Evolution of Enzymatic Activities in the Enolase Superfamily: Crystal Structures of the L-Ala-D/L-Glu Epimerases from *Escherichia coli* and *Bacillus subtilis*[†,‡]

Andrew M. Gulick,[§] Dawn M. Z. Schmidt,[||] John A. Gerlt,*[,||] and Ivan Rayment*[,§]

Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, and Departments of Biochemistry and Chemistry, University of Illinois, Urbana, Illinois 61801

Received August 9, 2001; Revised Manuscript Received October 22, 2001

ABSTRACT: The members of the enolase superfamily catalyze different overall reactions, yet share a partial reaction that involves Mg^{2+}-assisted enolization of the substrate carboxylate anion. The fate of the resulting enolate intermediate is determined by the active site of each enzyme. Several members of this superfamily have been structurally characterized to permit an understanding of the evolutionary strategy for using a common structural motif to catalyze different overall reactions. In the preceding paper, two new members of the superfamily were identified that catalyze the epimerization of the glutamate residue in L-Ala-D/L-Glu. These enzymes belong to the muconate lactonizing enzyme subgroup of the enolase superfamily, and their sequences are only 31% identical. The structure of YcjG, the epimerase from *Escherichia coli*, was determined by MAD phasing using both the SeMet-labeled protein and a heavy atom derivative. The structure of YkfB, the epimerase from *Bacillus subtilis*, was determined by molecular replacement using the muconate lactonizing enzyme as a search model. In this paper, we report the three-dimensional structures of these enzymes and compare them to the structure of *o*-succinylbenzoate synthase, another member of the muconate lactonizing enzyme subgroup.

The members of the "mechanistically diverse" enolase superfamily catalyze several different reactions that share divalent metal ion-assisted generation and stabilization of an enolate anion intermediate (*1*, *2*). This superfamily serves as a paradigm for understanding how an active site positioned at the C-terminal end of a $(\beta/\alpha)_7\beta$-barrel can be modified via the course of divergent evolution to catalyze "new" reactions. This process likely involves duplication of the gene encoding the "old" progenitor, with subsequent selective pressure directing changes in sequence and active site structure. In some cases, the progenitor may catalyze low levels of the desired new reaction, thereby providing a "toehold" for the evolutionary process (*3*).

The members of the muconate lactonizing enzyme (MLE) subgroup of the enolase superfamily share three conserved carboxylate ligands for the essential divalent metal ion (usually Mg^{2+}): an Asp at the end of the third β-strand of the barrel domain, a Glu at the end of the fourth β-strand, and an Asp at the end of the fifth β-strand (*4*). The members of the MLE subgroup also share functional groups that can serve as general acid/base catalysts: a Lys-X-Lys motif at the end of the second β-strand and a Lys at the end of the sixth β-strand. Despite these conserved residues, the members of the MLE subgroup catalyze different reactions. MLE catalyzes a reversible cycloisomerization reaction that utilizes the second Lys of the Lys-X-Lys motif at the end of the second β-strand as the catalyst that transfers protons to and/or from the α-carbon (*5*). *o*-Succinylbenzoate synthase (OSBS) catalyzes a highly exergonic dehydration reaction that likely utilizes the Lys at the end of the sixth β-strand as the general basic catalyst that mediates proton abstraction from the α-carbon and the second Lys of the Lys-X-Lys motif at the end of the second β-strand as the general acid catalyst (*6*). The promiscuous *N*-acylamino acid racemase/*o*-succinylbenzoate synthase (NAAAR/OSBS) catalyzes the OSBS reaction as well as a 1,1-proton transfer reaction that likely utilizes the Lys residues positioned at the ends of the second and sixth β-strands as the general acid/base catalysts in a two-base mechanism (*7*).

The ability of the active sites of members of the MLE subgroup to catalyze three different reactions provides the rationale for detailed mechanistic and structural studies of each aimed at understanding how changes in the active site

[†] This research was supported by Grants GM-40570 (to J.A.G.), GM-52594 (to J.A.G. and I.R.), and AR-35186 (to I.R.) from the National Institutes of Health. Use of the Argonne National Laboratory Structural Biology Center beamline at the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Energy Research, under Contract W-31-109-ENG-38.

[‡] The X-ray coordinates and structure factors for YcjG and YkfB have been deposited in the Protein Data Bank (1JPD and 1JPM, respectively).

* To whom correspondence should be addressed. I.R.: Department of Biochemistry, University of Wisconsin, 433 Babcock Dr., Madison, WI 53706; phone, (608) 262-0437; fax, (608) 262-1319; e-mail, ivan_rayment@biochem.wisc.edu. J.A.G.: Department of Biochemistry, University of Illinois, 600 S. Mathews Ave., Urbana, IL 61801; phone, (217) 244-7414; fax, (217) 265-0385; e-mail, j-gerlt@uiuc.edu.

[§] University of Wisconsin.

[||] University of Illinois.

[1] Abbreviations: MPD, 2-methyl-2,4-pentanediol; MLE, muconate lactonizing enzyme; SeMet, selenomethionine; MAD, multiwavelength anomalous dispersion; FOM, figure of merit; APS, Advanced Photon Source at Argonne National Laboratory, Argonne, IL; NAAAR, *N*-acylamino acid racemase; OSBS, *o*-succinylbenzoate synthase; OSB, *o*-succinylbenzoate; HEPPS, *N*-(hydroxyethyl)piperazine-*N*′-2-hydroxypropanesulfonic acid; BisTris, 2,2-bis(hydroxymethyl)-2,2′,2″-nitrilotriethanol.

10.1021/bi011641p CCC: $20.00
Published on Web 11/28/2001

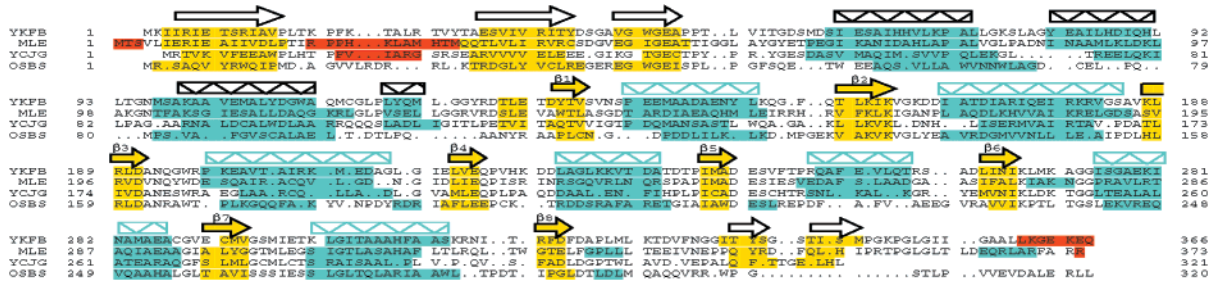

FIGURE 1: Structure-based amino acid alignment of members of the MLE subgroup. The sequences of YcjG, YkfB, MLE, and OSBS were aligned using a structure-based alignment (*18*, *19*) that minimizes the distance between the α-carbons of the structures. Disordered residues are shown in red, while the strands and helices are shown in yellow and blue, respectively. The location of the secondary structural elements is also shown in schematic form above the alignment. For these, the strands and helices of the capping domain are shown with open symbols with black lines, whereas the strands and helices of the barrel domain are shown in solid symbols in yellow and blue. The "extra" residues present in YkfB and MLE relative to YcjG and OSBS are distributed throughout the sequence.

architecture can direct enolate anion intermediates to different products with different chemistries. High-resolution structures are available for MLE (*5*, *8*) and OSBS from *Escherichia coli* (*6*). The structures of MLE do not contain a ligand in the active site; as such, the orientation of the active site functional groups that interact with the substrate and/or intermediate cannot yet be specified. The structure of the OSBS from *E. coli*, however, was determined with the OSB product bound in the active site; this structure suggests distinct roles for the Lys functional groups in the active site (*6*).

The discovery that the OSBS from *Amycolaptosis* also can catalyze low levels of the NAAAR reaction (*7*) requires that we understand the relationships between the mechanism and structure not only for this enzyme but also for a member of the MLE subgroup that catalyzes only a 1,1-proton transfer reaction. Detailed mechanistic and structural studies of the NAAAR/OSBS reaction are in progress. However, as described in the preceding article, we have discovered that the genomes of both *E. coli* and *Bacillus subtilis* encode L-Ala-D/L-Glu epimerases, YcjG and YkfB, respectively, that participate in murein peptide metabolism (*9*). In this article, we report the structures of both YcjG and YkfB (Figure 1) and compare these with the OSBS from *E. coli*.

MATERIALS AND METHODS

Cloning and Expression of YcjG and YkfB. The genes encoding YcjG and YkfB were cloned from genomic DNA as described in the preceding paper (*9*). The enzymes were purified to homogeneity as described and concentrated to ~10 mg/mL for crystallization studies. YcjG was dialyzed at 4 °C against 20 mM HEPPS (pH 8.0), 100 mM NaCl, and 1 mM NaN_3 and frozen in small aliquots in liquid nitrogen. Protein was thawed for crystallization experiments. While crystals grew from protein that was previously frozen, it was consistently observed that the best crystals were obtained from protein that had not been frozen.

Selenomethionine-labeled YcjG was prepared according to the following method. *E. coli* strain BL21(DE3) was transformed with the pET15B plasmid carrying the gene encoding YcjG (*9*). A 5 mL overnight culture in M9 minimal medium was used to inoculate a 2 L culture of minimal medium containing the following: 33 g of $K_2HPO_4 \cdot 3H_2O$, 3 g of KH_2PO_4, 2 g of $(NH_4)_2SO_4$, 18 g of glucose, 0.4 g of $MgSO_4 \cdot 7H_2O$, 20 mg of $CaCl_2 \cdot 2H_2O$, 100 μg of $FeSO_4 \cdot 7H_2O$, 0.1 g of thiamine, and 0.2 g of ampicillin. The culture was grown at 37 °C to an OD_{600} of 0.6. At this stage, 40 mL of a 0.22 μm filter-sterilized solution containing 200 mg of L-lysine, 200 mg of L-threonine, 200 mg of L-phenylalanine, 100 mg of L-leucine, 100 mg of L-isoleucine, and 100 mg of L-valine was added. The culture was returned to shaking at 37 °C for 30 min, at which time 0.1 g of SeMet was added (the SeMet was dissolved in cold water under N_2 and added to the culture through a sterile 0.22 μm filter). IPTG was added to a final concentration of 0.5 mM. The culture was returned to shaking at 37 °C and allowed to continue growing overnight. The labeled protein was purified under the conditions described for the unlabeled protein, with the addition of 2 mM β-mercaptoethanol to the sonication buffer and loading buffers for each column purification step. The SeMet-labeled protein was used directly for crystallization trials without freezing.

YkfB was dialyzed at 4 °C against 10 mM HEPPS (pH 8.0), 100 mM NaCl, and 1 mM NaN_3 (*9*). The protein was concentrated to 11.4 mg/mL and frozen as described for YcjG.

Analytical Gel Filtration Chromatography of YcjG and YkfB. The oligomeric states of both YcjG and YkfB were determined using a Bio-Sil SEC 250-5 300 mm × 7.8 mm gel filtration HPLC column (Bio-Rad) on a Beckman HPLC system. The mobile phase was 0.10 M sodium phosphate, 0.15 M NaCl, and 0.01 M sodium azide (pH 6.8) at a flow rate of 1.0 mL/min. The Bio-Rad gel filtration standard was used to generate the calibration curve.

Crystallization of YcjG. Crystals of YcjG were grown by hanging drop methods at room temperature. The precipitant was 0.4−0.6 M NaCl and 50 mM succinate (pH 5.0), and crystals grew in 3−5 days upon mixing equal volumes of protein and precipitant. The crystals grew as diamonds with maximal dimensions of 0.4 mm × 0.4 mm × 0.7 mm. Precession photography demonstrated that the crystals belonged to the $P4_12_12$ or $P4_32_12$ space group with the following unit cell dimensions: $a = b = 84.3$ Å and $c = 159.2$ Å. Prior to being frozen, crystals were transferred to a cryoprotectant consisting of 0.6 M NaCl, 5 mM $MgCl_2$, 50 mM succinate (pH 5.0), and 10% MPD for 3 min, followed by the same solution containing 20% MPD for an additional 3 min.

Determination of the Structure of YcjG. Crystals of YcjG were soaked for 20−24 h at room temperature in 0.6 M NaCl, 5 mM $MgCl_2$, 2.5 mM K_2PtCl_4, and 50 mM succinate (pH 5.0). After the soak, the crystals were transferred to

Table 1: Data Collection Statistics for YcjG and YkfB

	YcjG	SeMet YcjG			K_2PtCl_4 YcjG			YkfB
	native	peak	inflection	remote	peak	inflection	remote	native
wavelength (Å)	0.70008	0.97930	0.97917	1.01943	1.07186	1.07214	1.12059	0.95372
resolution (Å)	2.6	2.65	2.65	2.65	2.7	2.7	2.7	2.25
R_{merge} (%)[a]	5.3 (28.6)	6.6 (31.0)	5.5 (27.5)	4.0 (26.7)	7.0 (68.6)	6.2 (76.5)	5.4 (66.5)	7.2 (41.6)
R_{anom} (%)	−	7.1	4.8	2.1	5.8	4.9	3.5	−
completeness (%)[a]	95.3 (97.0)	99.8 (100.0)	99.8 (100.0)	99.7 (100.0)	99.8 (100.0)	99.8 (100.0)	99.3 (99.2)	99.9 (99.9)
no. of observations	101261	173646	176954	179995	228795	176397	173030	570691
no. of unique reflections	17379	17238	17220	17349	16453	16474	16486	97014
average intensity[a]	19035 (2003)	6185 (558)	6377 (591)	6935 (605)	4178 (87)	3843 (198)	4843 (240)	6290 (968)
average σ[a]	446 (219)	122 (78)	120 (74)	126 (71)	90 (95)	93 (71)	109 (80)	227 (216)
space group	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$I4$
unit cell ($a = b$, Å)	84.4	84.3	84.2	84.5	84.3	84.3	84.3	157.6
unit cell (c, Å)	159.2	158.7	158.7	159.1	159.6	159.6	159.6	168.3

[a] Values for the highest-resolution shell are shown in parentheses.

cryoprotectant solutions that were identical to the heavy metal soaking solution with the addition of 10 or 20% MPD. The crystals were soaked for 1 min in the 10% MPD solution followed by 2 min in 20% MPD.

The SeMet-labeled protein was crystallized and cryoprotected under the same conditions described for the unlabeled protein.

A native data set was collected at beamline ID19 of the Structural Biology Center-CAT of APS using two 135° scans with 1° oscillations. Exposure times were 7 s at a distance of 300 mm and 3 s at a distance of 350 mm.

Three-wavelength MAD experiments were performed with both the SeMet-labeled and Pt derivative crystals. Data were collected at beamline ID19 at the Structural Biology Center-CAT of APS. For crystals of the SeMet-labeled protein, 135° of data were collected at each wavelength at a distance of 235 mm with 4 s exposures. The same distance and time conditions were used for the Pt derivative; however, 180° of data were collected for the peak wavelength, while 137° of data were collected for the inflection and remote data sets. The data collection statistics are given in Table 1.

The data were processed with HKL2000 (*10*), and the output reflection files were input into SOLVE (*11, 12*) for heavy atom location and refinement. The Pt derivative data were locally scaled prior to submission to SOLVE, while the SeMet data were taken directly from HKL2000 into SOLVE. For the SeMet data, SOLVE found eight of nine Se positions and yielded a final FOM of 0.74 and a *z*-score of 28.15. For the Pt derivative, two major sites and one minor site were identified, yielding a FOM of 0.49 with a *z*-score of 11.1. In both cases, the resulting phases were improved by solvent flattening with DM (*13*), resulting in readily interpretable maps. The SeMet-derived map was generally of higher quality and was used for model building with TURBO-FRODO (*14*). The map derived from the Pt derivative was used in several locations where the SeMet map was ambiguous. The initial model contained 309 amino acids and 2274 atoms. Of the 309 residues, 294 were built with the proper side chain, while 15 residues were built as alanines because no side chain density was observed. Two breaks were present in the original model, between Pro 16 and Ser 25 and between Glu 37 and Lys 40.

The native data were also processed with HKL2000 (*10*). The intensities were converted to structure factors with TRUNCATE of the CCP4 package, which gave a Wilson *B* value of 56 Å^2. The high temperature factor is not surprising given the high solvent content of ~70% (*15*). The model was then subjected to atomic refinement using CNS (version 1.0) against the native data to 2.6 Å resolution (*16*). The initial refinement reduced the *R*-factor and R_{free} values from 40.6 and 39.0% to 27.1 and 30.2%, respectively. Atomic refinement continued against the native data using iterative cycles of manual model building and refinement with CNS. The final refinement statistics are listed in Table 2.

Table 2: Final Refinement Statistics for YcjG and YkfB

	YcjG	YkfB
no. of reflections	17409 (2726)[a]	96978 (15273)[a]
R-factor	20.9 (30.0)[a]	19.7 (23.2)[a]
R_{free}	24.5 (33.8)[a]	22.4 (27.6)[a]
rmsd for bond lengths (Å)	0.006	0.006
rmsd for bond angles (deg)	1.3	1.2
no. of protein atoms	2395	10745
no. of solvent atoms	144	520
Wilson *B* value	60.4	34.1
avg *B* factor, protein (all atoms/main chain)	49.7/48.7	32.0/30.8
avg *B* factor, solvent	50.2	34.0
avg *B* factor, heteroatoms	−	35.3
PDB entry	1JPD	1JPM

[a] Values in parentheses represent data for the highest-resolution shell (2.76−2.6 Å for YcjG and 2.39−2.25 Å for YkfB).

The surface areas of the YcjG monomer and dimers were calculated with GRASP (*17*) with a probe radius of 1.4 Å. Amino acid alignments were performed using ALIGN (*18*, *19*).

Crystallization and Determination of the Structure of YkfB. YkfB also crystallized at low pH; the crystals appeared as long rods. Crystals were grown by the hanging drop method using a precipitant of 8% PEG 400, 0.2 M $MgCl_2$, and 50 mM BisTris (pH 5.0) at room temperature. The crystals also belonged to a tetragonal space group, *I*4, with the following cell dimensions: $a = b = 158.2$ Å and $c = 168.6$ Å. The Matthews coefficient of 3.3 Å^3/Da suggested four subunits were present in the asymmetric unit. A crystal was soaked for 1 min each in 12% PEG 400, 0.25 M $MgCl_2$, and 50 mM BisTris, containing first 10% and then 25% glycerol. The crystal was then flash-frozen in a stream of nitrogen gas at −160 °C, and 135° of data were collected using 0.7° oscillations at the BM19 beamline at the Structural Biology Center-CAT at APS. The frames were processed and scaled with HKL2000 (*10*) and converted to structure factors with TRUNCATE (*13*). Final data collection statistics are given in Table 1.

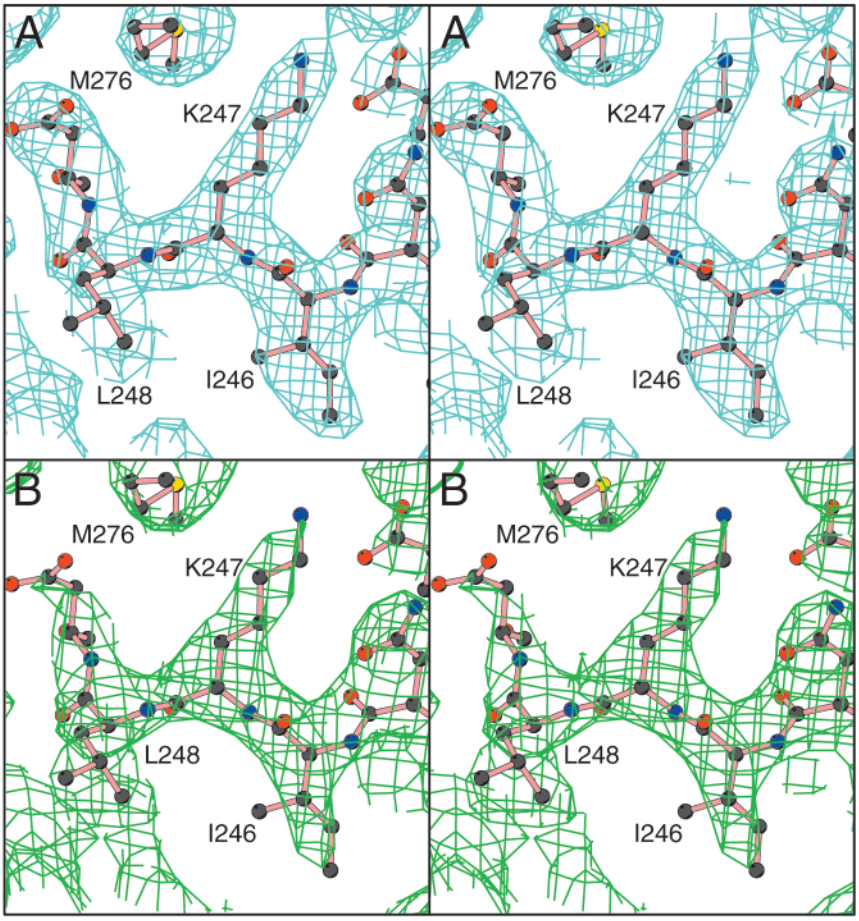

FIGURE 2: Electron density of YcjG MAD experimental maps. The heavy atom positions of the two MAD data sets were identified with SOLVE, and the resulting phases were improved with solvent flattening. The improved phases were used to generate electron density maps which were used for manual model building. Representative densities are shown from the same positions of both the SeMet (A) and K_2PtCl_4 (B) MAD data. The locations of the atoms from the final refined model are included.

Initial attempts to determine the structure with molecular replacement using YcjG as a search model proved to be unsuccessful. Attempts were made with AMORE (*13*), EPMR (*20*), and CNS (*16*), using either a monomer or dimer of YcjG. A correct solution was finally obtained using MOLREP (*13*) and a search model consisting of two adjacent subunits of the MLE octamer [chains A and C of PDB entry 1BKH (*8*)] with all side chains truncated to polyalanine. Two positions yielding a total of four subunits in the asymmetric unit were identified. Both dimers were aligned along the 4-fold rotation axis to generate two octamers in the crystallographic unit cell. A second MLE model was then used [PDB entry 1MUC (*5*)] from which only nonconserved side chains were removed. This model was rotated onto the four subunits of the molecular replacement solution and used as the starting model for model building and refinement.

An averaged map was generated with DM (*13*), and an improved model was generated by manual model building. The new model contained 240 complete residues, 99 residues built as alanines, and 27 disordered residues. This model was expanded back to the tetramer and subjected to atomic refinement with CNS. The *R*-factor dropped from 41.4 to 29.7% (R_{free} from 41.4 to 33.8%). After one more cycle of model building with an averaged map, the individual subunits were treated separately for the remaining refinement. Refinement continued with iterative cycles of model building and atomic refinement; final statistics are shown in Table 2.

RESULTS AND DISCUSSION

Structure of YcjG from E. coli. The structure of YcjG was determined from three-wavelength MAD experiments with both a Pt derivative and a SeMet-labeled protein. The data were processed, and heavy atom positions were determined with SOLVE. The resultant phases were improved with solvent flattening using DM, generating interpretable maps (Figure 2). The model was built primarily into the SeMet-derived map and subsequently refined against a higher-resolution native data set. The final refinement statistics are included in Table 2.

Native YcjG is 321 residues in length. The protein used for crystallization contained an additional N-terminal Gly-Ser-His tripeptide that remains after cleavage of the His tag with thrombin (*9*). The final model contains 318 residues, including this tripeptide. The only disordered region of the protein is a loop between the first two strands at the N-terminus of the protein consisting of residues Phe 17−Gly 22. The analogous loop was also disordered in MLE (*5*) and OSBS (*6*), the two previously characterized members of this subfamily. In the case of OSBS, this loop became ordered upon the binding of substrate. The side chains of two residues, Ser 23 and Arg 24, were disordered in the final model and were built into the model as alanine residues.

The structure of YcjG contains all of the basal secondary structural elements present in members of the enolase

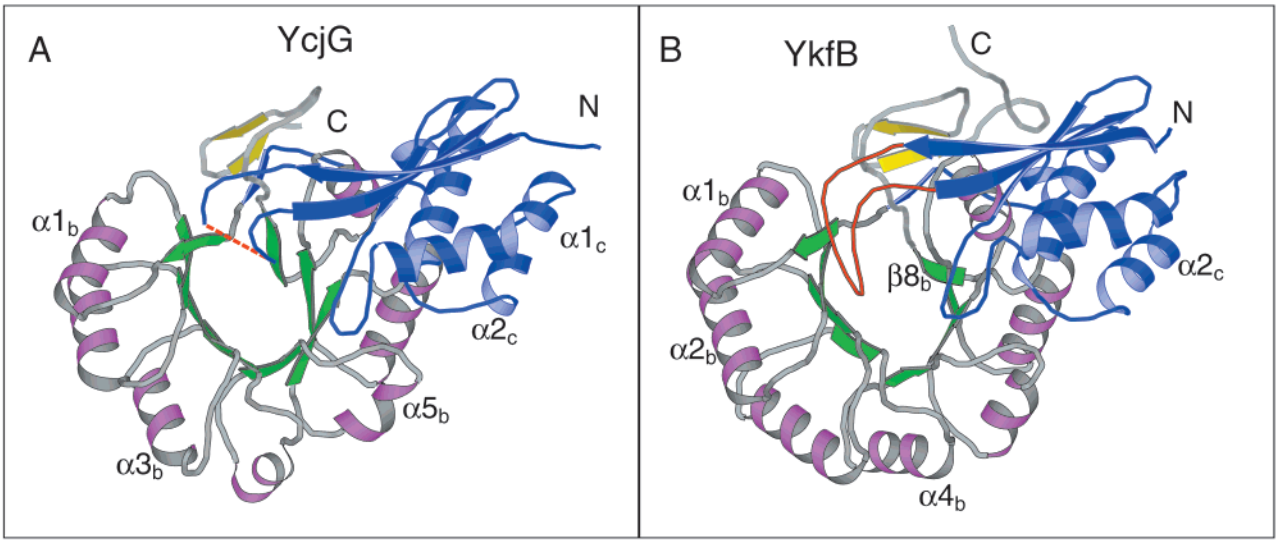

FIGURE 3: Ribbon diagram of YcjG and YkfB. Ribbon diagrams are shown for a single subunit of YcjG (A) and a subunit of YkfB (B). In both diagrams, the N-terminal capping domain is shown in blue. The barrel domain is shown with green strands and magenta helices. The loop between the first two strands of the capping domain is disordered in YcjG and is represented by a dashed line; in YkfB, the analogous loop is colored red.

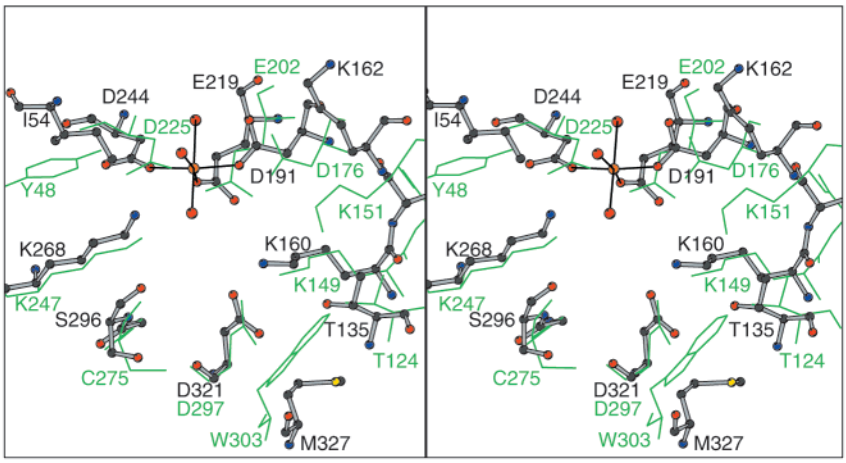

FIGURE 4: Active sites of both YcjG and YkfB. The active sites of YcjG and YkfB are shown. YkfB is shown in a ball-and-stick representation; YcjG is shown with green bonds. The octahedrally coordinated Mg^{2+} ion in YkfB is also shown. A residue located on each of the eight strands of the barrel is shown for both enzymes; these are thought to be important, based on comparison with other superfamily members and the proximity to the active site.

superfamily (Figure 3A). The N-terminal capping domain contains three β-strands followed by three α-helices. The first helix of the capping domain contains a proline residue (Pro 67) in the center which produces a significant kink. The capping domain concludes with a short strand before the barrel domain of the protein begins. The barrel contains a typical $(\beta/\alpha)_7\beta$-barrel with alternating strands and helices. The enzyme active site is expected to be located in the barrel domain, with residues from the strands and the loops that follow them contributing important catalytic residues. The KXK motif that includes Lys 149 and Lys 151 is located on strand β2, with the lysine side chains pointing toward the center of the barrel. Lys 247, which is expected to be the second catalytic base used in the epimerization reaction, is located on the opposite face of the active site on strand β6, as expected. The three metal ligands located on barrel strands β3, β4, and β5 also are located in an expected position that would allow them to coordinate a Mg^{2+} ion (Figure 4).

YcjG is a monomer in solution as determined by analytical gel filtration chromatography (data not shown). Despite this fact, the subunit present in the asymmetric unit makes extensive contacts with one of the neighbors that align around the crystallographic 2-fold axis. The molecular accessible surface area around the monomer is 14 020 Å^2. The amount of surface buried in forming the dimers is 1290 Å^2, or roughly 9% of the surface area. This value falls within the broad parameters for surface area of multimeric proteins defined by Jones and Thornton (*21*).

Structure of YkfB from B. subtilis. The structure of YkfB, a homologue from *B. subtilis* that also catalyzes the racemization of the glutamate residue in L-Ala-L/D-Glu, was determined. YkfB consists of 366 amino acids and is 45 residues longer than YcjG, and its sequence is only 31% identical to that of YcjG. Despite the expected structural homology between YcjG and YkfB, it was not possible to determine the structure of YkfB with molecular replacement with either a YcjG monomer or dimer as a search model. Instead, the structure of YkfB was determined by molecular replacement using two subunits of the octameric MLE structure as a search model. A correct solution was found

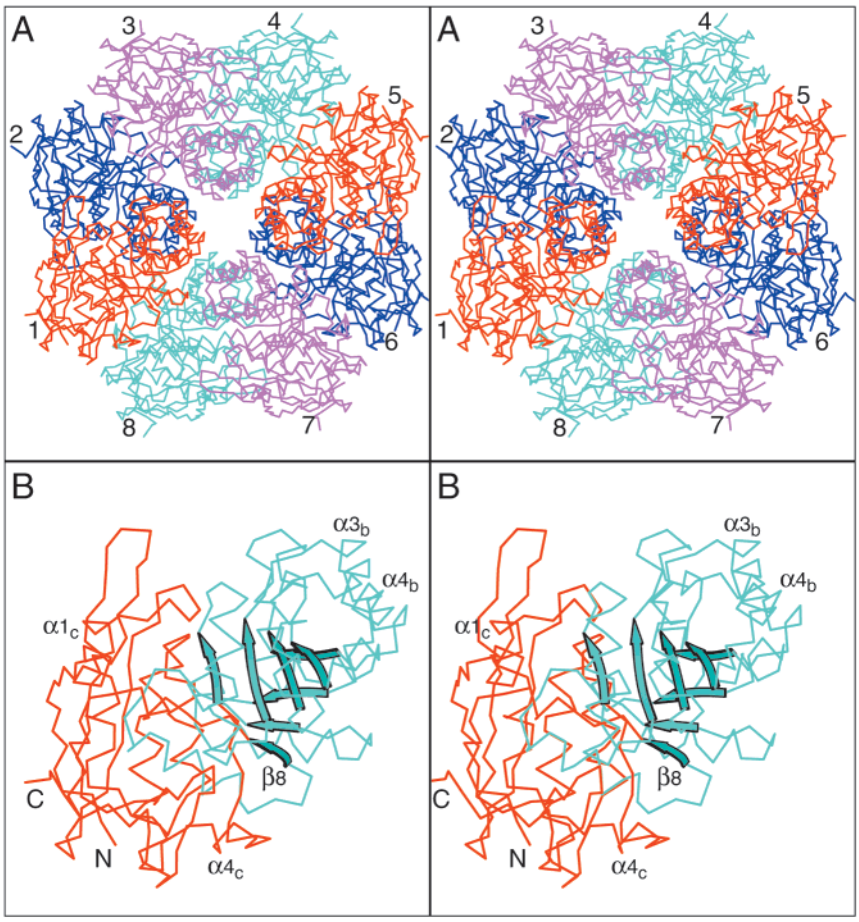

FIGURE 5: Octameric state of YkfB. (A) The octamer of YkfB is shown viewed down the 4-fold rotation axis. The protein consists of a tetramer of dimers where the dimers are formed by subunits 1 and 2. Subunit 1 makes considerable contacts with subunits 2 and 8, with which it is related by two 2-fold rotation axes. Subunit 1 also makes limited interactions with subunits 3 and 7. (B) Subunit 1 is shown enlarged in the precise orientation in panel A. This allows the identification of the regions of the protein that form the subunit interactions. The capping domain is shown in red, while the barrel domain is shown in cyan. The eight strands of the barrel domain are represented by ribbons. Interactions between subunits 1 and 2 come primarily from residues of the first and third helices of the capping domain and the loops that precede them but also include a hydrophobic interaction from a residue on the fifth helix of the barrel domain. Interactions with subunit 8 come from the start of the second helix of the capping domain, the loop, and the short two-turn helix that follows helix 3 of the capping domain, and between the fourth and fifth helices of the barrel domain. Last, minor interactions occur between subunits 1 and 3 (an ionic interaction between the third helix of the barrel and the loop preceding strand β5) and between subunits 1 and 7 (another ionic interaction between the end of the fifth helix of the barrel and the start of the fourth helix of the barrel). The first and fourth helices of the capping domain, as well as several elements from the barrel domain, are labeled; subscripts c and b denote the capping and barrel domains, respectively.

which demonstrated that YkfB is also an octamer (Figure 5), which is in agreement with the oligomeric state of YkfB determined by analytical gel filtration chromatography (data not shown). The asymmetric unit of the crystal contains two dimers that both associate around the crystallographic 4-fold rotation axis to form the two octamers present in the unit cell. The refinement statistics for YkfB are given in Table 2.

The overall secondary structure of YkfB is very similar to that found in both YcjG and MLE (Figures 1 and 3). All of the secondary structural elements present in YkfB are also present in MLE. In contrast, OSBS, the smallest member of this subgroup, is missing several of the conserved structural elements present in other members of the MLE subgroup. The capping domain of OSBS contains only two helices; it also is missing the short three-stranded β-sheet that is composed of one short strand that precedes strand β1 of the barrel domain and the two strands that follow strand β8 of the barrel (Figure 1). The C-terminus of YkfB extends beyond this small sheet to form an extended loop. The final eight residues in all four subunits are disordered. All four subunits contain several disordered residues for which the side chains are not observable. These residues have been modeled as alanine residues: Leu 23, Thr 25, Lys 153, Lys 165, Ser 184, and Leu 359 of subunit A; Lys 20, Lys 165, Lys 207, Lys 349, and Leu 359 of subunit B; Lys 20, Lys 165, and Lys 180 of subunit C; and Lys 20, Thr 25, Lys 165, Asp 167, and Lys 207 of subunit D.

While the crystallization conditions used for YcjG prohibited the binding of metal in the active site, because of the use of succinate as the buffer, those for YkfB allowed the binding of a Mg^{2+} ion to the three residues that coordinate the metal. These residues (Asp 191, Glu 219, and Asp 244) form close interactions with the Mg^{2+} (Figure 4). The structure shows that in all four subunits three water molecules complete the octahedral coordination. The three metal binding ligands of YcjG are spatially conserved, suggesting that under suitable conditions this enzyme will bind to Mg^{2+} with an analogous coordination geometry.

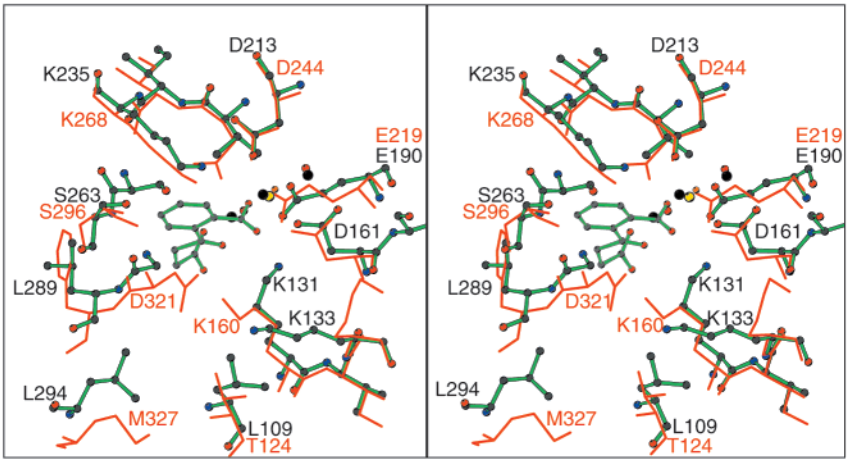

FIGURE 6: Active site overlay of YkfB and OSBS. The active site of YkfB is compared with that of OSBS with bound OSB [derived from PDB entry 1FHV (*6*)]. OSBS is shown as a ball-and-stick representation with green bonds; YkfB is shown with red sticks. OSB is present in the center of the figure. The Mg^{2+} ion and surrounding waters for YkfB are shown in yellow and black, respectively; those for OSBS are shown in orange and red, respectively.

Table 3: Domain Size of MLE Subgroup Members

	capping domain	barrel domain	avg helix length[a]
YkfB	Met 1−Glu 131, Phe 322−Gln 366 (176 amino acids)	Thr 132−Asp 321 (190 amino acids)	11.9
MLE	Met 1−Glu 137, Leu 328−Arg 373 (183 amino acids)	Val 138−Glu 327 (190 amino acids)	12.7
YcjG	Met 1−Ile 120, Gly 300−Leu 321 (142 amino acids)	Thr 121−Asp 299 (179 amino acids)	9.9
OSBS	Met 1−Arg 105, Thr 291−Leu 320 (135 amino acids)	Ala 106−Asp 290 (185 amino acids)	11.5

[a] Number of residues that make up the seven helices in the barrel domain. Note that OSBS contains one aberrant loop between strands β5 and β6 of the barrel domain which contains a short, single-turn helix; this helix was not included in the calculation.

Comparison of the Structures of the Members of the MLE Subgroup. Within the enolase superfamily, two other enzymes have been characterized from two different species. These are enolase, which has been crystallized from protein derived from both yeast (*22*−*24*) and lobster (*25*), and glucarate dehydratase (GlucD), which has been crystallized from protein from both *E. coli* and *Pseudomonas putida* (*26*, *27*). In these cases, the enzymes were highly conserved, with the sequences of the two enolases being 64% identical and the sequences of the GlucD enzymes being 72% identical. The two L-Ala-D/L-Glu epimerases studied here provide greater information about the plasticity of the sequence of these enzymes since they are only modestly conserved in sequence, with the sequences being 31% identical (Figure 1).

Despite the low level of sequence identity relating YcjG and YkfB, the active sites of the two enzymes are very well conserved in both sequence and structure (Figure 4). The α-carbon atoms of the residues that form the β-strands in the barrel and the important catalytic residues superimpose with an rmsd of 0.62 Å. In addition, the only residues at the presumed active site that differ in sequence between YcjG and YkfB are the substitutions of Tyr 48, Cys 275, and Trp 303 of YcjG with Ile 54, Ser 296, and Met 327 of YkfB, respectively (Figure 4). It is unclear which of these changes are responsible for the differences in substrate specificity that distinguish these enzymes (*9*).

As with other members of the enolase superfamily, each strand in the barrel contains one or more residues that are expected to be either involved in the catalytic mechanism or important for substrate binding. The eight strands from YcjG and YkfB donate the following residues to the active site: strand β1, Thr 124 for YcjG or Thr 135 for YkfB; strand β2, Lys 149 and Lys 151 for YcjG and Lys 160 and Lys 162 for YkfB (one of the catalytic bases); strand β3, Asp 176 for YcjG and Asp 191 for YkfB (a metal ion ligand); strand β4, Glu 202 for YcjG and Glu 221 for YkfB (a metal ion ligand); strand β5, Asp 225 for YcjG and Asp 244 for YkfB (a metal ion ligand); strand β6, Lys 247 for YcjG and Lys 268 for YkfB (the second catalytic base); strand β7, Cys 275 for YcjG and Ser 296 for YkfB; and strand β8, Asp 297 for YcjG and Asp 321 for YkfB. The elements from strands β1, β7, and β8 are likely to be involved in substrate binding. One expected feature of the reactions catalyzed by YcjG and YkfB is that a negatively charged group will interact with the amino group on the Ala residue of the substrate; the conservation of Asp 297 in YcjG and Asp 321 in YkfB suggests that these residues perform this function.

While all of the members of the MLE subgroup contain the same structural elements, the lengths of the sequences differ by as many as 54 residues. MLE and YkfB are the larger members of this subgroup, with 373 and 366 residues, respectively; YcjG and OSBS contain 321 and 320 residues, respectively. Approximately 20 residues present at the C-termini of MLE and YkfB account for some of this difference. The remaining 30 residues are distributed throughout the sequence. The barrel domains are longer for the larger enzymes, with this increase resulting from the lengths of the helices that connect the strands of the barrel (Table 3). YcjG, in particular, has very short helices that average 9.9 residues in length; this contrasts with the average lengths for the remaining members of this subfamily which range from 11.5 to 12.7 (Table 3). The capping domains, likewise, are longer for the larger members of this family (Table 3). Again, the

C-terminal extension accounts for some of this difference, but the N-terminal portion of the capping domain is also longer for both MLE and YkfB.

That the larger enzymes, MLE and YkfB, are both octamers while the smaller enzymes are either dimeric (YcjG) or monomeric (OSBS) suggests a relationship between these two features. A reduction in size by deleting extra residues on surface loops is thought to help stabilize proteins (*28*). Similarly, the formation of multimeric protein complexes may also be important for the stabilization of proteins. These two features may offset each other so that the two octameric proteins face less pressure to eliminate residues from surface loops.

The octameric YkfB has 422 symmetry and can be considered a tetramer of dimers (Figure 5). There are no interactions across the center of the octamer; in fact, a 14 Å pore is located in the center of the protein. The pore is lined by residues that form the third and fourth helices of the barrel domain. The protein contains four subunits that lie above the 2-fold axes (shown in red and magenta in Figure 5) and four subunits that lie below the 2-fold axes (in blue and cyan). Each subunit makes considerable contacts with the two closest subunits located across the 2-fold axes. The subunit interface that forms the minimal dimer (subunits 1 and 2) is composed of residues from the N-terminal capping domain. Interactions that form this dimer occur between the first and third helices of the capping domain and the loops that precede these helices. Specific interactions that occur between subunits 1 and 2 include hydrogen bonding interactions from the carbonyl oxygen of Leu 52 to the amide nitrogen of Thr 94, the carbonyl oxygen of Val 53 to the side chain hydroxyl of Thr 94, the carbonyl oxygen of Gly 56 to the side chain nitrogen of Asn 96, and the side chain carboxylate of Asp 60 and the side chain hydroxyl of Ser 64. Additionally, an ionic interaction occurs between Arg 251 of subunit 1 and Asp 225 of subunit 2. The entire C-terminal portion of the capping domain in YkfB is at the periphery of the octamer so that it makes no interactions with any other subunit.

Subunit 1 also makes significant contacts with subunit 8, the next closest neighbor across a different 2-fold axis. Interactions occur between residues that form the loop preceding and at the start of the second helix of the capping domain and the loop that follows the third helix of the capping domain. This interaction consists of a hydrogen bond from the amide nitrogen of Glu 83 from subunit 1 to the main chain carbonyl of Gly 124 of subunit 8. An additional hydrophobic stacking interaction occurs between the Phe 254 residues of both subunits.

Minor ionic interactions occur between subunit 1 and the two neighboring subunits that lie above the plane of the 2-fold axes, namely, subunits 3 and 7 in Figure 5. These involve the side chains of Lys 199 of subunit 1 and Asp 263 of subunit 3 and the side chains of Arg 260 of subunit 1 and Asp 226 of subunit 7. These interactions may help to maintain the octameric state of the protein.

The oligomerization state and size of YcjG and YkfB raise interesting evolutionary questions. It is anticipated that the primordial protein from which the enolase superfamily evolved was at one point a monomeric protein; however, given the abundance of $(\beta/\alpha)_8$-barrels, it is unknown if the first member of this group evolved from a more symmetric $(\beta/\alpha)_8$-barrel protein that was already oligomeric or evolved de novo from a monomeric fold. The vast bulk of proteins that contain a $(\beta/\alpha)_8$-fold are multimeric which suggests that there is an advantage to oligomerization. It has been suggested that oligomerization allows more efficient use of amino acid residues by generating proteins that utilize subunit interfaces as components of their active sites. This is not the case in the enolase superfamily where the active sites are maintained within a single protomer. Furthermore, as noted above, there is little correlation between the size of the protein and the oligomeric state within this superfamily since many of the larger subunits, including MLE and mandelate racemase, form octameric macromolecular assemblies. Thus, the small size of YcjG which is dimeric relative to the octameric Ykfb cannot be rationalized unless it reflects the evolutionary pressures of the temperature and environment of their respective organisms.

Comparison of the Active Sites of YkfB and OSBS. Because OSBS is the only member of the MLE subgroup to be structurally characterized with a ligand bound in the active site, a comparison of YkfB with OSBS may assist in understanding how YkfB catalyzes its reaction. The only major structural change is the closure of the loop between the first two strands of the capping domain over the active site in OSBS. While this loop is sufficiently ordered to be modeled in YkfB, it is expected that upon binding substrate, this loop will move closer to the substrate perhaps allowing interactions with the substrate. [This loop is disordered in MLE (*5*) and YcjG in the absence of substrate or inhibitor in the active site.] Because the sequence of this loop is not well conserved between YcjG and YkfB (only 3 of 10 residues are identical), this loop may also be responsible for the observed differences in substrate specificity (*9*).

All of the catalytic residues are similarly positioned in YkfB and OSBS with the exception of the side chain of Lys 162 in YkfB, the second lysine of the KXK motif which is expected to function as a general acid/base catalyst. This residue is well-ordered in all four subunits, but it is expected to adopt a different position in the presence of substrate. The difference in the lysine side chain may also reflect different requirements for the two enzymes. In the epimerase reaction, the lysines are expected to be located exactly opposite one another to allow both to donate a proton to the same intermediate. In the dehydration reaction catalyzed by OSBS, this is not a requirement as the two lysine residues perform different functions.

CONCLUSIONS

The structures of YcjG and YkfB, two members of the enolase superfamily of enzymes that recently were characterized as L-Ala-D/L-Glu epimerases (*9*), have been determined. These enzymes represent the two most divergent and structurally characterized members of the enolase superfamily that catalyze the same reaction. They differ significantly in size and yet maintain very similar active site architectures. Clearly, there are issues remaining concerning the optimal size of a polypeptide chain for a given enzymatic function. These are important within the context of protein engineering and directed evolution. Questions also remain with regard to the structural basis for the substrate specificity and the identification of the R- and S-specific bases for the epimer-

ization reaction. Additional mutagenic and crystallographic studies are in progress to address these issues.

REFERENCES

1. Babbitt, P. C., and Gerlt, J. A. (1997) *J. Biol. Chem. 272*, 30591−30594.
2. Gerlt, J. A., and Babbitt, P. C. (1998) *Curr. Opin. Chem. Biol. 2*, 607−612.
3. Gerlt, J. A., and Babbitt, P. C. (2001) *Annu. Rev. Biochem. 70*, 209−241.
4. Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R. J., Barrett, W. C., Reed, G. J., Rayment, I., Ringe, D., Kenyon, G. L., and Gerlt, J. A. (1996) *Biochemistry 35*, 16489−16501.
5. Helin, S., Kahn, P. C., Guha, B. L., Mallows, D. G., and Goldman, A. (1995) *J. Mol. Biol. 254*, 918−941.
6. Thompson, T. B., Garrett, J. B., Taylor, E. A., Meganathan, R., Gerlt, J. A., and Rayment, I. (2000) *Biochemistry 39*, 10662−10676.
7. Palmer, D. R., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., and Gerlt, J. A. (1999) *Biochemistry 38*, 4252−4258.
8. 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*, 10396−10401.
9. Schmidt, D. M. Z., Hubbard, B. K., and Gerlt, J. A. (2001) *Biochemistry 40*, 15707−15715.
10. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
11. Terwilliger, T. C. (1994) *Acta Crystallogr. D50*, 17−23.
12. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D55*, 849−861.
13. Collaborative Computational Project, No. 4 (1994) *Acta Crystallogr. D50*, 760−763.
14. Roussel, A., and Cabillau, C. (1991) in *Silicon Graphics Geometry Partners Directory*, Silicon Graphics, Mountain View, CA.
15. Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
16. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.
17. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins 11*, 281−296.
18. Cohen, G. H. (1997) *J. Appl. Crystallogr. 30*, 1160−1161.
19. Satow, Y., Cohen, G. H., Padlan, E. A., and Davies, D. R. (1986) *J. Mol. Biol. 190*, 593−604.
20. Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) *Acta Crystallogr. D55*, 484−491.
21. Jones, S., and Thornton, J. M. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 13−20.
22. Larsen, T. M., Wedekind, J. E., Rayment, I., and Reed, G. H. (1996) *Biochemistry 35*, 4349−4358.
23. Lebioda, L., and Stec, B. (1991) *Biochemistry 30*, 2817−2822.
24. Wedekind, J. E., Poyner, R. R., Reed, G. H., and Rayment, I. (1994) *Biochemistry 33*, 9333−9342.
25. Duquerroy, S., Camus, C., and Janin, J. (1995) *Biochemistry 34*, 12513−12523.
26. Gulick, A. M., Palmer, D. R. J., Babbitt, P. C., Gerlt, J. A., and Rayment, I. (1998) *Biochemistry 37*, 14358−14368.
27. Gulick, A. M., Hubbard, B. K., Gerlt, J. A., and Rayment, I. (2000) *Biochemistry 39*, 4590−4602.
28. Walden, H., Bell, G. S., Russell, R. J., Siebers, B., Hensel, R., and Taylor, G. L. (2001) *J. Mol. Biol. 306*, 745−757.

BI011641P