Evolution of Enzymatic Activities in the Enolase Superfamily: Crystallographic and Mutagenesis Studies of the Reaction Catalyzed by D-Glucarate Dehydratase from *Escherichia coli*[†,‡]

Andrew M. Gulick,[§] Brian K. Hubbard,[||] John A. Gerlt,*[,||] and Ivan Rayment*[,§]

Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53705, and Departments of Chemistry and Biochemistry, University of Illinois, Urbana, Illinois 61801

Received December 6, 1999; Revised Manuscript Received January 31, 2000

ABSTRACT: D-Glucarate dehydratase (GlucD) from *Escherichia coli* catalyzes the dehydration of both D-glucarate and L-idarate as well as their interconversion via epimerization. GlucD is a member of the mandelate racemase (MR) subgroup of the enolase superfamily, the members of which catalyze reactions that are initiated by abstraction of the α-proton of a carboxylate anion substrate. Alignment of the sequence of GlucD with that of MR reveals a conserved Lys-X-Lys motif and a His-Asp dyad homologous to the *S*- and *R*-specific bases in the active site of MR. Crystals of GlucD have been obtained into which the substrate D-glucarate and two competitive inhibitors, 4-deoxy-D-glucarate and xylarohydroxamate, could be diffused; D-glucarate is converted to the dehydration product, 5-keto-4-deoxy-D-glucarate (KDG). The structures of these complexes have been determined and reveal the identities of the ligands for the required Mg^{2+} (Asp^{235}, Glu^{266}, and Asn^{289}) as well as confirm the expected presence of Lys^{207} and His^{339}, the catalytic bases that are properly positioned to abstract the proton from C5 of L-idarate and D-glucarate, respectively. Surprisingly, the C6 carboxylate group of KDG is a bidentate ligand to the Mg^{2+}, with the resulting geometry of the bound KDG suggesting that stereochemical roles of Lys^{207} and His^{339} are reversed from the predictions made on the basis of the established structure−function relationships for the MR-catalyzed reaction. The catalytic roles of these residues have been examined by characterization of mutant enzymes, although we were unable to use these to demonstrate the catalytic independence of Lys^{207} and His^{339} as was possible for the homologous Lys^{166} and His^{297} in the MR-catalyzed reaction.

Enzyme superfamilies have long been thought to consist of a divergently evolved set of proteins whose members catalyze the same overall chemical reaction, although the precise identities of the substrates might differ among the family members. This divergence is thought to result from changes in the active site that alter substrate specificity while maintaining the same catalytic groups. For example, the serine protease family of trypsin, chymotrypsin, and elastase is commonly thought to have evolved through changes in the pocket that provide specific binding interactions with the side chain of the amino acid residue of the substrate that contributes the carbonyl group of the scissile peptide bond.

Recently, an enhanced definition of superfamilies was proposed to include those groups of divergently related enzymes that catalyze transformations involving the same or chemically similar intermediates (and/or transition states) (*1*, *2*), rather than using a classification that requires catalysis of the same chemical reaction. These superfamilies are termed "mechanistically diverse". The enolase superfamily is a good example of a mechanistically diverse superfamily (*1*).

All members of the enolase superfamily catalyze the initial abstraction of a proton from a carbon adjacent to a carboxylate group (α-proton). The fate of the enolic intermediate so generated is determined by the divergent features of the active site architecture, where the overall reactions include racemization, cycloisomerization, and the β-elimination of either water or ammonia using diverse carboxylate anions as substrates. A significant number of X-ray structures for this family has been determined, including enolase (*3*−*5*), mandelate racemase (*6*), and muconate lactonizing enzyme (*7*, *8*). Each of these enzymes contains an $(\beta/\alpha)_7\beta$-barrel domain where the catalytic functional groups involved in the formation of the enolic intermediate are located in the pocket formed by the C-termini of the β-strands of the barrel. A second domain consisting of the N-terminal third of the protein, as well as residues that follow the barrel domain in some enzymes, also contributes residues to the active site, interacting with regions of the substrate other than the carboxylate group.

[†] 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. A.M.G. was supported by NRSA Fellowship AR-08422. 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 of the different GlucD complexes have been deposited in the Protein Data Bank (native protein, file name 1EC7; product complex, file name 1EC8; hydroxamate inhibitor, file name 1EC9; and 4-deoxy inhibitor, file name 1ECQ).

* To whom correspondence may be addressed. I.R.: Department of Biochemistry, 1710 University Ave., Madison, WI 53705; phone, (608) 262-0529; fax, (608) 262-1319; e-mail, ivan_rayment@biochem.wisc.edu. J.A.G.: Department of Biochemistry, University of Illinois, Urbana, IL 61801; phone, (217) 244-7414; fax, (217) 265-0385; e-mail, j-gerlt@uiuc.edu.

[§] University of Wisconsin.

[||] University of Illinois.

10.1021/bi992782i CCC: $19.00
Published on Web 03/30/2000

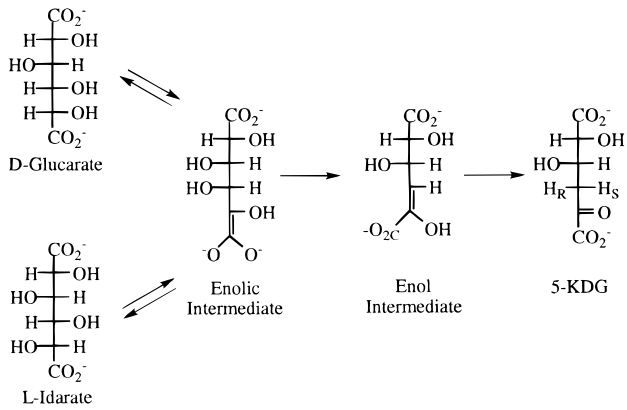
The enzymes in the enolase superfamily can be assigned to three subgroups largely on the basis of the identities of the general basic residues involved in formation of the enolic intermediate (*1*): (1) the mandelate racemase (MR) subgroup, (2) the muconate lactonizing enzyme (MLE) subgroup, and (3) the enolase subgroup. The enolase subgroup utilizes a single lysine residue; the MR subgroup utilizes one lysine and/or a histidine, and the MLE subgroup utilizes two lysine residues. In each subgroup, a required divalent metal ion is coordinated to three functional groups, with the metal ion thought to be coordinated to the substrate via at least one oxygen atom of the substrate's carboxylate group. This direct coordination provides significant stabilization of the enolic intermediate, thereby explaining the ability of these enzymes to catalyze the thermodynamically difficult abstraction of the α-proton of a carboxylate anion ($pK_a \geq 29$).

Recognition that these enzymes use a common structural framework for facilitating the formation and stabilization of an enolic intermediate has allowed the use of sequence relationships for identifying other members of this superfamily in protein databases (*1*). Numerous microbial proteins having unknown functions have been assigned to this superfamily (*9*). Although the types of reactions catalyzed by these unknown proteins, e.g., racemization/epimerization or dehydration, often can be specified on the basis of the number and identities of functional groups that can be assigned to the active sites at the C-terminal ends of the β-strands in the barrel domain, the exact substrate cannot be specified because the binding determinants are located in flexible loops in the N-terminal domain. With this approach, D-glucarate dehydratases (GlucDs)[1] from *Pseudomonas putida*, *Escherichia coli*, and *Bacillus subtilis* were assigned to the MR subgroup; alignment of the sequences of these GlucDs with that of MR revealed a conserved Lys-X-Lys (KXK) motif and a His-Asp dyad homologous to the *S*- and *R*-specific bases in the active site of MR. As a result, the GlucDs from *P. putida* (*9*) and *E. coli* (*10*) were predicted and found to catalyze not only the dehydration of D-glucarate, a naturally occurring substrate, but also the dehydration of L-idarate, an unnatural substrate, and the epimerization reaction that interconverts D-glucarate and L-idarate (Scheme 1).

GlucDs catalyze the dehydration of D-glucarate and L-idarate to form the same product, 5-keto-4-deoxy-D-glucarate (KDG). Solvent deuterium is incorporated stereospecifically into the KDG product at the 4-*proS* position, demonstrating that the final tautomerization step of the reaction occurs within the stereochemically constrained environment of the active site and that the displacement of the 4-OH group occurs with retention of configuration (*11*). Substantial substrate deuterium isotope effects (≥3) provide evidence that the abstraction of the α-proton is at least partially rate-determining in both dehydration reactions (*12*).

The dehydration reactions as well as the epimerization of D-glucarate and L-idarate can be rationalized by the participation of a common enolic intermediate (Scheme 1) (*13*). The

Scheme 1



ratio of the rates of dehydration and epimerization is ~4:1 with D-glucarate as the substrate and reflects the partitioning of the enolic intermediate between vinylogous β-elimination (dehydration) and protonation (epimerization). Epimerization occurs with incorporation of solvent hydrogen at C5 in the product; i.e., in D_2O, D-glucarate is converted to 5-[^{2}H]-L-idarate and L-idarate is converted to 5-[^{2}H]-D-glucarate. This behavior is diagnostic of a two-base mechanism for the inversion of configuration, with one active site base abstracting the α-proton from the substrate to generate the enolic intermediate and the conjugate acid of a second base protonating the opposite face of the intermediate to yield the product.

The structure of GlucD from *P. putida* confirmed both membership in the enolase superfamily and the fact that its putative active site is most similar to that of mandelate racemase (*14*); a KXK motif and a His-Asp dyad homologous to the *S*- and *R*-specific bases in the active site of MR, respectively, were found on opposite faces in the active site. Research on the *P. putida* enzyme demonstrated the enormous predictive value of a common structural framework for catalyzing similar chemical reactions that are related by formation of a common enolic intermediate.

Even so, many important questions were unanswered:

(1) The structure of GlucD from *P. putida* was obtained in the absence of a divalent metal ion; however, the putative site of the metal ion was found by soaking the crystal in $MnCl_2$. As such, the locations and identities of the expected three ligands provided by the protein for the catalytically essential divalent metal could not be specified unequivocally. This has now been achieved by cocrystallizing the enzyme with $MgCl_2$, the native metal ligand.

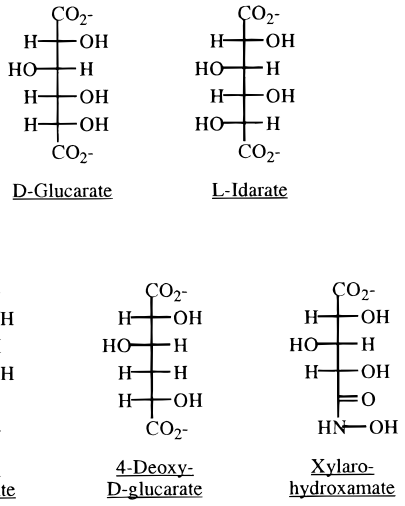
(2) In the absence of bound substrate, product, or inhibitor, the orientation of the bound substrate relative to the KXK motif and the His-Asp dyad in the active site of GlucD could not be predicted, so the stereochemical assignments of the roles of these putative acid/base catalysts in abstracting the α-proton from the D-glucarate and L-idarate substrates could not be determined.

(3) The vinylogous β-elimination of the hydroxyl on C4 from the enolic intermediate should require an acidic catalyst, but its location and identity could not be specified.

(4) The stereospecific tautomerization of the enol produced by the vinylogous β-elimination should also require an acid catalyst. It too could not be located.

[1] Abbreviations: GlucD, D-glucarate dehydratase; MePEG, methyl ether polyethylene glycol; MLE I, muconate lactonizing enzyme I; MR, mandelate racemase; SBC, Structural Biology Center at the Advanced Photon Source, Argonne, IL; KDG, 5-keto-4-deoxy-D-glucarate; SAA, *S*-atrolactate.

Scheme 2[a]



[a] All compounds are drawn with C1 at the top.

(5) The interactions of the distal portions of the substrates (presumably the same for D-glucarate and L-idarate since these share the same configurations at carbons 2−4) with flexible loops in the N-terminal domain could not be specified, so the structural basis for specificity was unknown. Defining and understanding the structural features that provide specificity is essential for predicting the identities of the substrates for unknown proteins that are members of this superfamily.

Efforts to obtain crystals of GlucD from *P. putida* in the presence of substrate, product, or substrate analogues failed. As such, we turned to the homologous (72% level of sequence identity) GlucD from *E. coli*. Like the GlucD from *P. putida*, the enzyme from *E. coli* also catalyzes the dehydration of L-idarate as well as the epimerization of D-glucarate and L-idarate with comparable kinetic constants (*10*). As reported here, the GlucD from *E. coli* crystallizes in a form that is amenable to a high-resolution structural study in the presence of Mg^{2+}; this crystal form can convert D-glucarate to KDG as well as bind two substrate analogues (Scheme 2). These studies define the geometric relationship of the substrates to the KXK motif and the His-Asp dyad as well as describe the interactions of the active site with the portions of the substrate distal from the sites of α-proton abstraction and β-elimination of the hydroxyl (carbons 1−3). The roles of residues that are proposed to interact with the substrate were examined by kinetic characterization of site-directed mutants. Taken together, the structural and mutagenic studies reveal an unexpected diversity of function that could not be predicted from our knowledge of the mechanism of the reaction catalyzed by MR, the structure−function "paradigm" for GlucD.

MATERIALS AND METHODS

Synthesis of 4-Deoxy-D-glucarate. D-Glucarate (5 g) reacted to completion with GlucD in 25 mL of 50 mM Tris-HCl (pH 8) containing 10 mM $MgCl_2$. Two equivalents of $NaBH_4$ dissolved in 25 mL of water was added to the solution and the mixture allowed to stir for 3 h. The reaction was stopped by addition of acetic acid to pH <7. The pH was adjusted to 10, and the resulting mixture of 4-deoxy-D-glucarate and 4-deoxy-L-idarate was purified and separated by anion exchange chromatography (Dowex AG1-X8, formate form) with a linear gradient of 0 to 3 M formic acid. The fractions containing acid sugars were identified by spotting aliquots on silica thin-layer chromatography plates and dipping the dried plate in an acidic ethanolic solution of *p*-anisaldehyde (5.1 mL of *p*-anisaldehyde, 6.9 mL of concentrated sulfuric acid, 2.1 mL of acetic acid, and 186 mL of 95% ethanol). 4-Deoxy-D-glucarate and 4-deoxy-L-idarate were separated by this procedure (*10*).

Syntheses of 2,5-[$^2H_2$]-L-Idarate. 5-Keto-D-fructose was prepared by a literature procedure (*15*). A solution of sodium borodeuteride (0.75 g) in 30 mL of water was added to a solution of 5-keto-D-fructose (5 g) in 50 mL of water. The reaction mixture was stirred for 3 h and the reaction stopped by the addition of acetic acid to pH <7. After addition of methanol (20 mL), the solvent was removed by rotary evaporation. Nitric acid (30 mL) was added along with a catalytic amount of sodium nitrate. The mixture was heated at 70 °C for 3 h. After removal of the nitric acid by lyophilization, the resulting solid was dissolved in water, and the pH was adjusted to 10. 2,5-[$^2H_2$]-L-Idarate was resolved from 2,5-[$^{22}H_2$]-D-glucarate and 2,5-[$^2H_2$]-D-mannarate by anion exchange chromatography as described in the previous section.

Synthesis of 2,3,4,5-[$^2H_4$]-D-Glucarate. 2,3,4,5-[$^2H_4$]-D-Glucarate was prepared from commercially available 2,3,4,5-[$^2H_4$]-D-glucose (Cambridge Isotope Laboratories) by nitric acid oxidation and purified as previously described.

Synthesis of D,L-Xylarohydroxamate. The pH of a 4 M solution of hydroxylamine-HCl was adjusted to 10.5 and added with stirring to a solution of xylarolactone [5 g (*16*)] in 15 mL of water. After 20 min, the pH was adjusted to 7.2, and the racemic xylarohydroxamate was purified by anion exchange (Dowex AG1-X8, chloride form) using a linear gradient of 0 to 2 M LiCl. The fractions containing the hydroxamate were evaporated to a syrup, and the product was precipitated by addition of 200 mL of a mixture of methanol and acetone (4:1) (*10*).

Purification of GlucD from E. coli. GlucD from *E. coli* was purified as described previously (*17*). The protein was dialyzed into 10 mM Tris-HCl (pH 8.0) at 4 °C and 5 mM $MgCl_2$ and concentrated to 14 mg/mL. The protein was frozen in small aliquots by pipetting directly into liquid nitrogen, and the resulting pellets of protein were stored at −80 °C and thawed for crystallization experiments as needed.

Crystallization of GlucD from E. coli. Purified protein was crystallized by microbatch experiments. An equal volume of precipitant containing 14% MePEG 5000 (Aldrich Chemical Co., Milwaukee, WI), 75 mM $MgCl_2$, 5% 2-propanol, and 50 mM Hepps (pH 8.0 at 4 °C) and protein were combined and macroseeded with a small crystal (0.05 mm × 0.05 mm × 0.05 mm) that was grown spontaneously by hanging drop methods. The crystals grew to a size of 0.6 mm × 0.6 mm × 0.3 mm over a period of 1 month. The crystals belong to triclinic space group *P*1 with the following unit cell dimensions: $a = 71.2$ Å, $b = 84.5$ Å, $c = 98.9$ Å, $\alpha = 103.3°$, $\beta = 94.0°$, and $\gamma = 113.1°$. A single tetramer is present in the unit cell, giving a Matthews coefficient of 2.65 Å^3/Da (53% solvent) (*18*).

Table 1: Data Collection Statistics

	native at 4 °C	native at −160 °C	glucarate at −160 °C	4-deoxyglucarate at −160 °C	xylarohydroxamate at −160 °C
no. of crystals	2	1	1	1	1
resolution (Å)	2.4	1.9	1.9	2.0	2.0
R_{merge} (%)	9.2	7.0	5.5	5.6	4.4
highest shell (%)	26.3	20.2	18.2	32.1	18.9
no. of unique reflections	96897	150262	152351	131655	133202
redundancy	2.5	2.5[a]	3.0[a]	3.9	3.9
completeness (%)	90.7	95.2	96.2	97.5	97.3
highest shell (%)	75.7	90.0	93.3	94.2	93.5
average intensity	5150	3186	3179	8060	9692
average σ	719	253	200	301	321
unit cell dimensions					
a (Å)	72.0	71.3	71.1	71.1	71.3
b (Å)	85.9	84.6	84.6	84.6	84.8
c (Å)	99.9	98.9	98.8	98.8	99.0
α (deg)	102.3	103.3	103.4	103.5	103.1
β (deg)	94.5	94.0	94.0	94.0	94.3
γ (deg)	113.3	113.1	113.1	113.0	113.2

[a] Redundancy for the frozen data sets was calculated to 1.7 Å.

Data Collection. Two data sets were used in the determination of the structure of the native GlucD protein. Initially, a low-resolution data set was collected with laboratory X-ray sources. This data set was used to determine the structure by molecular replacement with the program AMORE (*19*). Subsequently, a higher-resolution data set was obtained at the SBC ID19 line of the Advanced Photon Source at Argonne National Laboratory (Argonne, IL) for the native protein and for the three complexes with product and inhibitor molecules.

The laboratory data were collected with Cu Kα radiation from a Rigaku RU200 X-ray generator operating at 50 kV and 90 mA with Supper long focusing mirrors and a Siemens Hi-Star area detector. Data were processed with XDS (*20*, *21*) and were scaled with XSCALIBRE (*22*). Two crystals were used in the data collection; the merging *R*-factor for the two sets combined (9.8%) was only slightly higher than the R_{merge} values for the two individual data sets (7.5% and 7.0%). Data collection statistics are shown in Table 1.

High-resolution data sets were collected at the Advanced Photon Source from a crystal frozen to −160 °C. The crystal was transferred into the cryoprotectant ethylene glycol in incremental steps. The series of cryoprotectant solutions contained 15% methyl ether PEG 5000, 100 mM $MgCl_2$, 5% 2-propanol, and 1%, 10%, or 20% ethylene glycol, respectively. The crystal was incubated for approximately 1 min in each step of ethylene glycol. Additional complexes were studied by including substrate or inhibitor in the cryoprotectant solutions. In the first of the experiments in which small molecules were introduced into the crystals, the cryoprotectant solutions each contained 95 mM D-glucarate (Sigma Chemical Co., St. Louis, MO). For the first inhibitor complex, racemic xylarohydroxamate was included in the three cryoprotectant steps at concentrations of 44, 64, and 84 mM, respectively. For the second inhibitor complex, 4-deoxy-D-glucarate was included in the three cryoprotectant steps at concentrations of 21, 31, and 41 mM, respectively. After the brief incubation with cryoprotectant and substrate or inhibitor, lasting less than 5 min, the crystals were flash-frozen in a stream of nitrogen gas at −160 °C.

The high-resolution data sets were obtained at the SBC ID19 line of the Advanced Photon Source at Argonne National Laboratory. A full 360° of data were collected on a 3 × 3 mosaic area detector at a distance of 140 mm (*23*). The detector was oriented at a 2θ angle of 0°, which allows a theoretical maximum of four measurements of each reflection. The data were processed and scaled with HKL2000 (*24*) and remapped into the correct asymmetric unit with the TNT package (*25*, *26*).

Structure Determination. The structure of GlucD from *P. putida* (*14*) served as the search model for molecular replacement. Nonconserved side chains were removed, and the structure was expanded into a tetramer of 1580 amino acid residues on the basis of the crystallographic symmetry of the GlucD from *P. putida*. The temperature factors were all set to 25.0 Å^2. A structural solution for the data set taken at 4 °C was found with the molecular replacement program AMORE (*19*). Four symmetry-related solutions were identified. The final solution had a crystallographic *R*-factor of 36.7% for data from 20.0 to 4 Å and a correlation coefficient of 70.0% (*19*). The model was subjected to 20 cycles of least-squares refinement with TNT (*25*, *26*). The crystallographic *R*-factor dropped from 44.9% to 32.0% against all data to 2.4 Å resolution. One subunit of the model was manually built with FRODO (*27*) which allowed the inclusion of 33 side chains which were not conserved between the *Pseudomonas* and *E. coli* sequences. Six disordered side chains were removed. Additionally, seven residues of the missing C-terminus were built into the model. This improved monomer was expanded into a tetramer and again subjected to further least-squares refinement. This model then served as the molecular search model for molecular replacement (*28*) with data sets obtained with frozen crystals.

The structure of the crystals frozen in the presence of 94 mM D-glucarate was determined first. Molecular replacement with the partially refined model for the data recorded at 4 °C as a search model gave unambiguous solutions with an *R*-factor of 36.7% and a correlation coefficient of 64.1%. This model was refined with TNT to an *R*-factor of 33.4% against data to 2.2 Å resolution. Difference density was observed at the active site, although the substrate was not added at this time. In the first round of model building with FRODO (*27*), 119 side chains were added and 13 new residues were included. At the same time, 30 side chains

Table 2: Final Refinement Statistics for the Different GlucD Complexes

	native	KDG	4-deoxyglucarate	xylarohydroxamate
resolution (Å)	1.9	1.9	2.0	2.0
no. of reflections used	152199	153612	127330	130580
no. of total atoms	14443	14943	14643	14854
no. of protein atoms/residues	13321/1729	13637/1768	13666/1772	13861/1774
no. of solvent atoms	1102	1234	901	1097
no. of other atoms[a]	20	72	72	72
average *B*-factor (Å^2)				
protein	28.2	26.1	39.4	30.7
main chain	25.7	24.0	37.0	28.4
solvent	34.9	31.7	44.2	37.0
active site[b]	22.9	25.4	46.7	34.5
R-factor (%)	19.4	19.6	20.9	17.9
rms deviation				
lengths (Å)	0.015	0.016	0.014	0.013
angles (deg)	2.00	2.07	2.01	1.90
trigonal planes (Å)	0.004	0.005	0.004	0.004
general planes (Å)	0.013	0.014	0.012	0.012

[a] Other atoms refer to 2-propanol, Mg^{2+} ions, and product or inhibitor molecules. [b] Active site temperature factor includes the Mg^{2+} ion and the product or inhibitor molecule.

were removed where density was not apparent and three residues were deleted. Least-squares refinement reduced the *R*-factor to 29.5%. Mg^{2+} ions were included at this stage of the refinement. The cycles of refinement and model building continued in this manner with TURBO-FRODO (*29*). When the *R*-factor had dropped to 27.6% against data to 2.2 Å resolution, water molecules were added manually and with the PEKPIK program of the TNT package (*25*, *26*). The C-termini of each of the four subunits became visible as the crystallographic *R*-factor decreased. Similarly, it was possible to model $Asp^{93}-Leu^{106}$, the break in the N-terminal domain that was observed in the structure of GlucD from *P. putida* (*14*). When the *R*-factor had decreased to less than 23%, the substrate molecules were included. Initially, the model included only C1−C3 and their associated oxygens; the density for C4−C6 was ambiguous. As refinement continued, we realized that the active site of each polypeptide in the tetramer contained a molecule of KDG, the reaction product, not the substrate. Thereafter, data to 1.9 Å resolution were included in the refinement and model building to yield the final structure which contains 1768 amino acid residues, 4 metal ions, 4 molecules of product, and 1234 solvent molecules. Final refinement statistics for all product complex and native and inhibitor complexes described below are listed in Table 2.

Unliganded GlucD. The structure of GlucD in the presence of Mg^{2+} but in the absence of substrate, product, or inhibitors (unliganded GlucD) was determined to high resolution from a data set collected at −160 °C with the SBC ID19 line at the Advanced Photon Source. The structure was determined by Fourier methods beginning with a nearly complete model of the product complex, from which solvent, metal, and product atoms had been removed. Although the loop from Asp^{93} to Leu^{106} in the N-terminal domain was included in the starting model, it was immediately obvious that this loop is disordered in the native structure. The refinement and model building progressed as with the product complex. Mg^{2+} ions and solvent molecules were included where clear density and geometry allowed. The refinement continued to a final *R*-factor of 19.4% (Table 2).

Inhibitor Complexes. Data sets for the two inhibitor complexes also were obtained with the SBC ID19 line at the Advanced Photon Source. These data sets were determined and refined in essentially the same manner as described above. The phase problem was solved by molecular replacement with a search model based on the product complex from which solvent and product molecules had been removed. This method was used rather than difference Fourier methods to accommodate small shifts in the protein during freezing. The solution from AMORE (*19*) had a crystallographic *R*-factor of 30.5% and 29.4%, to 4.0 Å resolution, for the 4-deoxy-D-glucarate and xylarohydroxamate complexes, respectively. Following molecular replacement, a subset of 5% of the data was removed and used for a calculation of a value for *R*-free. The models were refined with TNT and manually built with TURBO-FRODO (*29*) on an SGI Indigo^2 workstation. When the *R*-factor had decreased below 23%, inhibitor molecules and metal ions were added to the active sites of each polypeptide in the tetramer. The loops from Asp^{93} to Arg^{107} in the N-terminal domains of each polypeptide remained well-ordered in the models of the inhibitor complexes. The *R*-free values, calculated at the penultimate refinement step, are 32.1% and 28.1% for the 4-deoxy-D-glucarate and xylarohydroxamate complexes, respectively. All of the data were then included for the final round of model-building and least-squares refinement.

Final Models. Each model is composed of four nearly identical subunits that form the tetramer. For the three complexes with product or inhibitors, the only missing residues are at the N-terminus: Met^1-Gln^4 for all four subunits in the product complex; Met^{1A}, Ser^{2A}, $Met^{1B}-Gln^{4B}$, $Met^{1C}-Ser^{3C}$, and $Met^{1D}-Gln^{4D}$ for the 4-deoxyglucarate complex; and $Met^{1A}-Gln^{4A}$, $Met^{1B}-Ser^{3B}$, $Met^{1C}-Ser^{3C}$, Met^{1D}, and Ser^{2D} for the xylarohydroxamate complex. The unliganded model contains the same missing N-terminal residues ($Met^{1A}-Gln^{4A}$, $Met^{1B}-Gln^{4B}$, $Met^{1C}-Gln^{4C}$, and $Met^{1D}-Ser^{3D}$) as well as the missing loop in the N-terminal domain, $Asp^{93A}-Asp^{105A}$, Ala^{96B}, Gly^{97B}, $Asp^{93C}-Leu^{106C}$, and $Asp^{95D}-Leu^{106D}$. In the refinement of the unliganded and product-bound models, two water molecules that were positioned close to each other refined to form a continuous piece of density with three lobes. It became apparent that this density was consistent with a molecule of 2-propanol,

Table 3: Root-Mean-Square Differences between Complexes (Å)[a]

	native	KDG	4-deoxyglucarate	xylaro-hydroxamate
native	0.23[b]	0.20[c]	0.16[c]	0.20[c]
KDG		0.19[b]	0.17[c]	0.14[c]
4-deoxyglucarate			0.24[b]	0.18[c]
xylarohydroxamate				0.19[b]

[a] The coordinates were superimposed with the program ALIGN (*42*). [b] Average rms difference in α-carbons of the four subunits within each complex. [c] The rms differences between the α-carbons of all four subunits in the cross-referenced complexes.

which was present in the crystallization conditions as well as the cryoprotectant soaks. Four 2-propanol molecules were found at equivalent positions in the interface of the four subunits of the tetramer in each model presented here. The geometry of each structure was checked with PROCHECK (*31*). The only residue found in the disallowed region of the Ramachandran plot was Thr^{318} in each of the four subunits ($\phi \sim 80°$, $\psi \sim 160°$); the density for this residue was unambiguous, and no reason is obvious why the residue should adopt this odd conformation. Unless otherwise specified, reported distances are the average of those found in the four subunits for a given model. Figures were generated with MOLSCRIPT (*32*, *33*). The structures for Figure 4 were aligned using LSQKAB (*34*, *35*) by aligning the Cα positions of the following residues from the β-strands of the barrel domain (the first residue for each pair is from MR): Lys^{164}−Lys^{205}, Ile^{165}−Leu^{206}, Val^{194}−Leu^{234}, Asp^{195}−Asp^{235}, Ile^{220}−Ala^{259}, Glu^{221}−Glu^{260}, Ser^{296}−Ser^{338}, and His^{297}−His^{339}. The rms differences between the subunits of each complex and between the complexes are shown in Table 3.

Assays for Dehydration of D-Glucarate and L-Idarate. The formation of KDG was assayed using semicarbazide to quantitate the α-ketoacid product. For routine assays, an aliquot containing GlucD was added to 1 mL of 10 mM acid sugar substrate in 50 mM Tris-HCl (pH 7.5) containing 5 mM $MgCl_2$ at 22 °C. The amount of α-ketoacid was quantitated by detection of its semicarbazone at 250 nm. Values of k_{cat} and k_{cat}/K_m were determined by varying the concentration of the acid sugar substrate.

NMR Assay for Epimerization of D-Glucarate and L-Idarate. Samples (600 μL) contained 100 mM acid sugar in 50 mM Tris-HCl (pH 7.5) containing 10 mM $MgCl_2$; the solvent was 95% H_2O and 5% D_2O. After shimming, GlucD was added to initiate the reaction, and spectra were acquired at 1 min intervals. Data were analyzed by integrating the resolved resonances for D-glucarate, L-idarate, KDG, and Tris-HCl (internal standard). Automated integration was performed using the program FINT developed by F. Lin in the NMR Spectroscopy Laboratory in the School of Chemical Sciences at the University of Illinois. The data were plotted and analyzed with Microsoft Excel.

Site-Directed Mutagenesis. Site-directed mutants were constructed with the following strategy. Two nonoverlapping oligonucleotide primers directing extension in opposite directions from the position of the desired mutation were used in separate PCRs to obtain two blunt-ended fragments that, when ligated, would reconstruct the full-length gene encoding GlucD containing the mutation. One reaction utilized a primer containing the mutation at its 5′-terminus and a primer complementary to a sequence outside the gene that contained a unique restriction site (*Xba*I at the 5′-end or *Xho*I at the 3′-end) in the expression plasmid; the second reaction utilized a primer containing a wild-type sequence and a primer complementary to a sequence outside the other end of the gene with the second unique restriction site. After purification, restriction digestion, and 5′-phosphorylation, these were ligated with the vector (pET15b that had been modified to encode an N-terminal His tag containing 10 His residues) that had been restricted with *Xba*I (following the His tag) and *Xho*I (preceding the T7 terminator sequence). The resulting ligation mixture was transformed into *E. coli* strain XL1-Blue; the plasmids contained in transformants were screened for the presence of the full-length gene encoding the putative mutant of GlucD by PCR analysis. Following transformation of these plasmids into *E. coli* strain BL21(DE3), the encoded GlucDs were isolated as described previously. After the mass of the isolated protein was confirmed by MALDI-TOF mass spectrometry, the entire gene encoding the mutant protein was sequenced to verify the presence of the desired mutation and to verify that no other changes had been introduced.

RESULTS AND DISCUSSION

The enolase superfamily of enzymes consists of a large group of enzymes with a limited degree of sequence homology that catalyze reactions that are initiated by abstraction of the α-proton of a carboxylate anion. On the basis of previous studies of enolase, MR, and MLE, it was postulated that the acidity of this proton is increased by the stabilization of the resulting enolate anion by coordination to a metal ion (*1*, *2*). The study presented here both defines the active site architecture of GlucD by X-ray crystallography and uses that information to interpret the phenotypes of several site-directed substitutions for active site residues. This new information is useful for assessing the structural strategies used for both abstraction of the α-proton and stabilization of the enolate anion intermediate in the enolase superfamily.

Structure of the Unliganded Enzyme. GlucD from *E. coli* was crystallized in the presence of 75 mM $MgCl_2$ but in the absence of substrate, product, or inhibitors. The structure was determined by molecular replacement utilizing the structure of the homologous GlucD from *P. putida* as the search model. The polypeptide of the GlucD from *E. coli* (446 residues) is five residues shorter than that of the GlucD from *P. putida*; their sequences are 72% identical over their entire sequences. The GlucD from *E. coli* crystallized in triclinic space group *P*1 (Table 1) with a complete tetramer in the asymmetric unit. The search model for the initial molecular replacement was a tetramer of the GlucD from *P. putida* with nonconserved side chains removed and temperature factors set to 25.0 $Å^2$. Unique solutions were obtained, and the model was built and refined as described in Materials and Methods. As shown in Table 3, the rms differences between the four independent subunits of the tetramer are quite small (0.23 Å) and suggest that the crystal packing interactions have not induced any major changes in the structure.

A monomer of GlucD from *E. coli* consists of two domains (Figure 1), as observed with the GlucD from *P. putida* (*14*).

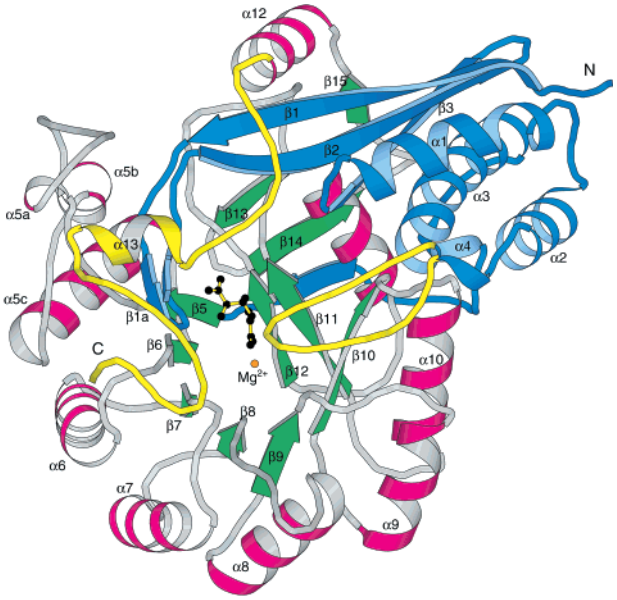


FIGURE 1: Three-dimensional structure of GlucD from *E. coli*. A ribbon tracing is shown for GlucD derived from the complex with KDG. Secondary structural elements are labeled throughout. The N-terminal domain is shown in blue; the barrel domain is shown with green strands and magenta helices. Two regions which were not well-ordered in the enzyme from *P. putida* (*14*) are shown in yellow. These include the N-terminal loop between α2 and α3 and the 28 C-terminal residues. The active site is shown at the C-termini of the strands to contain a Mg^{2+} ion and a molecule of 5-KDG.

An N-terminal domain that consists of three β-sheets and three α-helices precedes a C-terminal β/α-barrel domain. The loop between the two domains consists of residues Leu^{136}–Leu^{148} and contains a short β-strand. This strand, β4, forms a small antiparallel β-sheet with two strands (β13 and β14) that follow the barrel domain. Following these two strands at the C-terminus, the protein adopts a long loop and then an α-helix (α12). In the structure of GlucD from *P. putida* (*14*), the 30 C-terminal residues following α12 were disordered and could not be built into the model. Electron density for the entire C-terminus of the GlucD from *E. coli* is unambiguous. The C-terminus contains an extended loop that passes by the active site pocket in the barrel domain, a short helix, and then a coil domain that ends the polypeptide chain. This region, which interacts at the active site (see below), contributes to the binding of substrate and/or product by completing a cationic pocket around the C1 carboxylate group of the sugar.

The unliganded GlucD contains a Mg^{2+} ion in the active site that is coordinated by Asp^{235}, Glu^{266}, and Asn^{289} (Figure 2A).

The first two ligands for the Mg^{2+} were predicted from the sequence alignment with MR (*1*); however, the carboxamide third ligand was implicated only through crystallographic studies with the GlucD from *P. putida* (*14*).

The electron density for His^{339} suggests that this side chain may adopt two conformations when $\chi_1 = 60°$ and 180°. This residue is of interest because it is the putative general basic group that initiates the dehydration of D-glucarate (vide infra). The same disorder was observed in the structure of GlucD from *P. putida*, which also was determined in the absence of substrate or product.

Structure of the GlucD·KDG Complex. As noted above, the catalytically essential divalent metal ion is present but the remainder of the active site is empty in the unliganded enzyme. Thus, this crystal form provides an opportunity to investigate the interaction of the protein with substrates and inhibitors. The substrate D-glucarate was diffused into the crystal lattice. A priori, we did not know whether this experiment would yield meaningful results. However, several distinct possibilities could be imagined. (1) The reaction could occur in the lattice but induce a conformational change that would destroy the lattice. (2) The substrate could bind in the active site but be inert because of an inability of the protein to undergo a required conformational shift due to steric constraints in the crystal lattice (*36*, *37*). (3) The substrate could bind in a nonproductive manner due to its inability to induce a conformational change that is necessary for catalysis. (4) The reaction could occur in the lattice, and product would be either released or retained. (5) A mixture of substrate and product might be present in the active site and modeled during crystallographic refinement if the active site equilibrium constant is near unity (*5*, *38*). So, except for destruction of the crystal lattice or the inability of either substrate or product to bind, we anticipated that we would be able to obtain biochemical information relevant to the enzymatic reaction from this experiment. Additionally, the presence of four independent subunits of GlucD in the crystal lattice provides an additional internal control for the effects of crystal packing.

A crystal of GlucD was soaked in the presence of 94 mM glucarate for ~5 min during the cryoprotection procedure prior to freezing the crystal. Upon building and refinement of the model, the electron density revealed that the active sites of each of the four subunits contained the KDG product coordinated to a Mg^{2+} ion. The density for the product is unequivocal (Figure 2B). Thus, either the enzyme catalyzes a single turnover resulting in the bound product complex, or the enzyme has a sufficient affinity for the product such that it accumulates in the active site.

The structural model of the product complex includes the entire polypeptide except for the four residues at the N-terminus (Figure 1). Like the native enzyme, this complex contains an ordered C-terminus. Additionally, the loop in the N-terminal domain from Ala^{93} to Arg^{106} that was disordered in both the *E. coli* and *P. putida* unliganded structures is sufficiently ordered to be included in the final model. The average temperature factor for the loop is still somewhat high ($B_{ave} = 42.6$ Å^2) for residues Arg^{94}–Arg^{107}; however, the electron density is sufficient to model this region. The loop is an insertion relative to the sequences of MR, enolase, and MLE I, all of which have very short loops that connect α2 and α3. Enolase and MLE I have an extended loop between β3 and α1 which occupies some of the same space as this loop in GlucD (*4*, *5*, *7*, *8*). This region interacts with the product by forming a hydrogen bond between the Oγ1 of Thr^{103} and the C3 hydroxyl of the product. In addition, the side chain of Phe^{104} forms a hydrophobic stacking interaction with the methylene groups of Arg^{422}, which extends from the C-terminal loop of the protein to the active site. The position of Arg^{422} is further stabilized by an ionic interaction with Asp^{424}.

Both this structure and that of the unliganded enzyme depict the secondary structure of a 29-residue insertion (Gly^{155}–Met^{183}) in the sequence of GlucD relative to that

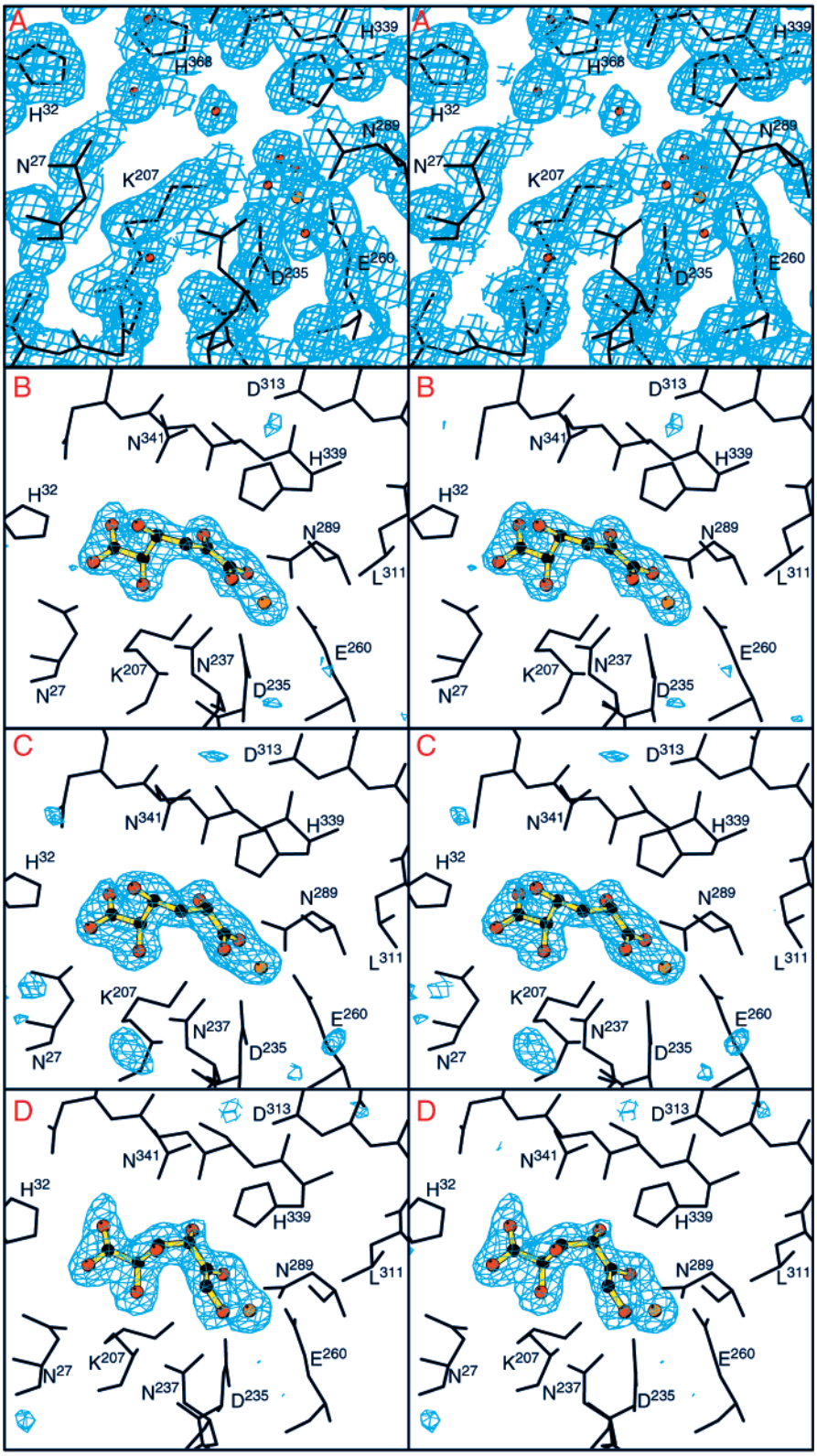


FIGURE 2: Representative electron density of all four complexes. (A) Electron density of the final model for the unliganded structure is shown. Density is derived from a map generated with coefficients in the form $2F_o - F_c$ and contoured at 1σ. The region that is shown is from subunit A of the native structure. Active site density is shown for (B) KDG, (C) 4-deoxyglucarate, and (D) xylarohydroxamate. For each figure, the product or inhibitor and Mg^{2+} ion were removed from the coordinate file and refinement was allowed to proceed for one cycle. An omit map was then calculated with the coefficients $F_o - F_c$ and contoured at 3σ. Active site residues are labeled.

of MR. In GlucD, an additional α-helix is located between the first and second strands of the barrel domain. Previously, sequence alignments were insufficient to identify the residues that comprise the first strand of the barrel, β5, due to a limited degree of homology. The structure now identifies a residue at the end of this strand, Tyr^{150}, that interacts with the substrate (vide infra).

The active site of the product complex is illustrated in schematic form in Figure 3a. Several interactions between GlucD and its substrates were predicted from the structures

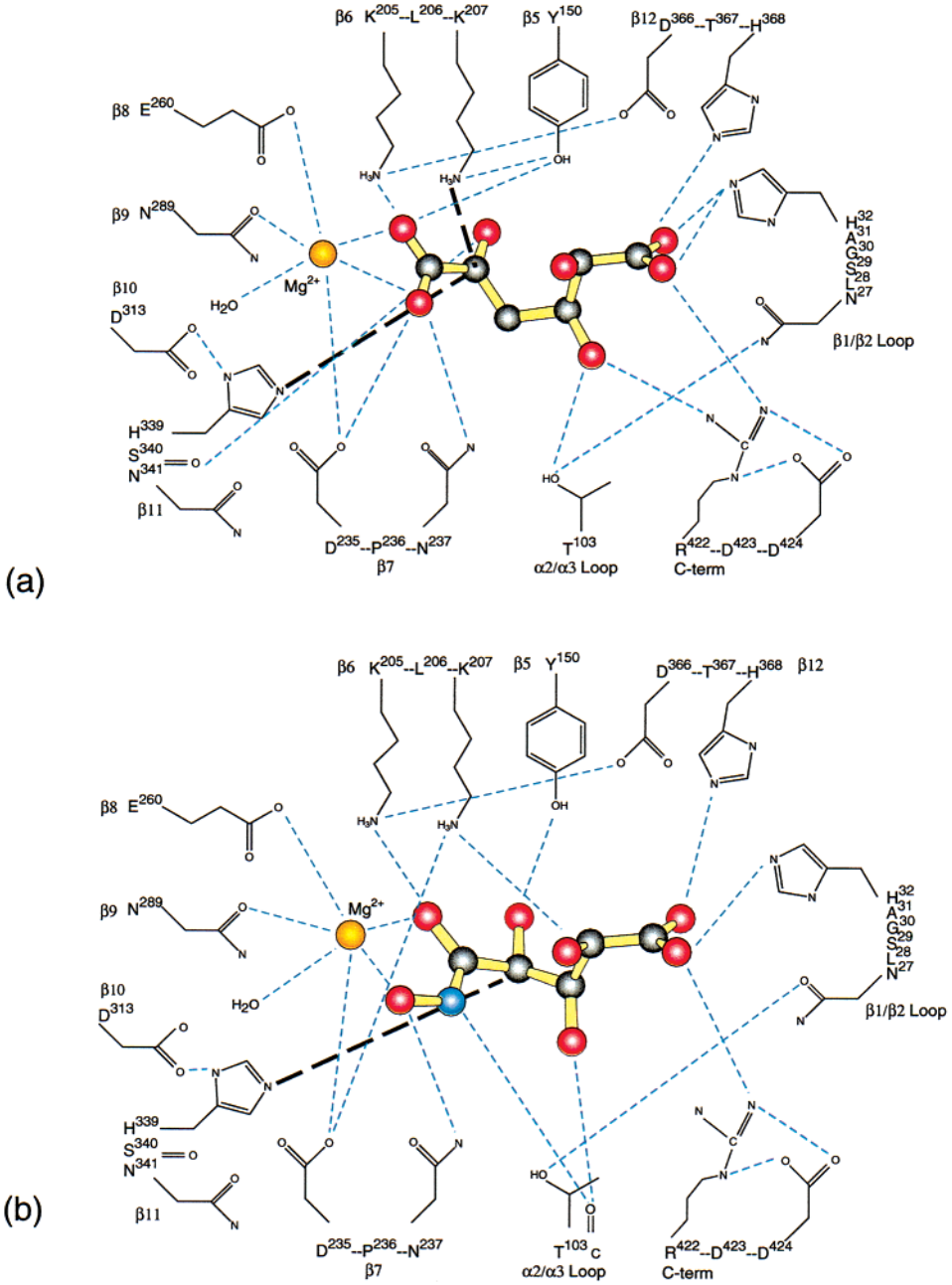


FIGURE 3: Interactions at the active site. (a) A schematic representation of the active site is shown for the complex of GlucD with KDG. Hydrogen bonding interactions that are less than 3.0 Å in length are shown with dashed blue lines. The thick black lines represent the positioning of the active site catalytic bases on opposite sides of C5 of the product. (b) A schematic representation of the active site in the presence of xylarohydroxamate. Similar interactions at C1−C3 are shown for the product and the inhibitor, while the network of interactions at C4 and C5 differs slightly. The movement of the C4 is apparent.

and sequence alignments of various enzymes of the enolase superfamily; however, the orientation of the product molecule in the active site was not known, i.e., its coordination geometry with the required Mg^{2+} and disposition relative to the KXK motif and the His-Asp dyad.

KDG is in an elongated conformation where the C6 carboxylate is a bidentate ligand of the metal ion. The distances between the carboxylate oxygens and the metal ion are 2.0 and 2.2 Å; the angle between the bonds from the oxygens to the metal ion measures 62°. This places C5 precisely between the functional groups of Lys^{207} and His^{339}, two catalytic bases that are properly positioned to abstract the proton from C5 of L-idarate and D-glucarate, respectively. Additionally, the plane formed by the sp^2 configuration of C5 is perpendicular to the interactions with the catalytic groups of Lys^{207} and His^{339}. Neither the bidentate coordination nor the orientation of C5 relative to the putative catalytic bases could have been predicted on the basis of its analogy with the structurally characterized MR.

In MR, one carboxylate oxygen and the α-OH group of the substrate are coordinated to the essential Mg^{2+}; the second carboxylate oxygen is hydrogen-bonded to Glu^{317} which has no functional homologue in the active site of GlucD. Presumably, this coordination geometry allows both neutralization of the negative charge of the substrate and stabilization of the enolate anion intermediate. Furthermore, in the MR-catalyzed reaction Lys^{166}, the homologue of Lys^{207} in GlucD, is the *S*-specific base and His^{297}, the homologue

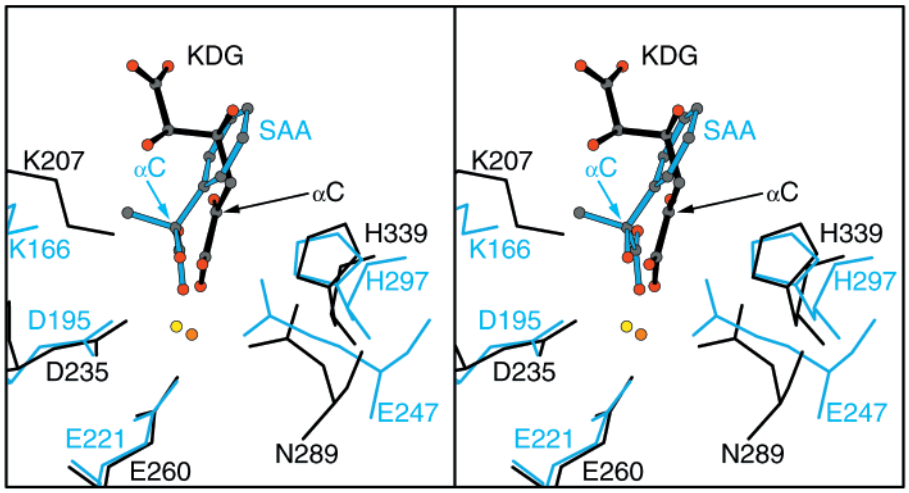


FIGURE 4: Comparison of the structures of the active sites of mandelate racemase (PDB file 1MDR) and GlucD. The active site of GlucD bound to KDG is shown with black bonds. The three residues that coordinate the metal ion (shown in orange for GlucD) as well as the two catalytic bases are shown with black bonds. The product molecule is shown as a ball-and-stick representation with black bonds. Atoms are colored gray (carbon) and red (oxygen). The orientation of *S*-atrolactate is shown with blue bonds. The equivalent active site residues are also included in blue. The metal ion for MR is shown in yellow. The orientation of Lys^{166} of MR is altered by the presence of the methyl group on the α-carbon of the inhibitor. The inversion of the α-carbons is illustrated which results in the reversal of chiral specificity of the enzymes' catalytic bases.

of His^{339} in GlucD, is the *R*-specific base (*6*). The configuration of C5 in L-idarate is *R*, and that of D-glucarate is *S*, thereby revealing that the stereochemical roles of the catalytic bases in the active sites of GlucD and MR are reversed (Figure 4). This reversal is a result of the monodentate nature of the binding of mandelate compared to the bidentate nature of glucarate binding. Because of this difference, the carboxylate group of *S*-atrolactate is projecting away from the viewer in Figure 4 while the carboxylate of glucarate is projected toward the viewer. This has the effect of inverting the chirality of this carbon relative to the protein. This fact could not have been predicted on the basis of sequence homology.

The active site of GlucD does contain an acidic residue, Asp^{366}, that is located at the C-terminal end of strand β12, the eighth strand in the barrel domain; in MR, this position is occupied by Glu^{317}. However, the carboxylate group of Asp^{366} does not interact directly with the C6 carboxylate group of KDG; a hydrogen-bonded water molecule intervenes. This suggests that Asp^{366} does not provide significant stabilization of the enolic intermediate.

Tyr^{150}, located at the C-terminal end of β5, the first strand of the barrel, forms a hydrogen bond to the amino side chain of Lys^{207} of the KXK motif. Tyr^{150} is also within hydrogen bonding distance of the C5 carbonyl of the product. The hydrogen bonding network with Lys^{207} suggests that Tyr^{150} contributes its hydroxyl proton to a hydrogen bond to the product. In MR, the C-terminal end of the first strand of the barrel domain does not contain a residue that contacts the bound substrate (recall that the α-OH group of mandelate is a ligand for the essential Mg^{2+}).

This structure identifies two residues that may act as general acid catalysts in the vinylogous β-elimination of the 4-OH group and subsequent stereospecific tautomerization to yield the KDG product. His^{339} could serve both of these functions in addition to abstracting the α-proton from D-glucarate, illustrating an extreme mechanistic economy of functional groups. Alternatively, the carboxamide group of Asn^{341} is hydrogen bonded to the 3-OH group of KDG; modeling suggests that it also is positioned within hydrogen bonding distance of the 4-OH group of D-glucarate and L-idarate. With this interaction, the carboxamide group of Asn^{341} would be appropriately positioned to function as the general acid in the departure of the 4-OH group; this suggestion is contrary to conventional thinking about the identities of viable acid/base catalysts in active sites. However, Asn-linked *N*-glycosyltransferases must be able to make the carboxamide group nucleophilic, although an unequivocal structural explanation for the nucleophilicity of Asn in these reactions is not available (*39*).

The C-terminal loop that was disordered in the model of the GlucD from *P. putida* completes the active site. A seven-residue loop follows α12 and contains Arg^{422}, the side chain of which interacts with both the C3 hydroxyl and the C1 carboxylate. A positively charged pocket is formed around the C1 carboxylate by the side chains of His^{32}, His^{368}, and Arg^{422}. Other interactions between the protein and product include a hydrogen bond between the carboxamide of Asn^{237} and one oxygen of the C6 carboxylate and also a hydrogen bond between the amino group of Lys^{205} and the second C6 carboxylate oxygen. In two of the subunits, the C5 carbonyl interacts with the main chain carbonyl oxygen of Ser^{340}. This suggests that the active site is not completely occupied by product and that some substrate is present with its C5 hydroxyl hydrogen bonded to Ser^{340}. Nonetheless, the omit-map density is well-defined for the inclusion of a KDG molecule in these complexes (Figure 2B) with no apparent density for the C4 hydroxyl of a substrate molecule.

The active site is entirely protected from solvent by four loops, the loop and two short antiparallel strands connecting residues Leu^{25} and His^{32}, the flexible loop between Ala^{92} and Arg^{107}, and the two C-terminal loops which have been identified in this work as residues His^{417}–Asp^{424} and Leu^{431}–Arg^{446}. Access to the active site is most likely provided by movement of the loop between Ala^{92} and Arg^{107} because this segment is disordered in the unliganded enzyme; however, since the entire C-terminal sequence beyond helix α12 was disordered in the structure of the GlucD from *P. putida* (*14*),

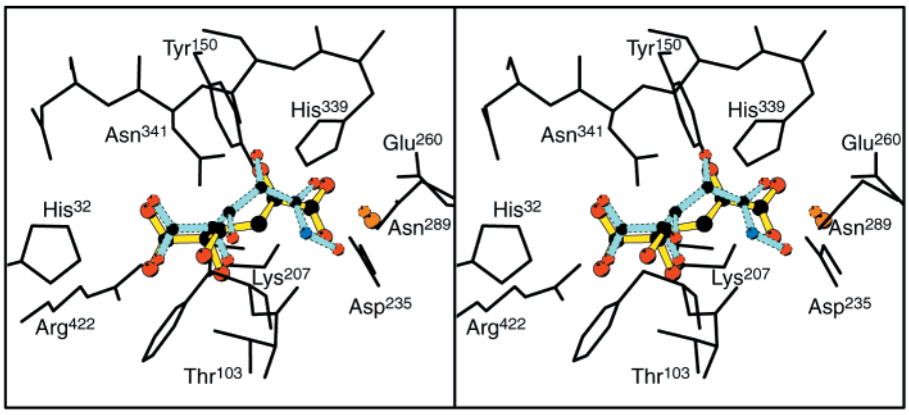


FIGURE 5: Overlay of active sites of the GlucD complexes with xylarohydroxamate and KDG. The protein atoms exhibit very little movement, and as such, only the protein atoms from the product complex are included. The changes in the orientation of the C4 groups are apparent which explains why the C4 hydroxyl of the hydroxamate approximates the location of the C5 carbonyl in the product complex. Several important active site residues are labeled.

it is probable that this region may contribute to conformational mobility.

Structure of the GlucD·4-Deoxy-D-glucarate Complex. The competitive inhibitor 4-deoxy-D-glucarate (K_i = 1 mM; K_m for D-glucarate = 0.16 mM) differs from KDG in the sp^3 configuration at C5 of D-glucarate; the product has a C5 carbonyl group, but the inhibitor contains a C5 hydroxyl, analogous to the substrate molecule. As would be expected, the interactions made between the active site and the inhibitor are virtually identical to those seen in the KDG complex (Figure 2C). In all four subunits, a hydrogen bond is formed between the C5 hydroxyl and the backbone carbonyl of Ser^{340}. No significant movements of protein atoms are observed.

Structure of the GlucD·Xylarohydroxamate Complex. Prior to the determination of the structures of the complexes with KDG and 4-deoxy-D-glucarate, racemic xylarohydroxamate was synthesized as a potential mimic of the enolic intermediate that would be obtained by abstraction of the α-proton from either D-glucarate or L-idarate; this strategy was based on the coordination of the enantiomers of mandelate to the Mg^{2+} in the active site of MR. However, given the bidentate coordination of the C6 carboxylate groups of KDG and 4-deoxy-D-glucarate to the essential Mg^{2+}, we subsequently realized that the xylarohydroxamate should be considered an analogue of a five-carbon diacid sugar. Xylarohydroxamate is a competitive inhibitor of GlucD (K_i = 0.8 mM), where its affinity is considerably lower than that expected if it were an analogue of the enolic intermediate. Regardless, the presence of the C4 hydroxyl in the xylarohydroxamate suggested that the structure of a complex with this inhibitor could assist in the identification of the expected general acidic catalyst that would facilitate the departure of the leaving group.

The xylarohydroxamate is well-ordered in the active site (Figure 2D). The inhibitor is in a position similar to that of the product at the distal C1−C3 positions (Figure 4); however, relative to the product, the remaining atoms of the inhibitor shift toward the Mg^{2+}, reflecting the shorter overall length of the inhibitor. The inhibitor and substrate diverge after C3 such that C4 of the product and inhibitor differ in position by ~1 Å. This movement, coupled with a rotation about the C3−C4 bond, places the C4 hydroxyl of the hydroxamate in a position close to that occupied by the C5 carbonyl of the product. This allows both the C5 carbonyl oxygen and the hydroxyamino oxygen of the hydroxamate to participate in bidentate coordination of the metal (Figure 3b). Thus, as shown in Figure 5, the structure of the xylarohydroxamate complex demonstrates that, with only small changes, C1−C3 of the inhibitor behave like C1 and C3 of the product, C4 of the inhibitor mimics C5 of the product, and the hydroxamate at C5 mimics the product carboxylate by forming a bidentate interaction with the Mg^{2+} ion. This change allows the C4 hydroxyl to interact with the main chain carbonyl oxygen of Ser^{340} and places it within hydrogen bonding distance of Tyr^{150}. This demonstrates again that Tyr^{150} may play a special role in coordinating the product.

The coordination of xylarohydroxamate by the metal ion in the active site of GlucD is similar to that observed in the structure of enolase complexed with phosphonoacetohydroxamate (*4*). In the case of GlucD, as was observed with enolase, the structure of the hydroxamate complex is misleading because the C5 carbonyl is involved in the metal coordination. As a consequence, the C4 hydroxyl on xylarohydroxamate is unlikely to be in the same position or orientation as the C4 hydroxyl of glucarate. Thus, we do not believe that this analogue is useful for identifying the putative general acidic catalyst in the dehydration reaction. However, this emphasizes the fact that caution must be adopted in using hydroxamates to study this superfamily of enzymes.

Mutagenic Analysis of the Mechanism of the GlucD-Catalyzed Reaction. As noted in an earlier section, the bidentate coordination of the essential Mg^{2+} by the C6 carboxylate was unexpected; in MR, one carboxylate oxygen and the α-OH group of the substrate are coordinated to the essential Mg^{2+}. A consequence of this coordination geometry is that the 5-proton of D-glucarate is presented to His^{339} and that of L-idarate is presented to Lys^{207}, the opposite predictions made on the basis of the structure−function relationships established for MR. Several site-directed substitutions for Lys^{207} and His^{339} were constructed so that it could be determined whether their kinetic and mechanistic phenotypes would support the active site geometry observed in the complexes with KDG and 4-deoxy-D-glucarate (Table 4). The structures allow the hypothesis that the dehydration of D-glucarate could be unaffected by substitutions for Lys^{207},

Table 4: Kinetic Phenotypes of GlucD and Mutants

enzyme	substrate	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat}^H/k_{cat}^D [a]
wild type	D-glucarate	35	5.8×10^5	3.5
	L-idarate	34	2.1×10^5	1.8
H339A	D-glucarate	0.016	0.014	0.013
	L-idarate	5.1×10^{-3}	75	2.3
H339N	D-glucarate	3.3×10^{-4}	7.2	
	L-idarate	2.5×10^{-4}	2.9	
H339Q	D-glucarate	2.0×10^{-2}	4.2×10^2	3.8
	L-idarate	2.0×10^{-2}	47	2.3
K207Q	D-glucarate	5.0×10^{-3}	110	3.7
	L-idarate	1.5×10^{-3}	17	1.9
K207R	D-glucarate	1.0×10^{-3}	10	0.9
	L-idarate	1.5×10^{-3}	6.5	2.0
Y150F	D-glucarate	1.9×10^{-1}	2.2×10^3	
	L-idarate	8.0×10^{-2}	3.8×10^2	
D366A	D-glucarate	1.0	5.1×10^2	
	L-idarate	1.5	3.2×10^2	
D366N	D-glucarate	1.8	3.1×10^3	
	L-idarate	8.0	9×10^2	

[a] Ratio of the substrate isotope effect on k_{cat}.

because the 5-proton is presented to His^{339} and the 4-OH group that is eliminated can be predicted to be located on the same side of the active site; i.e., the β-elimination reaction proceeds with a syn stereochemical course. In the MR-catalyzed reaction, it was established that the proton abstraction reactions catalyzed by Lys^{166} and His^{297} were independent reactions; e.g., the H297N mutant was inactive as a racemase but fully functional in catalyzing the exchange of the α-proton of (*S*)- but not (*R*)-mandelate with solvent hydrogen (*40*, *41*).

In contrast to these predictions, the substitutions for Lys^{207} and His^{339} were significantly and comparably inactive using either D-glucarate or L-idarate as the substrate in both dehydration (Table 4) and epimerization reactions; no exchange of the α-proton of either diacid substrate with solvent hydrogen was observed. Substrate deuterium isotope effects were measured, but these also failed to provide a basis for distinguishing between the presumed roles of Lys^{207} and His^{339} in abstracting the α-protons from L-idarate and D-glucarate, respectively. Thus, the catalytic roles of Lys^{207} and His^{339} could not be functionally dissected using substitutions as was achieved for the MR-catalyzed reaction (*40*, *41*). Presumably, the substitutions for Lys^{207} and His^{339} perturbed the structure of the active site so that the α-proton abstraction/reprotonation of D-glucarate and L-idarate could not be dissected from vinylogous β-elimination. Alternatively, the multistep nature of the reaction catalyzed by GlucD, α-proton abstraction, vinylogous β-elimination, and protonation/tautomerization, may obscure the functions of Lys^{207} and His^{339} in the initial partial reaction. It seems clear that both Lys^{207} and His^{339} are necessary for formation of the enolic intermediate irrespective of whether the substrate is D-glucarate or L-idarate. Attempts are underway to determine the structures of these mutants to determine if the substitutions dramatically alter the structure.

Mutations were also generated and characterized for two other residues that the structures revealed to be located in the active site of GlucD.

(1) Y150F. This substitution significantly impaired dehydration of both D-glucarate and L-idarate. This suggests that Y150 is involved in the dehydration reactions, perhaps by binding the 5-OH groups of the substrates as suggested by the structural data.

(2) D366A and D366N. Although these substitutions reduced the values of k_{cat} by factors ranging from 4- to 30-fold, these changes are consistent with an indirect effect of Asp^{366} on the stabilization of the enolic intermediate. Recall that in the active site of MR, Glu^{317}, which also is located at the C-terminal strand of the eighth β-strand in the barrel domain, participates in catalysis by stabilizing the enolic intermediate via direct hydrogen bonding.

CONCLUSIONS

Our structural and mutagenesis results provide unexpected insights into the catalytic mechanism of GlucD that are distinct from those previously established for MR. Unlike MR, the C6 carboxylate groups of both KDG and 4-deoxy-D-glucarate are bidentate ligands of the essential divalent metal ion. While Lys^{207} and His^{339} are positioned ideally to transfer protons to and from C5 of both D-glucarate and L-idarate in the dehydration and epimerization reactions catalyzed by GlucD, the complexes with KDG and 4-deoxy-D-glucarate suggest that Lys^{207} is positioned to abstract the 5-proton from L-idarate and His^{339} is positioned to abstract the 5-proton from D-glucarate. The enolate anion intermediate is stabilized by bidentate coordination with the metal ion and hydrogen bonds to the side chains of Lys^{205} and Asn^{237}. The identities of the general acid catalysts that promote the departure of the 4-OH groups in the dehydration reactions and the subsequent stereospecific tautomerization to the KDG product are restricted to His^{339}, the putative base for 5-proton abstraction from D-glucarate, and Asn^{341}, a priori an unlikely possibility given the ionization properties of the carboxamide functional group. These studies confirm the power of structure-based predictions within an enzyme superfamily, but also illustrate the limitations imposed by the subtle but important differences in the ways that substrates are bound in an active site.

ACKNOWLEDGMENT

We thank Drs. Frank Rotella, Norma Duke, and Randy Alkire at the Structural Biology Center at APS for assistance in data collection and processing. We are also grateful to Dr. Cary Bauer and Thomas Thompson who collected the data sets of the two inhibitor complexes.

REFERENCES

1. 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.
2. Babbitt, P. C., and Gerlt, J. A. (1997) *J. Biol. Chem. 272*, 30591−30594.
3. Lebioda, L., and Stec, B. (1991) *Biochemistry 30*, 2817−2822.
4. Wedekind, J. E., Poyner, R. R., Reed, G. H., and Rayment, I. (1994) *Biochemistry 33*, 9333−9342.
5. Larsen, T. M., Wedekind, J. E., Rayment, I., and Reed, G. H. (1996) *Biochemistry 35*, 4349−4358.
6. 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*, 635−643.
7. Helin, S., Kahn, P. C., Guha, B. L., Mallows, D. G., and Goldman, A. (1995) *J. Mol. Biol. 254*, 918−941.
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. Palmer, D. R. J., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., and Gerlt, J. A. (1999) *Biochemistry 38*, 4252−4258.
10. Hubbard, B. K. (1999) Ph.D. Thesis, University of Illinois, Urbana, IL.
11. Palmer, D. R. J., Wieczorek, S. J., Hubbard, B. K., Mrachko, G. T., and Gerlt, J. A. (1997) *J. Am. Chem. Soc. 119*, 9580−9581.
12. Palmer, D. R. J., Hubbard, B. K., and Gerlt, J. A. (1998) *Biochemistry 37*, 14350−14357.
13. Palmer, D. R., and Gerlt, J. A. (1996) *J. Am. Chem. Soc. 118*, 10323−10324.
14. Gulick, A. M., Palmer, D. R. J., Babbitt, P. C., Gerlt, J. A., and Rayment, I. (1998) *Biochemistry 37*, 14358−14368.
15. Mowshowitz, S., Avigad, G., and Englard, S. (1974) *J. Bacteriol. 118*, 1051−1058.
16. Bien, G. (1958) *J. Chem. Soc.*, 3189−3190.
17. Hubbard, B. K., Koch, M., Palmer, D. R., Babbitt, P. C., and Gerlt, J. A. (1998) *Biochemistry 37*, 14369−14375.
18. Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
19. Navaza, J. (1993) *Acta Crystallogr. D49*, 588−591.
20. Kabsch, W. (1988) *J. Appl. Crystallogr. 21*, 916−924.
21. Kabsch, W. (1988) *J. Appl. Crystallogr. 21*, 67−71.
22. Wesenberg, G., and Rayment, I. (1997) University of Wisconsin, Madison, WI.
23. Westbrook, E. M., and Naday, I. (1997) *Methods Enzymol. 276*, 244−268.
24. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
25. Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987) *Acta Crystallogr. A43*, 489−501.
26. Tronrud, D. E. (1997) *Methods Enzymol 277*, 306−319.
27. Jones, T. A. (1985) *Methods Enzymol. 115*, 157−171.
28. Rossmann, M. G. (1972) *The Molecular Replacement Method*, Gordon & Breach, New York.
29. Roussel, A., and Cabillau, C. (1991) in *Silicon Graphics Geometry Partners Directory 86*, Silicon Graphics, Mountain View, CA.
30. Deleted in proof.
31. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
32. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
33. Esnouf, R. M. (1999) *Acta Crystallogr. D55*, 938−940.
34. Kabsch, W. (1976) *Acta Crystallogr. A32*, 922−923.
35. CCP4 (1994) *Acta Crystallogr. D50*, 760−763.
36. Mozzarelli, A., and Rossi, G. L. (1996) *Annu. Rev. Biophys. Biomol. Struct. 25*, 343−365.
37. Gulick, A. M., Bauer, C. B., Thoden, J. B., and Rayment, I. (1997) *Biochemistry 36*, 14358−14368.
38. Malashkevich, V. N., Toney, M. D., and Jansonius, J. N. (1993) *Biochemistry 32*, 13451−13462.
39. Imperiali, B. (1997) *Acc. Chem. Res. 30*, 452−459.
40. Landro, J. A., Kallarakal, A. T., Ransom, S. C., Gerlt, J. A., Kozarich, J. W., Neidhart, D. J., and Kenyon, G. L. (1991) *Biochemistry 30*, 9274−9281.
41. Kallarakal, A. T., Mitra, B., Kazarich, J. W., Gerlt, J. A., Clifton, J. G., Petsko, G. A., and Kenyon, G. L. (1995) *Biochemistry 34*, 2788−2797.
42. Cohen, G. H. (1997) *J. Appl. Crystallogr. 30*, 1160−1161.

BI992782I