the program BONES (29). Examination of the BONES model indicated that OSBS contained of a fold similar to that found in MLE I. The structure of MLE I [PDB entry 1MUC (7)] was manually superimposed on the bones molecular model and used as a guide to build a model for OSBS. Only the most well-defined regions of the electron density were built that consisted of ~75% of the total mass. At that time, 5% of the native data was removed for an *R*<sub>free</sub> calculation. The model was subjected to several rounds of least squares refinement (TNT) (30) and manual model building with the program TURBO FRODO (31). The *R*-factor was reduced to 21% for 87.5% of the structure factors.

The final model from the MIR data was positioned in the frozen cell with the program AMORE to generate a starting model for the high-resolution refinement (27, 32). The model was refined with least squares refinement after 5% of the structure factors were removed for an *R*<sub>free</sub> calculation. The *R*-factor started at 31% before manual model building. After a few rounds of manual model building and least squares refinement, the *R*-factor dropped to 26%. Water molecules were added manually and with the program TNT (30). This improved the quality of the electron density map and several more residues along with several side chains absent in the model were added. The final refinement statistics are found in Table 3. The current *R*-factor is 19.5% for 94% of the data with an *R*<sub>free</sub> of 26.5% for all data from 30 to 1.65 Å. The rms deviations for bond length, bond angle, and groups of atoms expected to be coplanar are 0.013 Å, 2.73°, and 0.007 Å respectively. A representative section of the electron

density is shown in Figure 3. The X-ray coordinates have been deposited in the Protein Data Bank with file name 1FHU.

**Crystallization and Data Collection of OSB Cocrystals.** Crystals grown without substrate or product were soaked in 5 mM Mg<sup>2+</sup>•OSB overnight. Data collected from these crystals revealed electron density for the metal ion (unpublished results) but not for the product in the putative active site. Therefore, cocrystals were grown in the presence of both Mg<sup>2+</sup> and OSB. The largest crystals were grown by microbatch from 5% poly(ethylene glycol) 4000, 3% ethylene glycol, 15 mM MgCl<sub>2</sub>, 1 mM OSB, and 25 mM MES, pH 5.4, at room temperature with a protein concentration of 20 mg/mL. Microbatch experiments were nucleated by streak-seeding with a cat whisker (courtesy Harvey Thompson, fat domestic shorthair). Crystals grew as thin plates with dimensions of 0.25 × 0.25 × 0.03 mm. For data collection at −160 °C, crystals were transferred stepwise to a final cryoprotectant solution of 20% poly(ethylene glycol) 4000, 15% ethylene glycol, 50 mM MgCl<sub>2</sub>, 80 mM NaCl, 2 mM OSB, and 50 mM MES, pH 5.4. Crystals belong to the space group *P*2<sub>1</sub>2<sub>1</sub>2 with unit cell dimensions of *a* = 71.8 Å, *b* = 83.0 Å, *c* = 57.4 Å.

High-resolution data were collected at the Advanced Photon Source on the 19-ID beamline of the Structural Biology Center at Argonne National Laboratory. Diffraction maxima were observed to greater than 1.77 Å resolution. Two diffraction scans were collected with exposure times of 2 and 3 s and a crystal-to-detector distance of 250 and 325 mm, respectively. Both scans were collected at a wavelength of  $\lambda = 0.70090$  Å and a frame size of 1°. Data were processed with HKL2000 and internally scaled with SCALEPACK (22, 23). Data collection statistics are presented in Table 2.

**OSB Structure Determination and Refinement.** A molecular replacement solution was identified with CNS (version 0.5) using the apo OSBS structure as a search model (33). All subsequent refinement was carried out with CNS version 0.5 except as noted (33). Rigid body refinement yielded an *R*-factor of 48.4% and an *R*-free of 48.9%. The model was subjected to 200 steps of slow cooled simulated annealing from a starting temperature of 2500 K. After annealing, the *R*-factor and *R*-free dropped to 33.7 and 36.5%, respectively. Conjugate gradient minimization and grouped temperature factor refinement lowered the *R*-factor and *R*-free to 27.4 and 29.5%, respectively. The model was subjected to manual building and refinement with the aid of  $\sigma_a$  2*F*<sub>o</sub> − *F*<sub>c</sub> and *F*<sub>o</sub> − *F*<sub>c</sub> electron density maps (34). Electron density was visible for the two regions of disorder in the apo OSBS structure (Ala<sup>15</sup> to Arg<sup>23</sup>, Pro<sup>114</sup> to Gly<sup>126</sup>). After manual refinement and addition of solvent molecules, the *R*-factor and *R*-free dropped to 22.8 and 25.2%, respectively. The metal-product complex, Mg<sup>2+</sup>•OSB, was added last. Figure 3b shows the  $\sigma_a$  weighted omit electron density for OSB. The final model was refined with individual temperature factors with the program TNT to a final *R*-factor of 18.5% for all data from 30 to 1.77 Å resolution (30). The final refinement statistics for the Mg<sup>2+</sup>•OSB structure are found in Table 3. The rms deviations for the bond lengths, bond angles, and groups of atom expected to be planar are 0.014 Å, 2.03°, and 0.009 Å, respectively. The X-ray coordinates of the Mg<sup>2+</sup>•OSB complex have been deposited in the Protein Data Bank with

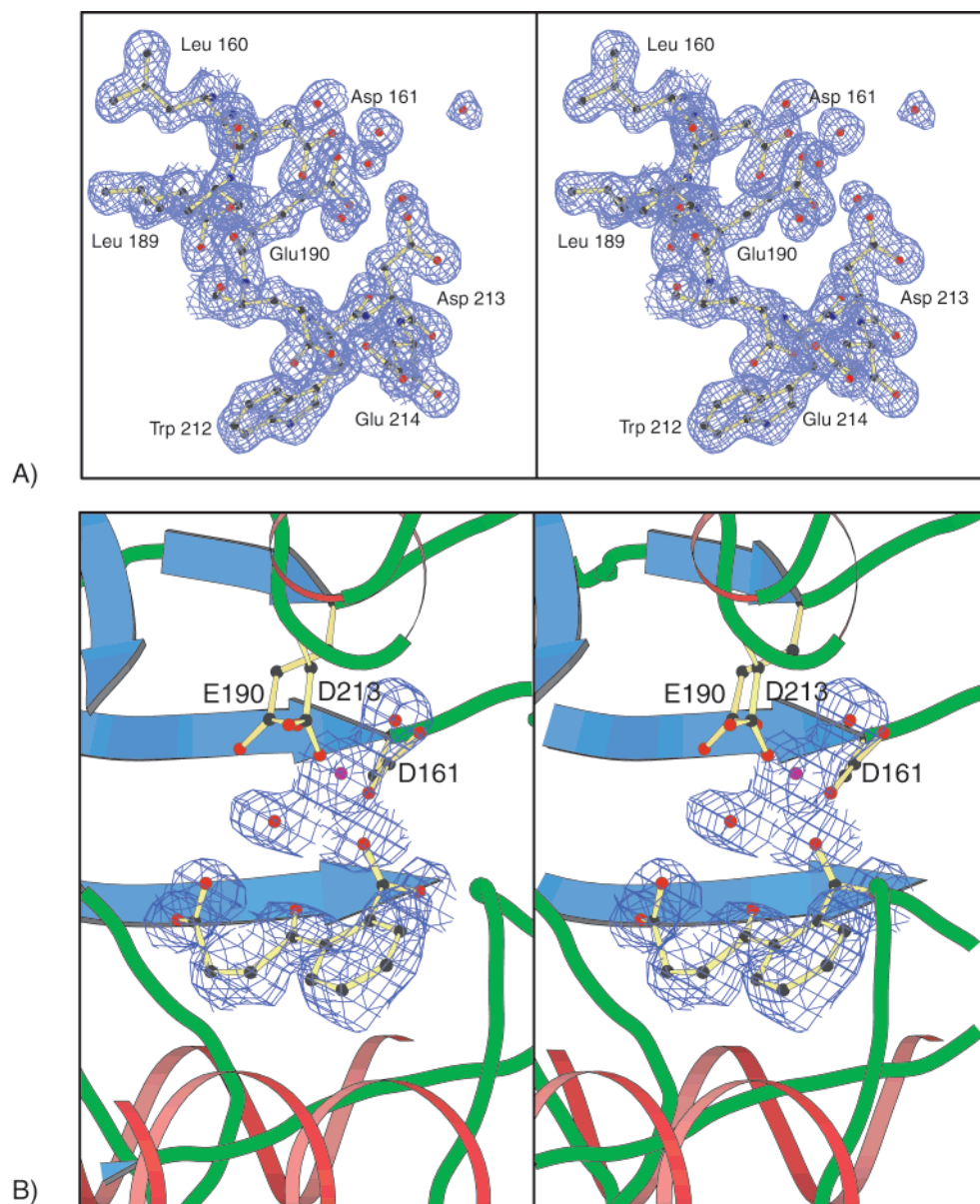


FIGURE 3: Stereoview of representative electron density: (A) The well-defined density for the three acidic metal amino acid ligands Asp<sup>161</sup>, Glu<sup>190</sup>, and Asp<sup>213</sup> in the apo structure. Coefficients for the electron density were of the form  $2F_o - F_c$ . (B) An omit map of the OSB molecule with Mg<sup>2+</sup> bound in the active site. Coefficients for the electron density were generated with  $\sigma_a$  weighting and were of the form  $2F_o - F_c$ . The figure was prepared with the program BOBSCRIPT (42).

the file name 1FHV.

## RESULTS AND DISCUSSION

The structure of OSBS in its apo form and in the presence of Mg<sup>2+</sup>•OSB were determined to 1.65 and 1.77 Å resolution, respectively. In both cases, OSBS is a monomer and crystallizes with one polypeptide chain in the asymmetric unit. The N-terminal Gly–Ser–His sequence that was appended to the 320-amino acid polypeptide encoded by the *menC* gene is disordered in the apo structure; the residue numbers that follow refer to those in the sequence of the genome-encoded protein.

The final high-resolution model of apo OSBS contains 298 of 323 residues with 350 water molecules and no metal ions. Representative electron density of the water structure and the metal ligands without metal bound is shown in Figure 3. The electron density is well-ordered except for two

regions, Ala<sup>15</sup> to Arg<sup>23</sup> and Pro<sup>114</sup> to Gly<sup>126</sup>. The first disordered section connects the first two  $\beta$ -strands of the  $\alpha/\beta$  domain. In most members of the superfamily studied to date, a loop provided by the capping domain undergoes a disorder/order transition during substrate binding (15, 35). This transition serves to shield the substrate and the active site from the bulk solvent. The disorder of residues Pro<sup>114</sup> to Gly<sup>126</sup> accounts for the absence of the first  $\alpha$ -helix of the barrel domain. In addition, no electron density is visible beyond the C $\beta$  for 10 other residues that have been modeled as Ala: Leu<sup>24</sup>, Lys<sup>25</sup>, Glu<sup>36</sup>, Glu<sup>38</sup>, Arg<sup>39</sup>, Glu<sup>54</sup>, Arg<sup>105</sup>, Asn<sup>111</sup>, Asp<sup>113</sup>, and Leu<sup>136</sup>. The electron density for atoms C $\delta$ , C $\epsilon$ , and N $\zeta$  of Lys<sup>133</sup> also is missing.

The final model of the Mg<sup>2+</sup>•OSB product complex contained 322 of 323 residues, 355 solvent molecules, and Mg<sup>2+</sup>•OSB. The two regions that are disordered in the apo structure are clearly present within the electron density. The



loop between Ala<sup>15</sup> to Arg<sup>23</sup> becomes ordered upon interaction with the benzene ring of the OSB molecule. The segment from Pro<sup>114</sup> to Gly<sup>126</sup> also is ordered and forms the first helix of the barrel domain. Electron density was not observed for the side chains of Asp<sup>21</sup>, Glu<sup>38</sup>, Glu<sup>127</sup>, Gln<sup>173</sup>, and Arg<sup>204</sup> and were all modeled as Ala. Similar to the apo structure, the electron density for atoms C<sub>δ</sub>, C<sub>ε</sub>, and N<sub>ζ</sub> of Lys<sup>133</sup> is missing.

The geometry of the models were analyzed with PROCHECK and TURBO-FRODO (31, 36). For the apo structure, 93% of the amino acids are in the allowed Ramachandran regions, whereas in the Mg<sup>2+</sup>•OSB complex, 91% of the residues are in the allowed regions. For both structures, examination of the Ramachandran plot revealed two residues that fall in more generously allowed regions. In the apo structure, these are Asp<sup>213</sup> ( $\Phi$  -140°;  $\Psi$  -81°) and Val<sup>300</sup> ( $\Phi$  +67°;  $\Psi$  -54°). The conformational angles of these residues in the Mg<sup>2+</sup>•OSB complex are essentially identical. These residues, which have anomalous conformational angles, are in sections of the model that have unambiguously defined electron density. The dihedral angles of Asp<sup>213</sup>, one of the metal ligands, are restricted by Trp<sup>212</sup>. Trp<sup>212</sup>, which has well-defined density, resides in a hydrophobic pocket and is conformationally restricted by a hydrogen bond with Glu<sup>226</sup>. As noted later, the position and orientation of Asp<sup>213</sup> is important for the integrity of the metal binding site. Val<sup>300</sup> is found in the beginning of C-terminal section of the capping domain. The dihedral angles are forced into their observed values by a well-defined salt bridge between Arg<sup>301</sup> and Glu<sup>93</sup>. This interaction inhibits  $\Psi$  from moving to a more favorable positive value. Also, the side chain of Val<sup>300</sup> points toward a small hydrophobic pocket formed by Trp<sup>42</sup> and Val<sup>313</sup> that may limit the freedom of  $\Phi$ . Although the exact thermodynamic reason for these unusual dihedral angles is difficult to assess, the presence of such conformational angles in an active site appears justified when they provide a structural or catalytic advantage to the protein (37).

A ribbon representation of OSBS (with Mg<sup>2+</sup>•OSB in the active site) is shown in Figure 4. Overall the fold of OSBS is very similar to those of all other members of the superfamily (1). To simplify comparison of this structure with other members of the superfamily, the secondary structural elements will be designated by their sequential location in either the capping or barrel domain rather than in the linear sequence. The secondary structural elements of the capping domain elements are subscripted with c whereas those of the barrel domain are subscripted with b as indicated in Figure 5. The boxed labels in Figure 5 indicate approximately where the secondary structural elements of the barrel differ from those observed in MLE (7).

**Domains of the OSBS Monomer.** As observed for all other structurally characterized members of the superfamily, OSBS contains two domains: a mixed  $\alpha/\beta$  capping domain and a larger ( $\beta/\alpha$ )<sub>7</sub> $\beta$  barrel domain. The active site is located at the C-terminal end of the barrel in close proximity to the capping domain. The barrel domain consists of residues Pro<sup>108</sup> to Leu<sup>294</sup>, and the capping domain consists of residues Met<sup>1</sup> to Ala<sup>107</sup> and Met<sup>295</sup> to Leu<sup>320</sup>. The capping domain covers more than half of the barrel at the C-terminal ends of its  $\beta$ -strands (Figure 4B). Most of the interdomain interactions come from the N-terminal segment of the

capping domain. Two loops in the capping domain interact with the C-terminal end of the barrel domain. First, the loop connecting  $\beta_{c1}$  and  $\beta_{c2}$  undergoes a disorder/order transition upon binding of product. This loop is disordered in the apo structure but becomes ordered when product (and presumably substrate) occupies the active site as is observed in enolase (15). The second loop, between  $\beta_{c3}$  and  $\alpha_{c1}$  or Pro<sup>47</sup>–Thr<sup>55</sup>, is shorter and has well-defined density in both structures. In GlucD, the loop that connects  $\alpha_{c2}$  and  $\alpha_{c3}$  undergoes the disorder/order transition during catalysis (35).

**Capping Domain.** The capping domain is built from elements at both the N- and C-termini of the polypeptide chain. At the N-terminus, the capping domain starts with three antiparallel strands that are common to many members of the enolase superfamily. As mentioned earlier, the loop between  $\beta_{c1}$  and  $\beta_{c2}$  becomes ordered upon product binding and presumably is ordered when substrate is bound. This loop is highly hydrophobic at its N-terminus with the sequence AGVVL and highly charged and flexible at its C-terminus with the sequence RDRR. The loop forms an apex around Leu<sup>19</sup> and Arg<sup>20</sup> that points into the active site. The N-terminal region of the loop forms a hydrophobic interaction with the C-terminal end of the barrel domain. The available structures for MLE demonstrate that metal binding does not order this loop (7). Likewise, OSBS from *E. coli* does not appear to change significantly when metal binds in the active site (unpublished results). A second loop between  $\beta_{c3}$  and  $\alpha_{c1}$  from residues Pro<sup>47</sup>–Thr<sup>55</sup> partially caps the C-terminal end of the barrel. This loop is maintained in the proximity of the active site by an interaction between  $\beta_{c3}$  and the loop connecting  $\beta_{b7}$  and  $\alpha_{b7}$  or Ser<sup>264</sup>–Gly<sup>270</sup> (Figure 5) and involves regions that are conserved among

<sup>3</sup> Sequence data for Figure 6 were derived from the following sources: *A. actinomycetem*: University of Oklahoma *Actinobacillus actinomycetemcommittans* genome sequencing project; accessed via <http://www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html>. *Amycolaptosis*: *Amycolaptosis* sp.; PIR I39598 and ref 1. *B. anthracis*: TIGR *Bacillus anthracis* genome sequencing project; accessed via <http://www.tigr.org/cgi-bin/BlastSearch/blast.cgi?B.subtilis:Bacillus subtilis>; EMBL AL009126 and ref 1. *C. diphtheria*: Sanger *Corynebacterium diphtheriae* genome sequencing project; accessed via [http://www.sanger.ac.uk/Projects/C\\_diphtheriae/blast\\_server.shtml](http://www.sanger.ac.uk/Projects/C_diphtheriae/blast_server.shtml). *C. tepidum*: TIGR *Chlorobium tepidum* genome sequencing project; accessed via <http://www.tigr.org/cgi-bin/BlastSearch/blast.cgi?E.coli:Escherichia coli>; GenBank g2507062, accession number P29208. *E. faecalis*: TIGR *Enterococcus faecalis* genome sequencing project; accessed via <http://www.tigr.org/cgi-bin/BlastSearch/blast.cgi?H.influenzae:Hemophilus influenzae>; accession number P44961. *M. avium*: TIGR *Mycobacterium avium* sequencing project; accessed via [http://www.ncbi.nlm.nih.gov/Microb\\_blast/unfinishedgenome.html](http://www.ncbi.nlm.nih.gov/Microb_blast/unfinishedgenome.html). *M. leprae*: Sanger Center *Mycobacterium leprae* sequencing project; accessed via [http://www.sanger.ac.uk/Projects/M\\_leprae/blast\\_server.shtml](http://www.sanger.ac.uk/Projects/M_leprae/blast_server.shtml). *M. tuberculosis*: *Mycobacterium tuberculosis* H37Rv; EMBL AL123456. *P. gingivalis*: TIGR *Porphyromonas gingivalis* genome sequencing project; accessed via <http://www.tigr.org/cgi-bin/BlastSearch/blast.cgi?P.multocida>; University of Minnesota *Pasteurella multocida* genome sequencing project; accessed via <http://www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html>. *S. aureus*: *Staphylococcus aureus*; GenBank U51132. *S. putrefaciens*: TIGR *Shewanella putrefaciens* genome sequencing project; accessed via <http://www.tigr.org/cgi-bin/BlastSearch/blast.cgi?S.typhi>; Sanger Center *Salmonella typhi* genome sequencing project; accessed via [http://www.sanger.ac.uk/Projects/S\\_typhi/blast\\_server.shtml](http://www.sanger.ac.uk/Projects/S_typhi/blast_server.shtml). *Synechocystis*: *Synechocystis* PCC6803; GenBank D64001 and ref 2. *V. cholerae*: TIGR *Vibrio cholerae* genome sequencing project, accessed via <http://www.tigr.org/cgi-bin/BlastSearch/blast.cgi?Y.pestis>; Sanger Center *Yersinia pestis* genome sequencing project; accessed via [http://www.sanger.ac.uk/Projects/Y\\_pestis/blast\\_server.shtml](http://www.sanger.ac.uk/Projects/Y_pestis/blast_server.shtml).

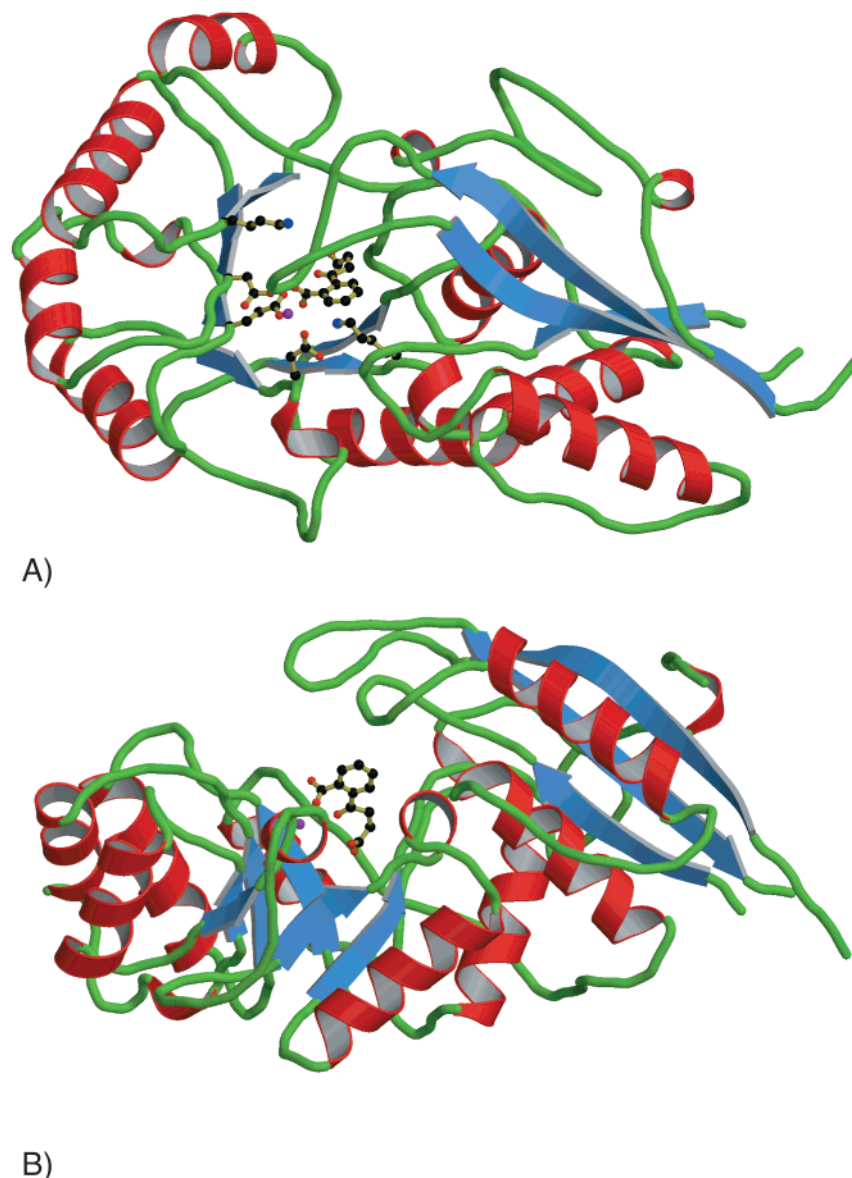


FIGURE 4: Ribbon representation of OSBS. (A) View looking down the  $\beta$ -barrel with the capping domain on the right. The three acidic metal ligands along with the putative active site lysine general acid/bases are represented as ball-and-stick. (B) View rotated 90° on the horizontal axis such that the active site is pointed up and the capping domain remains positioned to the right. The OSB molecule is represented as ball-and-stick. This figure was prepared with the program MOLSCRIPT and RASTER3D (41, 43).

the amino acid sequences of 20 orthologous bacterial OSBSs that can now be identified in the sequence databases (Figure 6).<sup>3</sup> Similar to other members of the enolase superfamily, this loop interacts directly with the product in the active site. After the three  $\beta$ -strands, the chain proceeds through two more  $\alpha$ -helices ( $\alpha_{c1}$  and  $\alpha_{c2}$ ) before an extended section of coil connects the capping domain to the barrel domain.

The C-terminal segment of the capping domain is highly variable among the orthologous OSBSs (Figure 6). In OSBS from *E. coli*, this segment consists of the final 26 residues from Met<sup>295</sup> to Leu<sup>320</sup> and is mostly random coil except for a helical turn shown as  $\alpha_{c3}$  in Figure 5.

The capping and barrel domains interact via Glu<sup>44</sup> that hydrogen bonds to the two backbone amide hydrogens of Ser<sup>267</sup> and Gly<sup>270</sup> and the side chain of Ser<sup>267</sup> in the barrel domain. Alignment of the sequences of the orthologous OSBSs (Figure 6) suggests that a hydroxylic functional group is required at the position equivalent to Ser<sup>267</sup>. In the Mg<sup>2+</sup>•OSB complex, the hydroxyl group of Ser<sup>267</sup> hydrogen

bonds to the absolutely conserved Glu<sup>44</sup>. As will be discussed later, Ser<sup>267</sup> also forms part of the active site and interacts with OSB. Conversely, a small flexible amino acid side chain is required at residue 270 (Gly<sup>270</sup> in this structure) to allow hydrogen bonding from the backbone. Two hydrophobic residues contained in the Glu<sup>44</sup> loop, Leu<sup>48</sup> and Phe<sup>51</sup>, point toward the active site and are within 4 Å of the product OSB. The residues equivalent to Leu<sup>48</sup> in the orthologous OSBSs are large hydrophobic residues, leucine, phenylalanine, or methionine; the homologues of Phe<sup>51</sup> are aromatic with a single Leu exception (Figure 6). As will be discussed later, these residues participate in formation of a hydrophobic pocket for the benzene ring portion of OSB (and, presumably the cyclohexadienyl ring of the SHCHC substrate).

**Barrel Domain.** The  $(\beta/\alpha)_7\beta$  barrel domain is structurally similar to members of the triosephosphate isomerase (TIM) superfamily of  $(\beta/\alpha)_8$  barrel motifs (38). However, in all of the members of the enolase superfamily characterized to date the barrel deviates from the standard  $(\beta/\alpha)_8$  barrel by



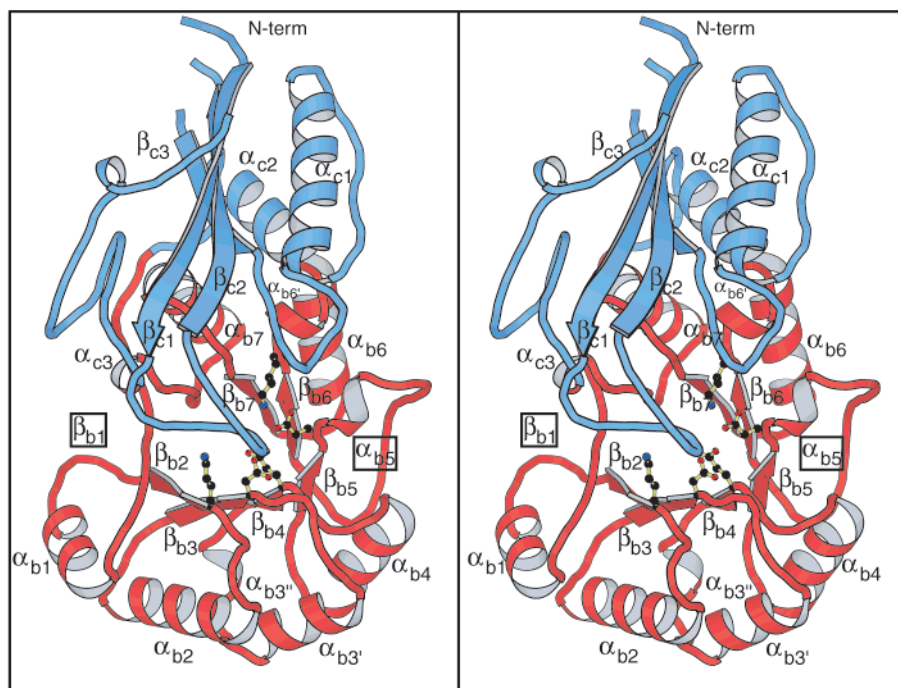


FIGURE 5: Stereoview of a ribbon representation of OSBS where the secondary structural elements have been labeled. The capping domain is colored blue, and the secondary structural elements are denoted by a “c” subscript, whereas the barrel domain is colored red, and the secondary structural elements are denoted with a “b” subscript. Secondary structural elements enclosed within boxes are different from those observed in MLE and are indicative of structural divergence within the enolase superfamily. These are marked at the locations where the corresponding elements are found in the structure of MLE. Figures 5 and 7–9 were prepared with MOLSCRIPT (41).

replacing the last helix with chain elements that interact with the capping domain.

Unlike other structures in the enolase superfamily, the barrel domain of apo OSBS contains disordered secondary elements. In many  $(\beta/\alpha)_8$  barrels a tendency for a discontinuity is observed between  $\beta_{b1}$  and  $\beta_{b8}$ ; in the case of OSBS this is extreme, leading to a “split barrel”. In most  $(\beta/\alpha)_8$  barrels, three residues from each  $\beta$ -strand typically contribute to the hydrogen-bonding pattern of the barrel. In the  $\text{Mg}^{2+}$ •OSB product complex,  $\beta_{b1}$  makes only one hydrogen bond between the carbonyl oxygen of Pro<sup>108</sup> and the amide hydrogen on Lys<sup>131</sup>, with other interactions occurring between side chains. This results in the observed splitting of the barrel. The apo structure has only six ordered  $\beta$ -strands and six ordered  $\alpha$ -helices.  $\beta_{b1}$  and  $\alpha_{b1}$  are not well-ordered in the apo structure, possibly due to a lack of substrate or product binding or crystal packing interactions. Helix  $\alpha_{b3}$  is broken into two segments,  $\alpha_{b3'}$  and  $\alpha_{b3''}$ , by a kink created by Pro<sup>180</sup>. This structural characteristic may be restricted to the OSBS from *E. coli* since this region of the sequence is highly variable in the orthologous OSBSs (Figure 6). All other elements of the barrel are characteristic of the  $(\beta/\alpha)_8$  barrel motif except for  $\alpha_{b5}$  and  $\beta_{b8}$ . Helix  $\alpha_{b5}$  consists only of a single helical turn followed by a large loop connecting  $\alpha_{b5}$  to  $\beta_{b6}$ . Likewise, the extended polypeptide chain from residue Ile<sup>286</sup> to Leu<sup>292</sup> exhibits very little  $\beta$ -strand characteristic with only one true strand–strand interaction with  $\beta_{b7}$  but is classified as  $\beta_{b8}$  by its spatial orientation in the barrel. Strand  $\beta_{b8}$  also may be fragmented in some but not all of the orthologous OSBSs as judged by the highly conserved Pro<sup>287</sup>–Gly<sup>288</sup> sequence (Figure 6) that comes in close proximity to the active site. After Leu<sup>292</sup> the chain becomes slightly helical and even completes one helical turn before advancing into the C-terminal segment of the capping

domain. This could be a remnant of  $\alpha_{b8}$  and would complete the more common  $(\beta/\alpha)_8$  barrel that is disrupted by its interaction with the C-terminal segment of the capping domain.

**Domain Movement upon OSB Binding.** The structure of the  $\text{Mg}^{2+}$ •OSB product complex reveals that the angle between the capping and  $\beta$ -barrel domains is increased relative to that in the apo structure by  $\sim 10^\circ$ . This change gives the impression that the substrate/product binding induces an opening of the domains even though two sections of the polypeptide chain become more ordered in the OSB cocrystal lattice. A superposition of the structures of the apo and  $\text{Mg}^{2+}$ •OSB bound enzymes with the capping domain used as a reference is shown in Figure 7. The rms difference of 290 C $\alpha$  pairs within the entire model is 1.4 Å with a maximum deviation of 4.4 Å. The largest difference occurs in a translation of 3.7 and 3.5 Å in  $\alpha_{b2}$  and  $\alpha_{b3}$ , respectively. The C-terminal ends of  $\beta_{b2}$ ,  $\beta_{b3}$ , and  $\beta_{b4}$ , where independent active site groups are located (Lys<sup>133</sup>, Asp<sup>161</sup>, and Glu<sup>190</sup>), move approximately 2.0 to 2.5 Å away from the capping domain. However, not all the structural elements in the barrel domain move. The strands in close proximity ( $\beta_{b6}$  and  $\beta_{b7}$ ) to the capping domain remain fairly static, suggesting that the barrel itself forms an open and closed state. The difference in structural states is not induced when divalent metals are diffused into the apo crystals even though a metal ion is observed to be coordinated by the carboxylates (unpublished structural results). In the case of enolase, the opposite effect is observed where substrate binding causes the two homologous domains to move toward one another (39).

At least three explanations may explain the open and closed forms of the enzyme. The change may be an artifact of different crystal packing in the two crystal forms.

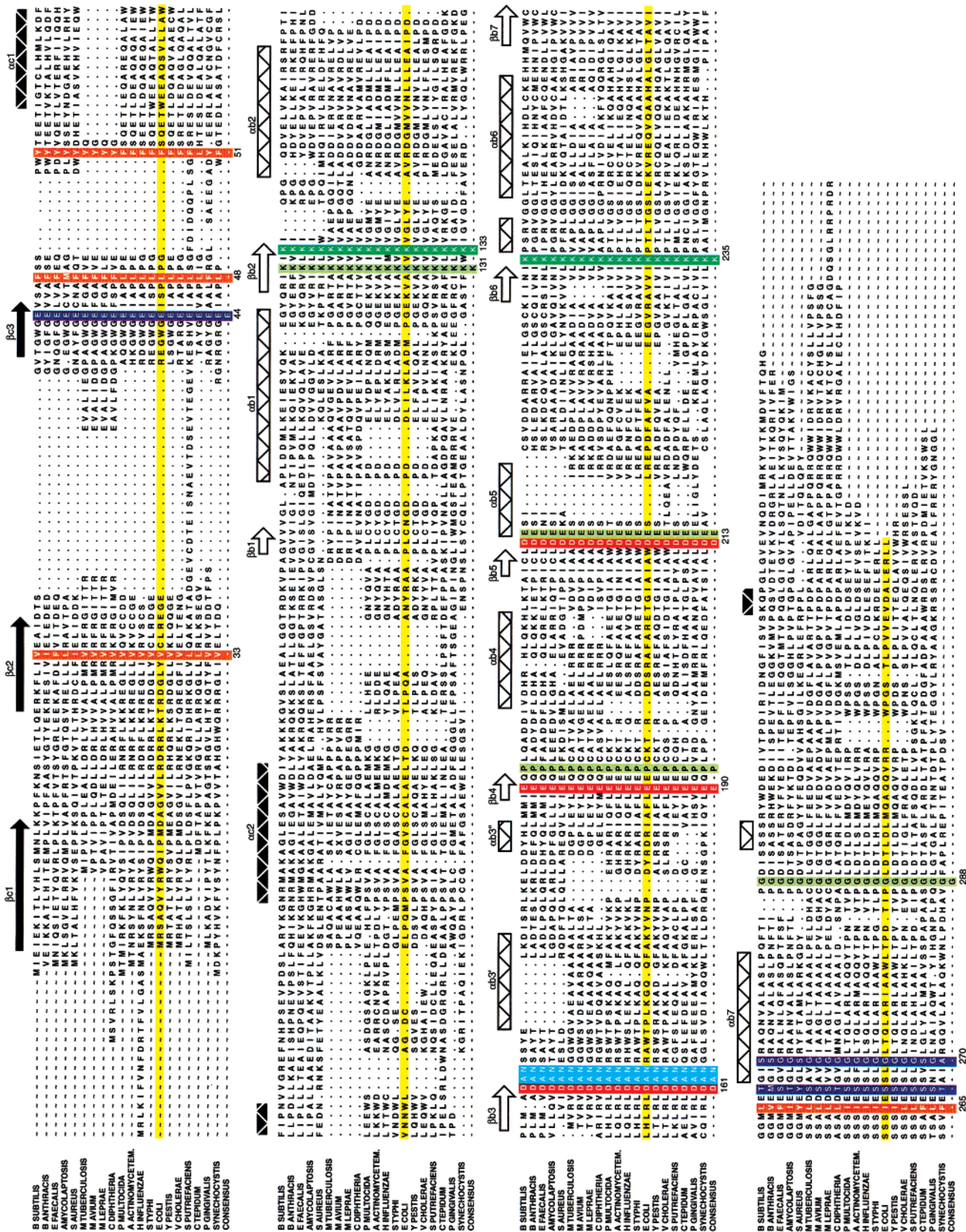


FIGURE 6: Sequence alignment of orthologous *o*-succinylbenzoate synthases. (See footnote 3). Important sections of the sequence alignment are colored as follows; red, metal binding residues; green, putative general acid/base Lys residues; orange, hydrophobic residues that align the active site cavity, aqua, conserved Ala and Asn of the DAN motif; blue, residues involved in a highly conserved domain-domain interaction; line-green, other conserved amino acid residues. The sequence for *E. coli* OSBS is highlighted in yellow. Selective residues are numbered according to the *E. coli* OSBS sequence. Secondary structural elements from *E. coli* OSBS are also included. Secondary structural elements in the capping and barrel domain are depicted in black and white, respectively. The  $\beta$ -strands are shown as arrows, whereas the  $\alpha$ -helices are illustrated as crosshatched blocks. Sequence alignment was performed with the PILEUP algorithm of the GCG software package (44).



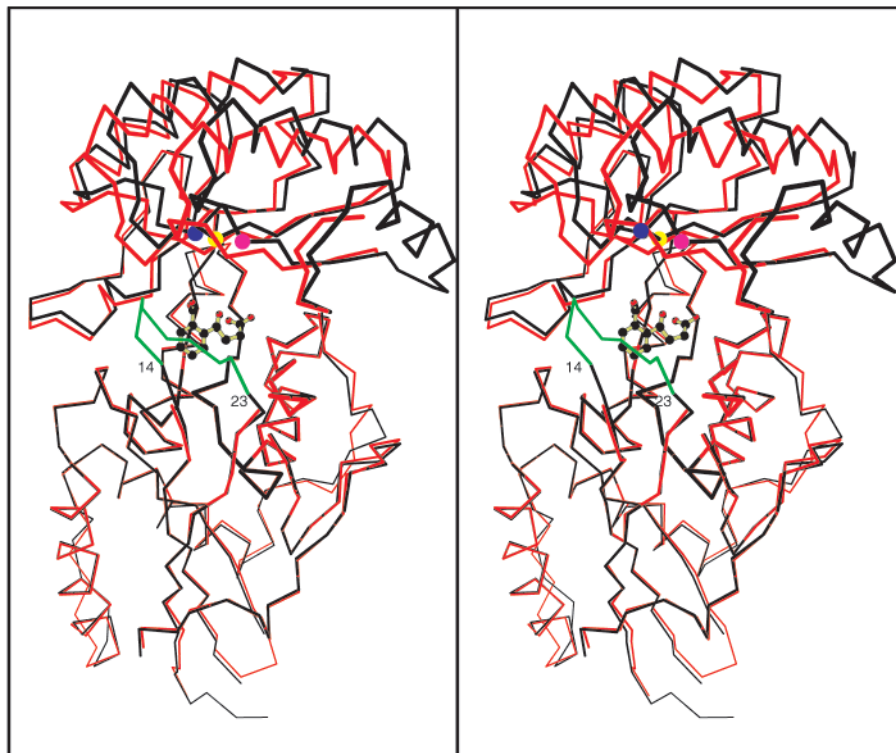


FIGURE 7: Stereoview of the  $C_{\alpha}$  alignment between the apo structure of OSBS (red line) and the  $Mg^{2+}$ •OSB complex (black line) performed with the program ALIGN (45) where the largest deviation was 4.4 Å. The loop between  $\beta_{c1}$  and  $\beta_{c2}$ , which is ordered in the  $Mg^{2+}$ •OSB complex, is colored green. For reference, the  $C_{\alpha}$ 's of Lys<sup>133</sup>, Asp<sup>161</sup>, and Glu<sup>190</sup> have been colored magenta, yellow, and blue, respectively.

Alternatively, the large loop between  $\beta_{c1}$  and  $\beta_{c2}$ , which extends into the active site and is colored green in Figure 7, could induce the barrel domain to open. Last, the binding of  $Mg^{2+}$ •OSB could be responsible for open and closed forms due to OSB's interaction with the active site residues located at the C-terminal ends of the  $\beta$ -strands (the segments that show the greatest relative movement). It is conceivable that a different relationship might be observed between the domains when substrate is bound and that these transitions may compensate for the pronounced disorder/order transitions observed between the apo and the product complexes. A deeper understanding of the conformational changes will require further structural studies to determine whether substrate binding or crystal packing forces are responsible for these domain movements.

**Active Site.** The structure of the  $Mg^{2+}$ •OSB complex reveals that the active site is located at the C-terminal ends of the  $\beta$ -strands of the barrel domain. The active sites of all other enzymes whose active sites are located in  $(\beta/\alpha)_8$  barrel domains are analogously positioned, including the other structurally characterized members of the enolase superfamily. The sequence alignment for the orthologous OSBS's (Figure 6) reveals that few residues are absolutely conserved; these are the ligands for the divalent metal ion (Asp<sup>161</sup>, Glu<sup>190</sup>, and Asp<sup>213</sup>), the putative acid/base catalysts (Lys<sup>133</sup> and Lys<sup>235</sup>), and those residues that are important for the structural reasons that were described earlier. The  $Mg^{2+}$  binding site is structurally analogous to those found in other members of the enolase superfamily. Lys<sup>133</sup> and Lys<sup>235</sup>, totally conserved in all the orthologues, reside on opposite sides of the active site with the OSB molecule positioned between them. The presence of these Lys residues confirms the membership of OSBS in the MLE subfamily.

**Divalent Cation Binding Site.** A stereoview of the active site is shown in Figure 8; various interresidue distances are shown in Figure 9. The  $Mg^{2+}$  is octahedrally coordinated, with Asp<sup>161</sup>, Glu<sup>190</sup>, and Asp<sup>213</sup>, two water molecules, and one benzoate carboxylate oxygen from the OSB molecule serving as ligands. Three acidic residues commonly form the high affinity metal binding site found in members of the enolase superfamily. Of these, only Asp<sup>161</sup> and Glu<sup>190</sup> are invariant in the enolase superfamily. In the previously determined structures of members of the superfamily, the metal ligands are located separately on the third, fourth, and fifth  $\beta$ -strands of the barrel or on the loops connecting these  $\beta$ -strands to the following  $\alpha$ -helices (1, 5, 7–9). OSBS is no exception as the three carboxylmetal ligands, Asp<sup>161</sup>, Glu<sup>190</sup>, and Asp<sup>213</sup>, are located separately on  $\beta_{b3}$ ,  $\beta_{b4}$ , and  $\beta_{b5}$ , respectively (Figure 5).

Asp<sup>161</sup> is part of an absolutely conserved DAN motif found in all OSBS orthologues ("AN" aqua in Figure 6). The small hydrophobic side chain of Ala<sup>162</sup> is required to maintain the correct position of Asp<sup>161</sup>.  $C_{\beta}$  of Ala<sup>162</sup> points into a small hydrophobic pocket formed by Trp<sup>166</sup>, Leu<sup>189</sup>, and Pro<sup>192</sup>; Leu<sup>189</sup> and Pro<sup>192</sup> line the sides of the pocket, whereas Trp<sup>166</sup> lies perpendicular to Ala<sup>162</sup> preventing substitution by a larger residue. The conformation of Asn<sup>163</sup> differs in the apo and  $Mg^{2+}$ •OSB bound structures. The  $\chi_1$  torsional angle rotates by 110° which points the side chain toward the active site when it is occupied. The positions of the second and third metal ligands, Glu<sup>190</sup> and Asp<sup>213</sup>, respectively, are stabilized by a hydrogen bond between the side chain of Glu<sup>191</sup> and the amide nitrogen of Asp<sup>213</sup>. The residue that follows the second metal ligand in the orthologous OSBS's is either a Glu or Gln (Glu<sup>191</sup> in this structure), suggesting that this interaction is conserved in all orthologous OSBSs. The third



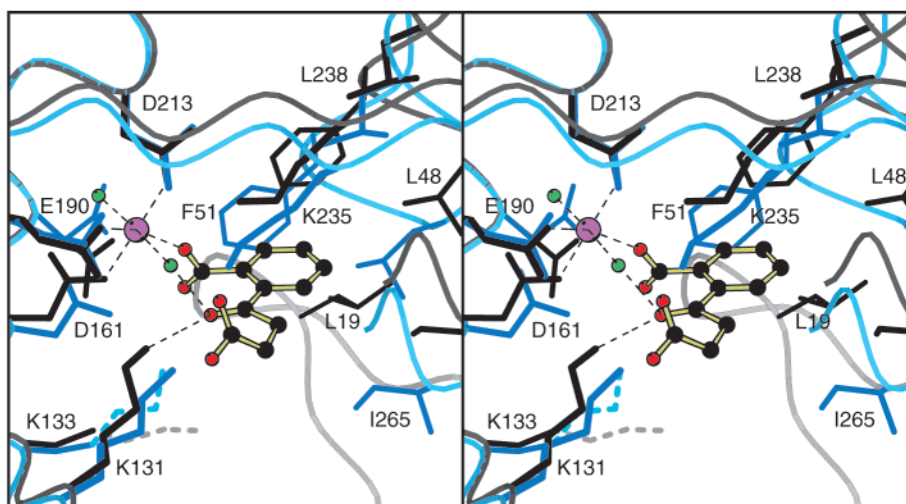


FIGURE 8: Stereoview of the active site alignment between the apo structure of OSBS (blue line) and the  $\text{Mg}^{2+}$ •OSB complex (black line) performed with the program ALIGN (45). The alignment was performed by superimposing the  $\text{C}_\alpha$ 's of the three metal ligands Asp<sup>161</sup>, Glu<sup>190</sup>, and Asp<sup>213</sup>. The  $\text{Mg}^{2+}$  ion (colored purple) is hexacoordinated to the three residue mentioned above, two water molecules (colored green), and one of the carboxyl oxygens of OSB. The loop between  $\beta_{c1}$  and  $\beta_{c2}$ , which is ordered in the OSB complex, is colored gray and contains Leu<sup>19</sup> which forms part of the OSB binding pocket. The  $\text{C}_\delta$ ,  $\text{C}_\epsilon$ , and  $\text{N}_\zeta$  atoms of Lys<sup>133</sup> are disordered in both the apo and the OSB structure and are represented with a dashed line.

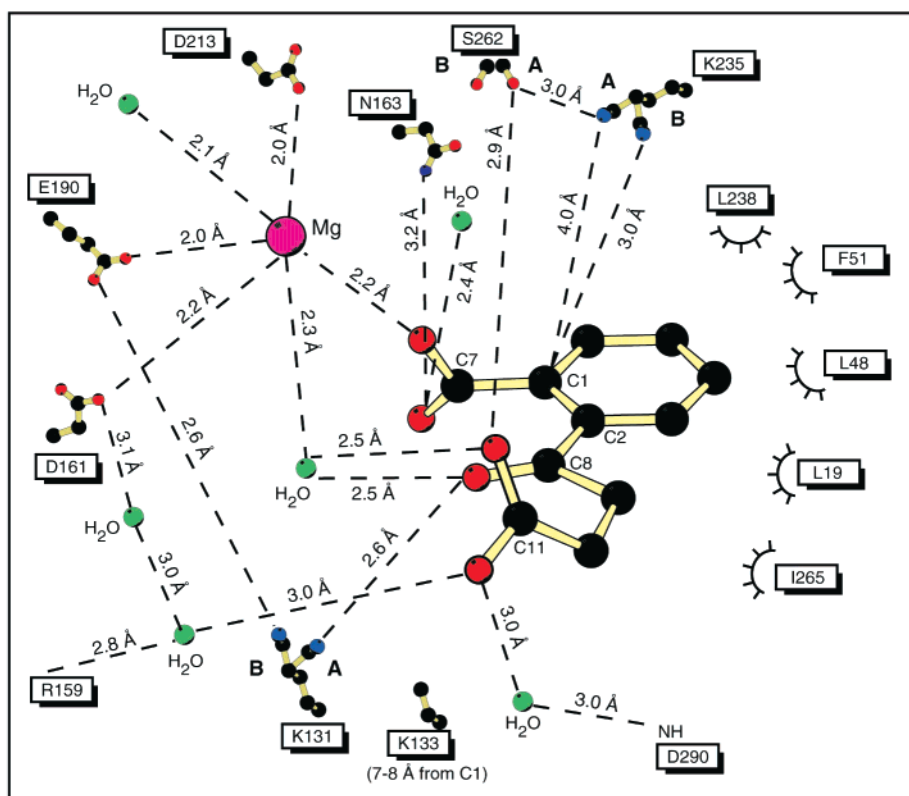


FIGURE 9: Ligand diagram of the OSBS active site. Lys<sup>131</sup>, Lys<sup>235</sup>, and Ser<sup>262</sup> have multiple conformations that are indicated by the letters A and B.

metal ligand, Asp<sup>213</sup>, is variable among the members of enolase superfamily but is totally conserved as an Asp in the members of the MLE subfamily.

**Residues Involved in Catalysis.** The chemical step common to all members of the superfamily is abstraction of the  $\alpha$ -proton of a carboxylate anion substrate by conserved general acid/base catalytic groups located on opposite sides of the active site (I). In OSBS, the conserved Lys<sup>133</sup> and Lys<sup>235</sup> are the most likely candidates and are situated on opposite sides of the active site at the C-terminal ends of

$\beta_{b2}$  and  $\beta_{b6}$ , respectively. Figures 8 and 9 detail the positions of these putative acid/base residues relative to the OSB molecule.

The C<sub>δ</sub>, C<sub>ε</sub>, and N<sub>ξ</sub> atoms of Lys<sup>133</sup> are disordered in both the apo and Mg<sup>2+</sup>•OSB bound structures (probably due to the lack of a hydrogen bonding partner). When Lys<sup>133</sup> is modeled in an extended conformation, the N<sub>ξ</sub> atom is a significant distance away (~8 Å) from the C6 of the bound OSB; however, the torsional angles of the side chain of Lys<sup>133</sup> can be adjusted such that N<sub>ξ</sub> can approach within 4 Å of C6

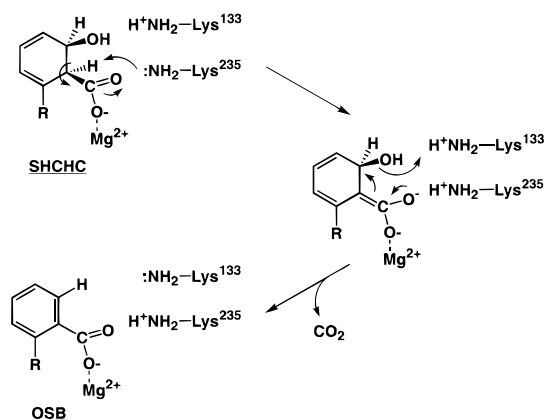


FIGURE 10: Proposed mechanism for the OSBS-catalyzed reaction.

of OSB. In contrast, Lys<sup>235</sup> is substantially closer to C1 and C6 of OSB. In this complex, Lys<sup>235</sup> exhibits two conformations (Figure 9). In one conformation, the side chain is hydrogen bonded to Ser<sup>262</sup> (which also exhibits multiple conformations) and points away from the OSB molecule. The alternative conformation positions atom N<sub>ε</sub> of Lys<sup>235</sup> within 3.0 Å of C1 of OSB. Even though the SHCHC substrate may bind in a slightly different conformation, the structural evidence presented here suggests that Lys<sup>235</sup> is the base that abstracts the α-proton from C1, and Lys<sup>133</sup> is the acid that donates a proton to the 6-OH leaving group of SHCHC to yield the OSB product (Figure 10).<sup>4</sup>

**Binding Pocket for OSB.** The binding pocket for OSB can be divided into three regions: (i) the carboxylate/divalent metal interaction; (ii) the succinyl group/OSBS interaction; and (iii) the hydrophobic interactions of OSBS with the benzene ring of OSB. Figure 3B shows the  $\sigma_a$  weighted omit electron density for OSB. The OSB molecule is well-defined except for the benzoate carboxyl group on C1. This carboxylate group likely lies perpendicular to the carbonyl group of the succinyl group and benzene ring, based solely on steric and charge conflicts. Apparently, the carboxyl adopts multiple conformations in the structure whereby a water replaces the carboxyl oxygen as the sixth metal ligand.

Figure 9 describes the interactions of the succinyl portion of OSB with the active site of OSBS. The carbonyl at C8 is an indirect ligand to the metal via an intervening water molecule. This same water molecule hydrogen bonds to one of the carboxylate oxygens of the succinyl group. The interaction of the succinyl portion of OSB with the metal ion via a water molecule may explain the unusual monodentate coordination of the metal ion by the benzoate carboxylate group at C1 since other enolase superfamily members usually exhibit a bidentate coordination involving at least one of the carboxylate oxygens of the substrate.

Lys<sup>131</sup>, which is part of the conserved KXK motif (Figure 6), has two conformations. In one conformation, the side

chain hydrogen bonds the carbonyl oxygen at C8. In the alternate conformation, Lys<sup>131</sup> hydrogen bonds to Glu<sup>190</sup>, one of the metal ligands. In the structures of mandelate racemase and GlucD the lower-numbered Lys of the Lys–X–Lys motif at the end of the second β-strand is hydrogen-bonded to one of the carboxylate oxygens of the substrate. The carboxyl group at C11 points toward the interior of the protein but only forms one direct interaction with the protein at Ser<sup>262</sup> (the same serine that interacts with Lys<sup>235</sup>). The remaining interactions between OSB and the active site are mediated via water molecules. Arg<sup>159</sup>, which is highly conserved, and the amide nitrogen of Asp<sup>290</sup> hydrogen bond to two separate water molecules that interact with one of the oxygens of the succinyl carboxyl group (Figure 9).

The hydrophobic section of the binding pocket is composed from several regions of the protein. Figure 8 shows the superposition of the active sites in the presence and absence of Mg<sup>2+</sup>•OSB shown in black and blue, respectively. The superposition was achieved by aligning the amino acid residues that bind the metal ion. The main difference induced by the presence of the OSB product occurs in the hydrophobic pocket that opens to allow OSB to bind. The pocket consists of Leu<sup>19</sup> from the large loop (β<sub>c1</sub> – β<sub>c2</sub>) of the capping domain that is ordered in the Mg<sup>2+</sup>•OSB complex, Leu<sup>48</sup> and Phe<sup>51</sup> from the following loop (β<sub>c3</sub> and α<sub>c1</sub>) in the capping domain, and Leu<sup>238</sup> and Ile<sup>265</sup> from separate regions of the barrel domain. Hydrophobic residues are conserved at these positions in the orthologous OSBSs (orange in Figure 6).

This binding pocket for the aromatic ring of OSB is comparable in structure to the binding site for the aromatic ring of mandelate in MR. In MR, a large loop that extends into the active site contributes a valine residue (Val<sup>22</sup>) and the loop after the third sheet of the capping domain contains Phe<sup>52</sup> and Tyr<sup>54</sup> in comparable positions to Leu<sup>48</sup> and Phe<sup>51</sup> of OSBS (13). Also, Leu<sup>298</sup> which follows one of the catalytic bases (His<sup>297</sup>) is analogous to Leu<sup>238</sup> which follows Lys<sup>235</sup> of OSBS and is on the same strand of the barrel domain (β<sub>bs</sub>).

**A Possible Structural Basis for the Evolution of Enzymatic Function.** The shape and volume of the active site appear well-suited for evolution of different reactions and substrate specificity. The binding interactions between C11, the remote carboxylate group, and the active site are mediated by a single interaction with an amino acid residue, Ser<sup>262</sup>, with the remaining interactions involving water molecules; it is expected that these water molecules can be displaced or repositioned to allow the substrate specificity to be altered. The interactions between the aromatic ring of OSB and the α/β capping domain are entirely hydrophobic: a substrate of differing dimensions could be accommodated by changes in the volumes and shapes of these side chains. A further consequence of “readjustments” in the shape and volume of the active site cavity is that the relative orientations of the α-proton, α-carbon, and/or β-leaving group of a bound substrate also can be altered such that homologues of Lys<sup>133</sup> and Lys<sup>235</sup> in other enzymes can be utilized to catalyze different reactions.

Thus, the initially surprising catalytic promiscuity documented for the NAAAR/OSBS from *Amycolaptosis* (2) can be rationalized by the structural plasticity of the substrate binding pocket. This plasticity also may explain the evolution

<sup>4</sup> An alternate mechanism for the OSBS-catalyzed reaction is general acid-catalyzed (by Lys<sup>235</sup>) departure of the 6-OH group to yield a stabilized carbocation intermediate. Abstraction of the α-proton by Lys<sup>133</sup> would yield the OSB product. In this mechanism, the benzoate carboxylate group would not be required to be in the plane of the cyclohexadienyl ring of SHCHC since an enolic intermediate is not involved and would not need to be stabilized. However, we regard both membership of OSBS in the enolase superfamily and the direct coordination of OSB to the Mg<sup>2+</sup> as persuasive evidence for the enolic intermediate mechanism.

of the dipeptide epimerases, e.g., L-Ala-D/L-Glu epimerases, that have been identified as members of the MLE subfamily.<sup>5</sup> Detailed mechanistic and structural studies of the NAAAR/OSBS and dipeptide epimerases are in progress; these should provide additional insights into the structural strategy for evolution of catalytic activities in the superfamily.

**A Monomeric Member of the Superfamily.** Thus far, OSBS is the only monomeric member of the enolase superfamily. The symmetry of the crystal lattice in which the apo form of *E. coli* OSB crystallizes precludes any simple rotational symmetry between neighboring molecules. In the Mg<sup>2+</sup>•OSB complex, the crystal lattice has P2<sub>1</sub>2<sub>1</sub>2 symmetry and could allow the presence of a molecular 2-fold axis; however, in this case, the buried surface area of the interactions between two subunits across the crystallographic 2-fold is 633 Å<sup>2</sup> or 4.2% of the total surface area, thereby indicating a very weak intersubunit interaction (40).

As mentioned earlier, crystal structures have been determined for enolase, mandelate racemase, muconate lactonizing enzyme, D-glucarate dehydratase (GlucD), and D-galactonate dehydratase (GalD). Each of these contains multiple subunits in their quaternary structures, forming dimers (enolase), tetramers (GlucD), and octamers (MR, MLE, and GalD). The subunit–subunit interface is variable among these proteins and contains a mixture of hydrophobic–hydrophobic and hydrogen bonding interactions. Furthermore, these proteins utilize different regions of the surfaces of their subunits to form the oligomeric structures. For instance, yeast enolase forms a dimer with interactions occurring between the N-terminal segment of the capping domain and the C-terminal section of the barrel domain. GlucD forms a tetramer with subunit–subunit interactions occurring at both the N-terminal segment of the capping domain and the start of the barrel domain. MLE and MR form octamers, with subunit–subunit interaction occurring between the loops that join the three helices of the N-terminal segment of the capping domain with the C-terminal segment of the capping domain of an adjacent subunit. In addition, helices 3, 4, and 5 of the barrel domains (the part of the barrel opposite the capping domain) in MR and MLE point toward the center of the octamer forming a variety of hydrogen bonds and hydrophobic interactions. In contrast, the surface of OSBS consists, for the most part, of hydrophilic residues that are not conducive to oligomerization. The observation that this OSBS is monomeric in the crystal lattice does not preclude the possibility that it might dimerize weakly in solution;<sup>6</sup> however, our crystallographic studies indicate that any interpolypeptide interactions must be very weak and cannot be predicted on the basis of the current structures. Thus, OSBS either is a primitive member of the superfamily or has lost the need to form oligomeric structures.

## CONCLUSIONS

The distantly diverged members of the enolase superfamily share an extraordinary spatial relationship, even though they catalyze different enzymatic reactions. The crystal structure of OSBS in complex with Mg<sup>2+</sup>•OSB both verifies the predicted identities of the three ligands for a catalytically

essential Mg<sup>2+</sup> (Asp<sup>161</sup>, Glu<sup>190</sup> and Asp<sup>213</sup>) as well as two Lys residues (Lys<sup>133</sup> and Lys<sup>235</sup>) that likely function as general acid/base catalysts. The roles of the latter residues are being probed by mutagenesis, kinetic, and structural studies.

The apparent structural plasticity within the active site of OSBS suggests that this has been exploited in the evolution of new enzymes and implies that this could be utilized in the design of new catalytic activities. It is expected that further studies of orthologous OSBSs, including the OSBS/NAAAR from *Amycolaptosis* (2), as well as studies of the L-Ala-D/L-Glu epimerases will provide additional insights into the evolution of function in the enolase superfamily.

## ACKNOWLEDGMENT

We thank Dr. Andrew Gulick for discussions on the enolase superfamily. We also thank the following SBC staff members: Dr. Frank Rotella, Dr. Norma E. C. Duke, Dr. Ruslan Sanishvili, and Jack Lazarz for help with data collection.

## REFERENCES

- Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L., and Gerlt, J. A. (1996) *Biochemistry* 35, 16489–501.
- Palmer, D. R., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., and Gerlt, J. A. (1999) *Biochemistry* 38, 4252–8.
- Farber, G. K., and Petsko, G. A. (1990) *Trends Biochem. Sci.* 15, 228–34.
- Wedekind, J. E., Reed, G. H., and Rayment, I. (1995) *Biochemistry* 34, 4325–30.
- Neidhart, D. J., Kenyon, G. L., Gerlt, J. A., and Petsko, G. A. (1990) *Nature* 347, 692–4.
- Larsen, T. M., Wedekind, J. E., Rayment, I., and Reed, G. H. (1996) *Biochemistry* 35, 4349–58.
- Helin, S., Kahn, P. C., Guha, B. L., Mallows, D. G., and Goldman, A. (1995) *J. Mol. Biol.* 254, 918–41.
- Lebioda, L., and Stec, B. (1988) *Nature* 333, 683–6.
- Gulick, A. M., Palmer, D. R., Babbitt, P. C., Gerlt, J. A., and Rayment, I. (1998) *Biochemistry* 37, 14358–68.
- Wieczorek, S. J., Kalivoda, K. A., Clifton, J. G., Ringe, D., Petsko, G. A., and Gerlt, J. A. (1999) *J. Am. Chem. Soc.* 121, 4540–51.
- Gulick, A. M., Hubbard, B. K., Gerlt, J. A., and Rayment, I. (2000) *Biochemistry* 39, 4590–4602.
- Hasson, M. S., Schlichting, I., Moulai, J., Taylor, K., Barrett, W., Kenyon, G. L., Babbitt, P. C., Gerlt, J. A., Petsko, G. A., and Ringe, D. (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95, 10396–401.
- Neidhart, D. J., Howell, P. L., Petsko, G. A., Powers, V. M., Li, R. S., Kenyon, G. L., and Gerlt, J. A. (1991) *Biochemistry* 30, 9264–73.
- Landro, J. A., Gerlt, J. A., Kozarich, J. W., Koo, C. W., Shah, V. J., Kenyon, G. L., Neidhart, D. J., Fujita, S., and Petsko, G. A. (1994) *Biochemistry* 33, 635–43.
- Wedekind, J. E., Poyner, R. R., Reed, G. H., and Rayment, I. (1994) *Biochemistry* 33, 9333–42.
- Sharma, V., Meganathan, R., and Hudspeth, M. E. (1993) *J. Bacteriol.* 175, 4917–21.
- Meganathan, R. (1996) in *Escherichia coli and Salmonella, Cellular and Molecular Biology* (Neidhardt, F. C., Ed.) pp 642–655, ASM Press, Washington, DC.
- Reed, G. H., Poyner, R. R., Larsen, T. M., Wedekind, J. E., and Rayment, I. (1996) *Curr. Opin. Struct. Biol.* 6, 736–43.
- Bodem, G. B., and Leete, E. (1976) *J. Am. Chem. Soc.* 98, 6321–6325.
- Kabsch, W. (1988) *J. Appl. Crystallogr.* 21, 67–71.

<sup>5</sup> B. K. Hubbard, D. Z. Schmidt, and J. A. Gerlt, unpublished observations.

<sup>6</sup> R. Meganathan, unpublished observations.



21. Wesenberg, G., and Rayment, I. (1997), manuscript in preparation.
22. Otwinowski, Z. (1986), Ph.D. Thesis, Yale University.
23. 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.
24. Terwilliger, T. C., Kim, S.-H., and Eisenberg, D. (1987) *Acta Cryst. A* **43**, 1–5.
25. Terwilliger, T. C., and Berendzen, J. (1996) *Acta Cryst. D* **52**, 749–757.
26. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–61.
27. CCP4. (1994) *Acta Crystallogr. D* **50**, 760–763.
28. Cowtan, K., and Main, P. (1998) *Acta Crystallogr. D* **54**, 487–493.
29. Kleywegt, G. J., and Jones, T. A. (1994) in *From First Map to Final Model* (Bailey, S., Waller, R. H. D., Ed.) pp 59–66, SERC, Daresbury Laboratory, Warrington.
30. Tronrud, D. E. (1997) *Methods Enzymol.* **277**, 306–19.
31. Roussel, A., and Cambillau, C. (1991) in *Silicon Graphics Geometry Partners Directory*, Silicon Graphics.
32. Navaza, J. (1993) *Acta Crystallogr. D* **49**, 588–591.
33. 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. D Biol. Crystallogr.* **54**, 905–21.
34. Read, R. J. (1986) *Acta Crystallogr. A* **42**, 140–149.
35. Gulick, A. M., Hubbard, B. K., Gerlt, J. A., and Rayment, I. (2000), manuscript in preparation.
36. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291.
37. Herzberg, O., and Moul, J. (1991) *Proteins: Struct. Funct. Genet.* **11**, 223–229.
38. Farber, G. K., and Petsko, G. A. (1990) *Trends Biochem. Sci.* **15**, 228–234.
39. Wedekind, J. E., Poyner, R. R., Reed, G. H., and Rayment, I. (1994) *Biochemistry* **33**, 9333–9342.
40. Miller, S., Lesk, A. M., Janin, J., and Chothia, C. (1987) *Nature* **328**, 834–836.
41. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950.
42. Esnouf, R. M. (1999) *Acta Crystallogr. D* **55**, 938–940.
43. Merrit, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. D* **50**, 869–873.
44. Devereux, J., Haeblerli, P., and Smithies, O. (1984) *Nucleic Acid Res.* **12**, 387–395.
45. Cohen, G. H. (1997) *J. Appl. Crystallogr.* **30**, 1160–1161.

BI0008550