Evolution of Enzymatic Activities in the Enolase Superfamily: Crystal Structure of (*D*)-Glucarate Dehydratase from *Pseudomonas putida*[†,‡]

Andrew M. Gulick,[§] David R. J. Palmer,[||] Patricia C. Babbitt,[⊥] John A. Gerlt,*[,||] and Ivan Rayment*[,§]

The Institute for Enzyme Research and Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, Department of Biochemistry, University of Illinois, Urbana, Illinois 61801, and Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446

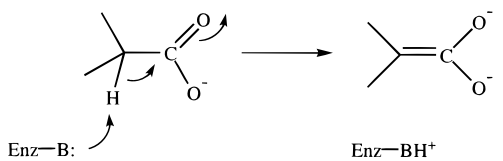
Received May 14, 1998; Revised Manuscript Received August 14, 1998

ABSTRACT: The structure of (*D*)-glucarate dehydratase from *Pseudomonas putida* (GlucD) has been solved at 2.3 Å resolution by multiple isomorphous replacement and refined to a final *R*-factor of 19.0%. The protein crystallizes in the space group *I*222 with one subunit in the asymmetric unit. The unit cell dimensions are $a = 69.6$ Å, $b = 108.8$ Å, and $c = 122.6$ Å. The crystals were grown using the batch method where the primary precipitant was poly(ethylene glycol) 1000. The structure reveals that GlucD is a tetramer of four identical polypeptides, each containing 451 residues. The structure was determined without a bound substrate or substrate analogue. Three disordered regions are noted: the N-terminus through residue 11, a loop containing residues 99 through 110, and the C-terminus from residue 423. On the basis of primary sequence alignments, we previously concluded that GlucD is a member of the mandelate racemase (MR) subfamily of the enolase superfamily [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]. This prediction is now verified, since the overall fold of GlucD is strikingly similar to those of MR, muconate lactonizing enzyme I, and enolase. Also, many of the active site residues of GlucD can be superimposed on those found in the active site of MR. The implications of this structure on the evolution of catalysis in the enolase superfamily are discussed.

The limited number of unique protein folds allows the conclusion that new enzymes likely evolve from preexisting (older) enzymes via gene duplication followed by "retooling" of the active site to catalyze the new reaction. However, our understanding of this retooling process has been controversial. One hypothesis is that both the old and new enzymes catalyze the same overall reaction but with a different substrate, so the retooling primarily involves only changes in substrate specificity (*1*, *2*). An example of this strategy is provided by the serine proteases having the trypsin fold (e.g., trypsin, chymotrypsin, elastase, and their structural homologues), each member of which catalyzes peptide bond hydrolysis but recognizes a different primary sequence context. In this view, conscription of enzymes for new functions is dominated by substrate binding, with constraints on preserving the ability to deliver the correlated chemistry something of an evolutionary "afterthought".

An alternate hypothesis is that the old enzyme provides a structural strategy for catalyzing a difficult chemical transformation, and the retooling involves the addition or deletion of functional groups necessary to utilize that strategy to catalyze a different overall reaction (*3*, *4*). This strategy is more easily distinguished in superfamilies of highly divergent proteins, e.g., those in which both the substrates and overall chemical reactions may be substantially different. This latter strategy apparently is employed in the evolution of a number of enzyme superfamilies, including the enolase superfamily (*3*). From this perspective, studies of enzymes that perform a specific chemical step allow identification of both (1) the conserved structural features of the active site that mediate the common partial reaction and (2) the dominant chemical constraint conserved in the evolutionary retooling to generate the new catalytic activity.

To date, the enolase superfamily includes nine different enzymes of assigned function (*3*, *4*), each of which initiates its reaction by divalent metal-assisted, general base-catalyzed abstraction of the α-proton of a carboxylate anion to generate a stabilized enolate anion intermediate:



The fate of the intermediate is determined by the structure of each active site, so that the overall reactions differ and may involve 1,1-proton transfer (racemization), β-elimination of either water or ammonia, or cycloisomerization. Although the superfamily also includes at least eight additional open reading frames (orfs)[1] of uncertain or unassigned function,

[†] 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.

[‡] The X-ray coordinates have been deposited in the Brookhaven Protein Data Bank (file name 1BQG).

* Address correspondence to these authors (e-mail: j-gerlt@uiuc.edu and ivan@enzyme.wisc.edu).

[§] University of Wisconsin.

[||] University of Illinois.

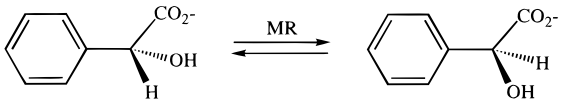
[⊥] University of California.

S0006-2960(98)01123-4 CCC: $15.00
Published on Web 09/25/1998

our working hypothesis is that the reaction catalyzed by each of these involves the formation of a stabilized enolate anion by divalent metal-assisted, general base-catalyzed abstraction of the α-proton of a carboxylate anion substrate.

Although an insufficient number of primary sequences are now available to allow construction of detailed phylogenetic relationships in the enolase superfamily, the alignment of the primary sequences of the members of the superfamily allows three distinct subgroups to be identified on the basis of conserved sequence motifs as well as catalytic residues inferred to be in the active sites of the various proteins. Coincidentally, a high-resolution X-ray structure is available for one member of each of these subgroups, i.e., mandelate racemase from *Pseudomonas putida* (MR) (*5*, *6*), muconate lactonizing enzyme I from *P. putida* (MLE I) (*7*; M. Hasson, I. Schlicting, J. Moulai, D. Ringe, and G. Petsko, unpublished observations), and enolase from *Saccharomyces cerevisiae* (*8*, *9*). As such, the subgroups are designated by the names of these enzymes.

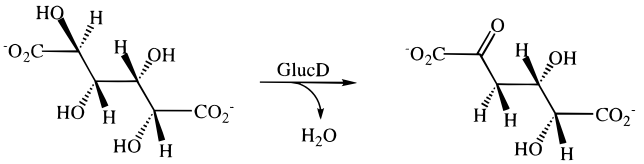
The reaction catalyzed by MR is a 1,1-proton transfer reaction that equilibrates the enantiomers of mandelate.



(*R*)-Mandelate (*S*)-Mandelate

Mechanistic analyses of the MR-catalyzed reaction (*10*, *11*) revealed that the reaction is stepwise with initial general base-catalyzed abstraction of the α-proton from one enantiomer of mandelate to generate a stabilized enolate anion intermediate (Scheme 1). Protonation of the intermediate on the opposite face by the conjugate acid of a second general basic catalyst completes the 1,1-proton transfer reaction. The results of structural and mechanistic analyses are consistent with Lys 166 and His 297 functioning as the (*S*)- and (*R*)-specific general basic catalysts, respectively (*5*, *6*, *10*, *12*). The enolate anion intermediate is stabilized by hydrogen-bonding interactions with Glu 317 and electrostatic interactions with Lys 164 and a single essential Mg^{2+}. The Mg^{2+} is coordinated to the carboxylate groups of Asp 195, Glu 221, and Glu 247.

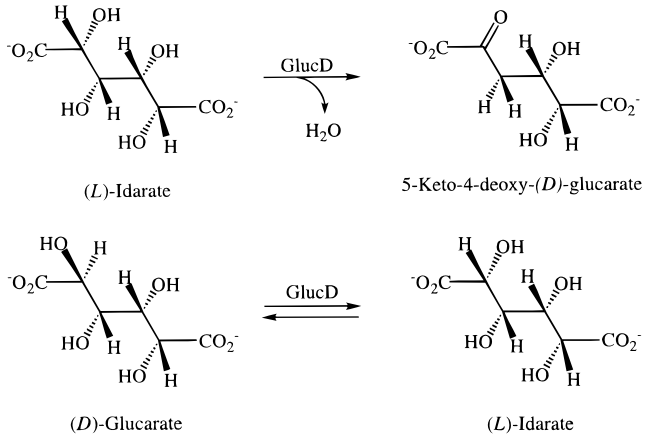
The MR subgroup includes homologous glucarate dehydratases (GlucDs) from *P. putida*, *Escherichia coli*, and *Bacillus subtilis*. GlucD catalyzes the *β*-elimination of water from (*D*)-glucarate to yield 5-keto-4-deoxy-(*D*)-glucarate (*13*, *14*).



(*D*)-Glucarate 5-Keto-4-deoxy-(*D*)-glucarate

Examination of the alignment of the primary sequences of MR and the homologous GlucDs (panel A in Figure 1) suggests the identities of several conserved functional groups that are likely to participate in the reactions catalyzed by GlucD (*3*). Referring to the GlucD from *P. putida*, these include (1) Lys 213 and His 345, homologues of Lys 166 and His 297 in MR; (2) Lys 211, the homologue of Lys 164; and (3) Asp 241 and Glu 266, homologues of Asp 195 and Glu 221. The sequence alignment does not allow identification of homologues for either Glu 247, the third ligand for Mg^{2+}, or Glu 317, the putative hydrogen-bonding electrophilic catalyst.

On the basis of the conserved general basic functional groups, namely, an analogue for both His 297 and Lys 166 of MR, we predicted and confirmed that the GlucD from *P. putida* would catalyze not only the dehydration of (*D*)-glucarate but also both the dehydration of (*L*)-idarate and the epimerization of (*D*)-glucarate and (*L*)-idarate (*13*, *14*).



(*L*)-Idarate 5-Keto-4-deoxy-(*D*)-glucarate

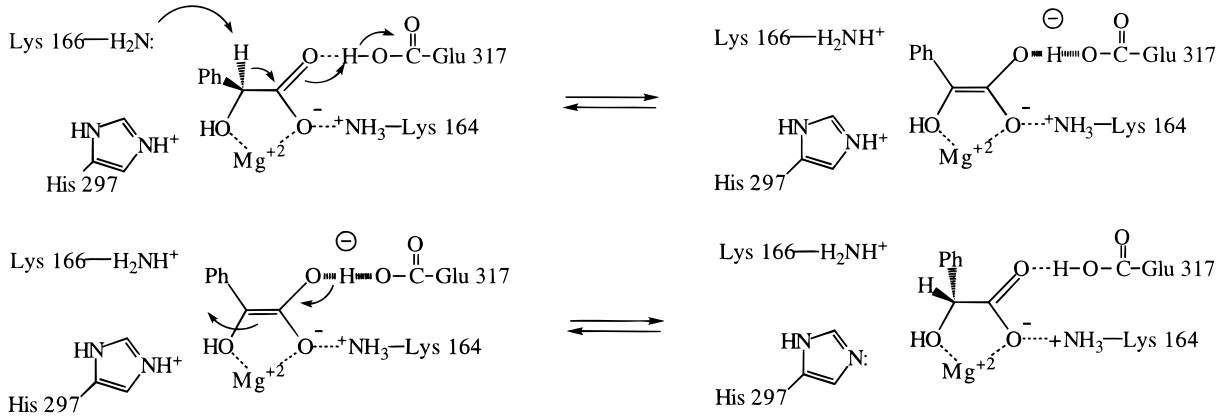
(*D*)-Glucarate (*L*)-Idarate

While the epimerization reaction catalyzed by GlucD provides a formal biosynthetic pathway for (*L*)-idarate, to the best of our knowledge (*L*)-idarate is not found in the biosphere. As such, the ability of GlucD to utilize this acid sugar as the carbon source is metabolically cryptic but undoubtedly important for the eventual understanding of the evolution of this superfamily of enzymes. This observation provides support for the hypothesis that GlucD evolved from a progenitor whose active site supported chemical reactions on enantiomers/diastereomers of its substrate.

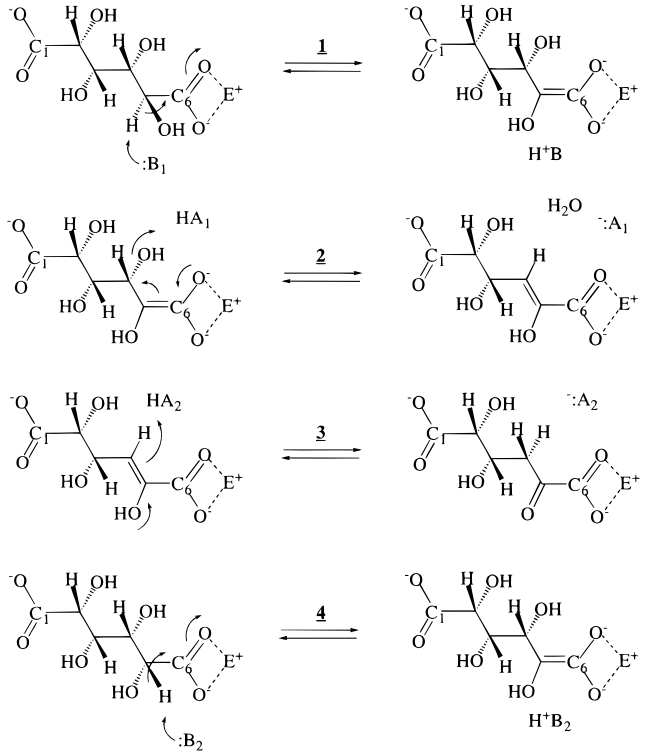
Each of the GlucD-catalyzed reactions is envisaged to involve the formation of a common, stabilized, enolic intermediate via general base-catalyzed abstraction of an α-proton from the substrate (Scheme 2; (*D*)-glucarate, reaction 1; (*L*)-idarate, reaction 4). In the case of the dehydration reactions, vinylogous elimination of the *β*-OH group from the enolate anion intermediate (Scheme 2, reaction 2) is expected to require general acid catalysis, chemistry not catalyzed by MR. Ketonization of the resulting enol intermediate (Scheme 2, reaction 3) is also expected to require general acid catalysis, chemistry also not catalyzed by MR. In support of this expectation, GlucD catalyzes the stereospecific incorporation of a solvent-derived hydrogen into the 4*S*-position of the product (*15*). In the case of the epimerization reaction, as in the MR-catalyzed reaction, the product is formed by protonation of the intermediate on the opposite face by the conjugate acid of the second general basic catalyst (Scheme 2; reactions 1 and 4). The previous paper describes a quantitative analysis of the partitioning of the common enolic intermediate between dehydration and

[1] Abbreviations: GalD, (*D*)-galactonate dehydratase; GlucD, (*D*)-glucarate dehydratase; MLE I, muconate lactonizing enzyme I; MR, mandelate racemase; orf, open reading frame; BTP, 1,3-bis[[tris(hydroxymethyl)methyl]amino]propane; TIM, triose phosphate isomerase.

Scheme 1



Scheme 2



epimerization (*14*). These observations lend strong support to the hypothesis that the dominant strategy for the evolution of new enzymes is retention of the structural strategy for catalyzing the chemically difficult step of α-proton abstraction and addition of functional groups to catalyze a different overall reaction.

In this paper we describe the three-dimensional structure of the GlucD from *P. putida*. This structure confirms the presence of the functional groups predicted by the sequence alignment with MR (*3*) and suggests the identities of functional groups that function as (1) the third ligand for the Mg^{2+} ion, (2) the hydrogen-bonding electrophilic catalyst that stabilizes the enolic intermediate, and (3) the general acid catalyst that facilitates the departure of the 4-hydroxyl group in the *β*-elimination reaction and, perhaps, stereospecific protonation of the enol intermediate.

MATERIALS AND METHODS

Protein Crystallization. Homogeneous GlucD from *P. putida* (*4*, *13*) was dialyzed into 10 mM Hepps, pH 8.0 at 4 °C, 25 mM NaCl, 1 mM $MgCl_2$, and 1 mM NaN_3, concentrated to 10.4 mg/mL, and frozen as small aliquots by dropping the solution directly into liquid nitrogen. Individual pellets of frozen protein were thawed daily for crystallization experiments.

Crystals of GlucD were grown by microbatch methods in a nine-well Plexiglas plate at room temperature after equal volumes of protein and precipitant solution were mixed, typically 6 μL of each. The precipitant contained 19% PEG 1000, 100 mM sodium formate, and 100 mM BTP, pH 7.0 at 22 °C. Crystals were grown by immediately streak-seeding from a crushed crystal with a cat whisker. Crystals appeared within 3 days and grew to full size within 2 weeks. The crystals grew to a maximum size of 0.8 × 0.3 × 0.3 mm and diffracted to 2.3 Å. The GlucD crystallized in the orthorhombic space group *I*222 as determined by precession camera methods with unit cell dimensions $a = 69.6$ Å, $b = 108.8$ Å, $c = 122.6$ Å, and $\alpha = \beta = \gamma = 90°$. GlucD is a tetramer that exhibits 222 symmetry with one subunit in the asymmetric unit. There are, therefore, two complete GlucD tetramers in the unit cell, and the Matthews coefficient is 2.33 $Å^3$/Da (solvent = 47%) (*16*).

To obtain reproducible X-ray diffraction results, the crystals first were transferred from the crystallization mother liquor to a solution with a higher PEG concentration, thereby stabilizing the crystals for further manipulation. In this procedure, a solution containing 20% PEG 1000, 0.1 M sodium formate, and 100 mM BTP, pH 7.0 at 22 °C, was added slowly over a 24 h period to the wells in which the crystals had grown. A total of 25 μL of stabilization solution was added to the original 12 μL at a rate of 1−2 μL/h. After stabilization, the crystals could be maintained at room temperature for several weeks and could be transferred to the same solution containing heavy atom compounds for trial soaking experiments. The crystals were mounted in thin-walled quartz capillary tubes and cooled to 0 °C for data collection.

Data Collection and Model Building. Data were collected with Cu Kα radiation from a Rigaku RU200 X-ray generator operating at 50 kV × 90 mA with Supper long focusing mirrors and a Siemens Hi-Star area detector. Data were processed with XDS (*17*, *18*) and were scaled with XSCALIBRE (*19*).

The protein structure was determined by multiple isomorphous replacement (MIR) with two heavy atom derivatives. The heavy atom compounds used were trimethyllead acetate

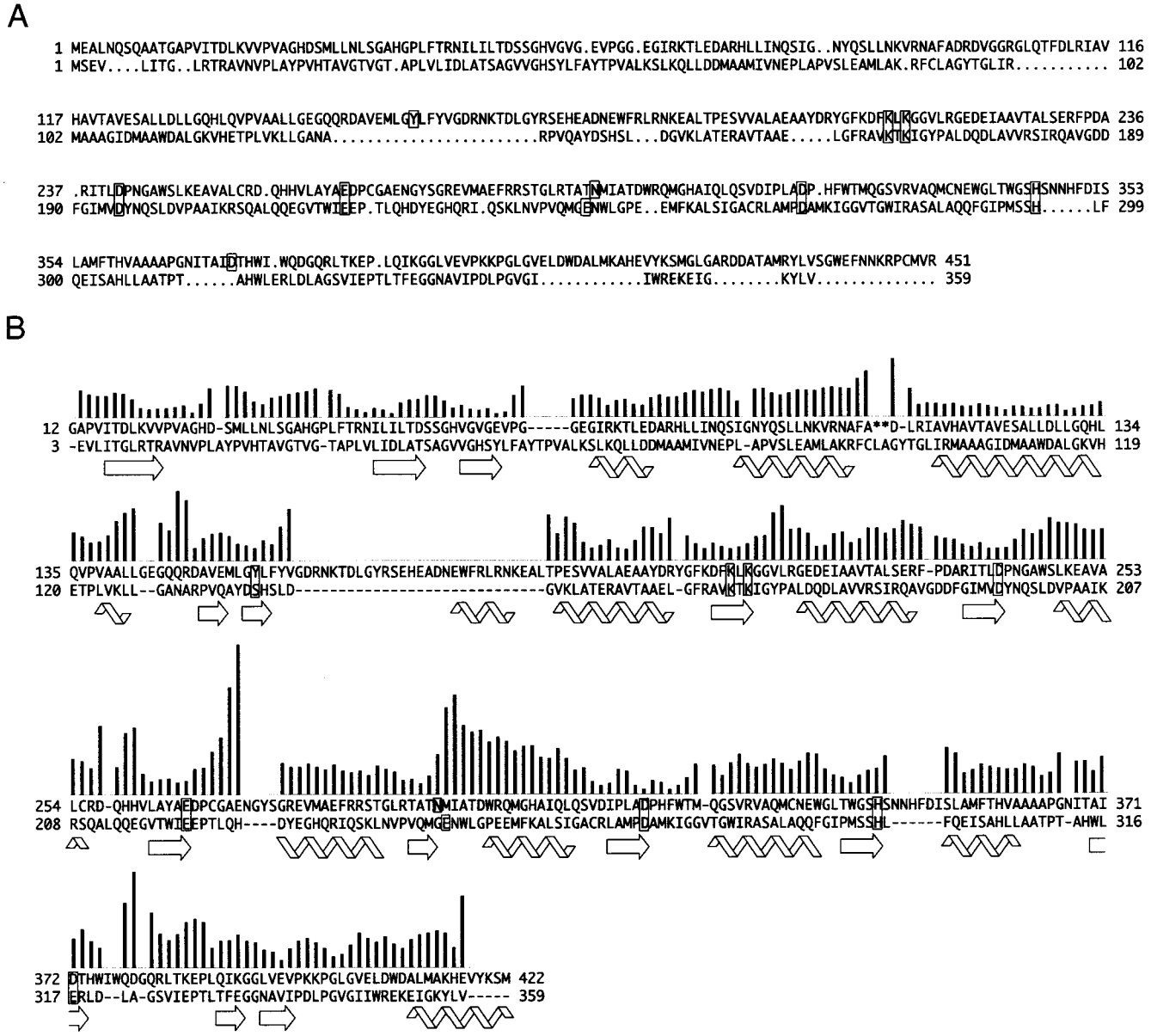


FIGURE 1: Amino acid alignments of MR and GlucD. Panel A: Genetics-based alignment. Panel B: Structure-based alignment. In both panels, the active site residues are enclosed in boxes; in panel B, the secondary structure elements in GlucD are shown below the alignment. The coordinates of MR with sulfate bound to the active site Mg^{2+} (PDB accession code 2MNR; ref *33*) were used in the structure-based alignment. The distance between the Cα atoms of each pair is represented by the bars above the alignment. The rms deviation for the two structures calculated by ALIGN (*24*) is 2.3 Å. The bar above Gly 244 in GlucD and Gln 198 in MR is also equal to 2.3 Å. Two proteins having the level of sequence identity relating GlucD and MR (23%) would be expected to have an rms deviation of 1.7 Å (*35*). The slightly higher value resulting from our calculations arises from the inclusion of all Cα's rather than simply using those of the core of the protein. The two asterisks in the first line of the GlucD sequence represent the disordered break between residues Ala 98 and Asp 111.

and K_2PtCl_4 (Table 1). Two sites from the platinum derivative were identified by visual analysis of the Patterson functions. The positions of the platinum atoms were refined with HEAVY (*20*). A difference map generated with structure factor amplitudes of the form $|F_{der} - F_{nat}|$ and native protein phases obtained from the platinum derivative identified the location of a major site for the lead derivative. Refinement of the positions and occupancies of these three heavy atom sites with HEAVY followed by difference Fourier analysis identified an additional minor site for each metal. The anomalous signal collected for both derivatives was included in the final refinement. The native phases were improved by solvent flattening (*21*), resulting in the final native MIR map that was used to build the protein model. The final figure of merit for all data to 2.6 Å was 0.66. Model building was done with FRODO (*22*) on an Evans and Sutherland PS390.

Polyalanine was built into the native MIR map. The presence of nine tryptophan residues in the 451 amino acid structure facilitated the identification of the sequence along the polyalanine chain, allowing side chains to be identified and built into the protein model. Initially, 377 residues (83% of the total sequence) were built into the model; over 300 of these residues were built in with side chains while the remaining residues were built in as alanine residues where no obvious side chain density was present. The temperature factors were set to 25 $Å^2$, and the model was subjected to 25 cycles of least-squares refinement with TNT (*23*). The crystallographic *R*-factor was reduced from 44% to 25% against data to 2.6 Å following the refinement. Continued

Table 1: X-ray Diffraction Data

	native	Mn^{2+}•GlucD	TMLA	K_2PtCl_4
resolution (Å)	2.2	2.6	2.7	2.8
R_{merge}[a]	4.0	4.0	7.3	6.9
highest shell[b] (%)	20.1	18.1	22.6	21.4
no. of reflections	72242	53491	70813	72776
indep reflections	21142	12712	12185	10486
completeness (%)	88.0	86.3	92.6	88.6
highest shell[b] (%)	70.8	70.9	79.7	74.1
unit cell				
a (Å)	69.6	69.6	69.7	69.8
b (Å)	108.8	109.2	108.9	108.8
c (Å)	122.6	122.9	122.7	122.4
heavy atom concn (mM)			15	1.1
length of soak			4 days	15 h
R_{iso}[c] (%)			16.5	13.4
phasing power[d]			2.20	1.31
no. of sites			2	3
binding sites			(I) D241, E266 (II) C255	(I) M296 (II) M413, H416 (III) M422, H27

[a] $R_{merge} = \sum(||I_{hi}| - |I_h||)/\sum I_{hi} \times 100$, where I_{hi} and I_h are the intensities of individual and mean structure factors, respectively. [b] The highest resolution shell for native is 2.4−2.3 Å, for Mg•GlucD is 2.7−2.6 Å, for trimethyllead acetate is 2.8−2.7 Å, and for K_2PtCl_4 is 2.9−2.8 Å. [c] $R_{iso} = \sum(||F_h| - |F_n||)/\sum|F_n| \times 100$, where F_h and F_n are the heavy atom and native structure factors, respectively. [d] The phasing power is defined as the mean value of the heavy atom structure factor divided by the lack-of-closure error. The anomalous signals were collected for the TMLA and K_2PtCl_4 derivative data sets and were used to assist in the refinement of the heavy atom positions. The overall mean figure of merit for the data to 2.6 Å for the native and two heavy atom derivatives was 0.66.

model building with FRODO and least-squares refinement while extending the resolution to 2.3 Å reduced the *R*-factor to 23.5%; thereafter, solvent molecules were added to the model. The positions of solvent molecules were identified with the PEKPIK program of the TNT package (*23*) and checked manually for good geometry. The final structure contains 147 water molecules.

Computational Structural Analyses. The molecular surface area was determined with EDPDB (*24*) using a probe of radius 1.4 Å. The structural superposition in Figure 5 was performed using ALIGN (*25*). The superposition of the four proteins in Figure 6 was performed with LSQKAB (*26*, *27*) using Cα positions in strands β6, β7, and β9 of the β-barrel domain, three strands that contain important active site residues, as the basis of the alignment. Figures 2−6 were made with MOLSCRIPT (*28*). Electron density was included in Figure 2 with MOLDED (*29*).

Crystallization in the Presence of Metal Ions. To locate the position of the metal ions in the active site, a native crystal of GlucD was incubated in the presence of 1 mM $MnCl_2$ for 1 week. The soaking experiment was carried out under the same experimental conditions as the search for heavy atom derivatives. A data set was collected for a Mn^{2+} soaked crystal that diffracted to 2.6 Å (Table 1). The crystal was isomorphous with the native crystals. The structure was solved by difference Fourier techniques and submitted to 25 rounds of TNT refinement (*23*). The starting model for refinement was the final structure of the native GlucD protein from which all solvent molecules were removed and for which all temperature factors were set to 25.0 Å^2. The crystallographic *R*-factor started at 31.7% against data to 2.7 Å and dropped to 19.9% after 25 rounds. Electron density maps were examined with FRODO (*22*), and density was observed in the active site. The highest peak was located at the position coordinated by Asp 241, Glu 266, and Asn 295. The protein was manually fitted prior to an additional round of least-squares refinement. During this procedure, the side chains of Arg 75 and Glu 220 were removed because little electron density was observed for these atoms. Following the next round of refinement, a difference map was made at the active site showing the location of the Mn^{2+} ion (Figure 2B). The final structure was obtained by inclusion of the metal ion and subjecting the structure to five cycles of TNT refinement. The rms deviation between the Mn^{2+} structure and the native structure for all 1532 main chain atoms is 0.2 Å using ALIGN (*25*).

RESULTS AND DISCUSSION

Structure Determination of GlucD. The GlucD protein is a tetramer of four identical polypeptides, each containing 451 amino acids. The protein structure was solved by MIR from two heavy atom derivatives (Table 1). The model was built into a native MIR map that was improved by solvent flattening (*21*). The final structure of the glucarate dehydratase contains 399 amino acids out of 451. Three disordered regions were located: one break in the polypeptide chain at residues 99−110 that is located at the C-terminal end of the barrel (the location of the active site) and both the N- and C-termini are disordered. The final model consists of residues Gly 12−Ala 98 and Asp 111−Met 422. The side chains of 14 residues are disordered (Lys 69, Asp 111, Leu 112, Arg 113, Glu 175, Asp 177, Lys 186, Lys 249, Glu 386, Lys 399, Lys 400, Asp 410, Lys 414, and Glu 417) and have been built as alanine residues. Nine of these residues are located near the mobile portions of the protein, namely, the C-terminus and the break between residues Ala 98 and Asp 111. The final model also includes 144 water molecules incorporated where $F_o - F_c$ density was greater than 2.5σ and hydrogen-bonding geometry consistent with a water molecule was observed.

The geometry of the protein model was analyzed with PROCHECK (*30*) and TNT (*23*). Ninety percent of the amino acids fall in the most favored portion of the Ramachandran plot; 10% fall in the additionally allowed regions. One residue, Asp 267, falls in the generously allowed region of the plot ($\phi = 31°$, $\psi = 61°$); however, the electron density for this residue is unambiguous. This residue is located at the active site, and this orientation may be necessary to allow the neighboring side chain of Glu 266 to interact with the metal ion or the substrate (see below). The average temperature factor is 38.3 Å^2 for all atoms and 34.8 Å^2 for all main chain atoms (Table 2). The final *R*-factor for the data to 2.3 Å is 19.0%. Representative electron density is shown in Figure 2.

Structure of the GlucD Subunit. As expected on the basis of its membership in the enolase superfamily, GlucD contains two domains, an N-terminal domain, Met 1−Val 136, and a C-terminal domain, Pro 137−Arg 451, that contains the TIM β-barrel (*3*, *31*, *32*). The residues that make up the major secondary structure elements are listed in Table 3.

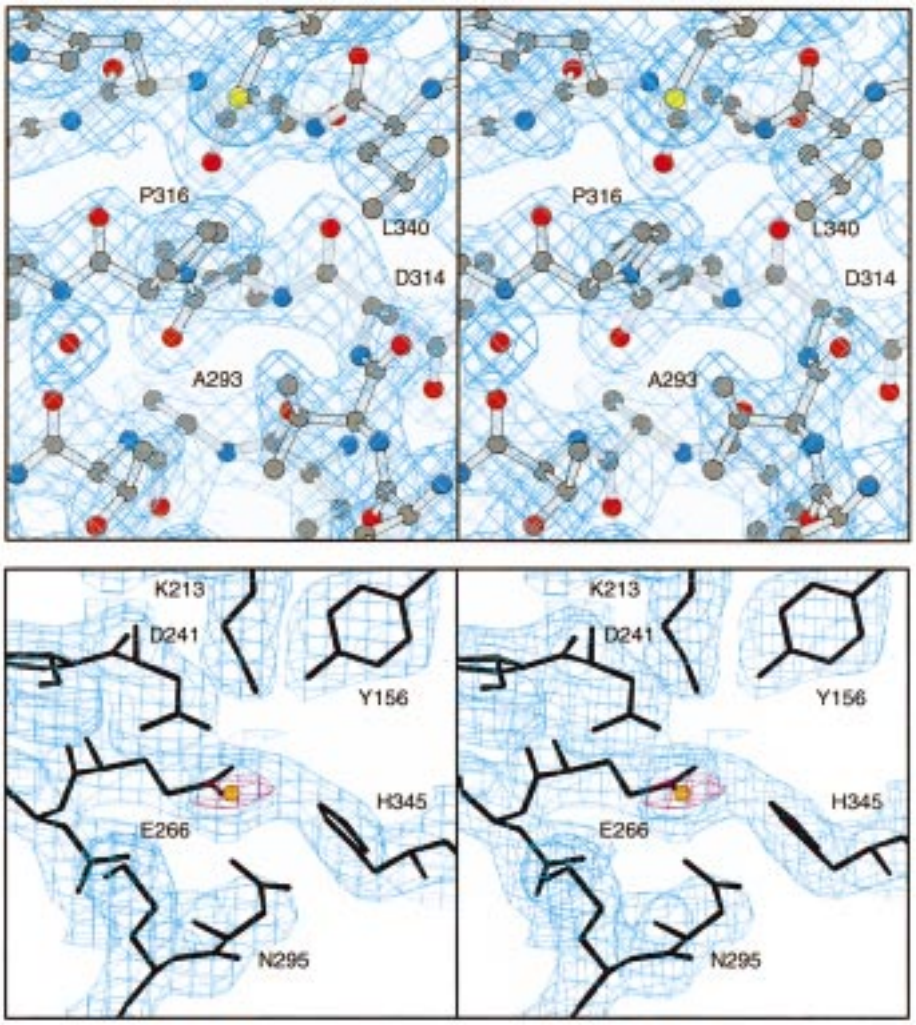


FIGURE 2: Representative electron density. Panel A (top): Final electron density from the native data set calculated from coefficients of the form $2F_o - F_c$ and contoured at 1σ. A portion of the C-terminal domain is shown. Protein atoms are colored gray, blue, red, and yellow for carbon, nitrogen, oxygen, and sulfur atoms, respectively. Panel B (bottom): Density of the metal binding site of the Mn^{2+} data set. The Mn^{2+} atom is shown in orange. The three oxygens which coordinate the metal ion are located 2.1 Å (Asp 241), 1.9 Å (Glu 266), and 2.4 Å (Asn 295) from the metal ion. The density depicted in blue was calculated with coefficients of the form $2F_o - F_c$ and is contoured at 1σ. The electron density depicted in red was calculated with $F_o - F_c$ coefficients determined prior to inclusion of the Mn^{2+} ion in the structure. This map is contoured at 6.5σ. This figure and Figures 3−6 were generated with MOLSCRIPT (*28*).

Table 2: Refinement Data

	native	Mn^{2+}·GlucD
resolution used for refinement (Å)	30.0−2.3	30.0−2.7
no. of reflections used	20284	12098
total atoms	3188	3028
no. of protein atoms	3041	3027
no. of solvent atoms	147	0
av *B*-factor (Å$^{2)}$	38.3	35.0
av *B*, protein (Å$^{2)}$	37.9	35.0
av *B*, main chain (Å$^{2)}$	34.8	32.0
av *B*, solvent (Å$^{2)}$	47.2	—[a]
R-factor (%)	19.0	19.4
rms deviation		
bond lengths (Å)	0.017	0.014
bond angles (deg)	2.53	2.23
trigonal planes	0.005	0.005
general planes	0.015	0.012

[a] No solvent molecules are present in the Mn^{2+} structure.

The N-terminal domain contains three strands ($\beta1-\beta3$) that form an antiparallel β-sheet (Figure 3). The loop joining the first and second strands extends from residues Val 24 to Thr 43 and reaches from the N-terminal domain to the β-barrel region of the protein, interacting at the carboxy-terminal ends of the strands. This allows components of the N-terminal domain to interact with residues that are believed to be part of the active site. Three helices ($\alpha1-\alpha3$) follow the antiparallel sheet. The connection between $\alpha2$ and $\alpha3$ is disordered and is not included in the final model.

The C-terminal two-thirds of the molecule form the β-barrel domain. Two short structural elements, $\alpha4$ and $\beta4$, precede the barrel. This domain is a typical β-barrel with the exception of an extended loop and two helices between the first and second strands of the barrel ($\beta5$ and $\beta6$). The loop between the fourth strand of the barrel ($\beta8$) and the following helix is somewhat longer than the remaining loops, perhaps because this region of the protein is involved in interactions with another subunit of the tetramer. Following the final strand of the β-barrel, two short strands together with strand $\beta4$ form a short three-stranded antiparallel sheet. The molecule ends with a helix that precedes the disordered 28 residues at the C-terminus of the protein.

Table 3: Secondary Structure

N-terminal domain		C-terminal (α/β-barrel) domain					
element	residues	element	residues	element	residues	element	residues
β1	Ile 16−Val 22	α4	Val 138−Leu 141	β7	Arg 237−Asp 241	α10	Gln 326−Trp 338
β2	Ile 46−Asp 51	β4	Ala 150−Glu 152	α7	Leu 248−Cys 255	β11	Thr 341−His 345
β3	Val 56−Glu 60	β5	Gly 155−Leu 157	β8	Leu 262−Glu 266	α11	Ser 353−Ala 361
α1	Gly 66−Asp 73	α5a	Glu 179−Asn 185	α8	Gly 277−Thr 288	β12	Ala 370−Thr 373
α2	Ile 83−Ala 96	α5b	Pro 191−Arg 204	β9	Thr 292−Thr 294	β13	Gln 389−Lys 391
α3	Ala 115−Leu 134	β6	Asp 209−Lys 213	α9	Trp 301−Leu 310	β14	Leu 394−Val 397
		α6	Gly 219−Arg 232	β10	Ile 315−Asp 319	α12	Ala 411−Lys 420

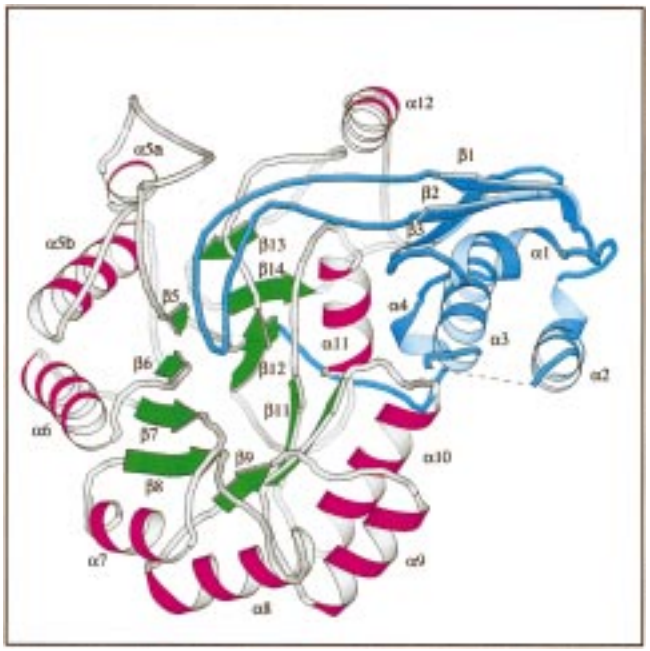


FIGURE 3: Ribbon diagram of the GlucD subunit and ribbon representation of the GlucD monomer. The N-terminal domain is shown entirely in blue. The sheets of the β-barrel domain are depicted in green while the helices are shown in red.

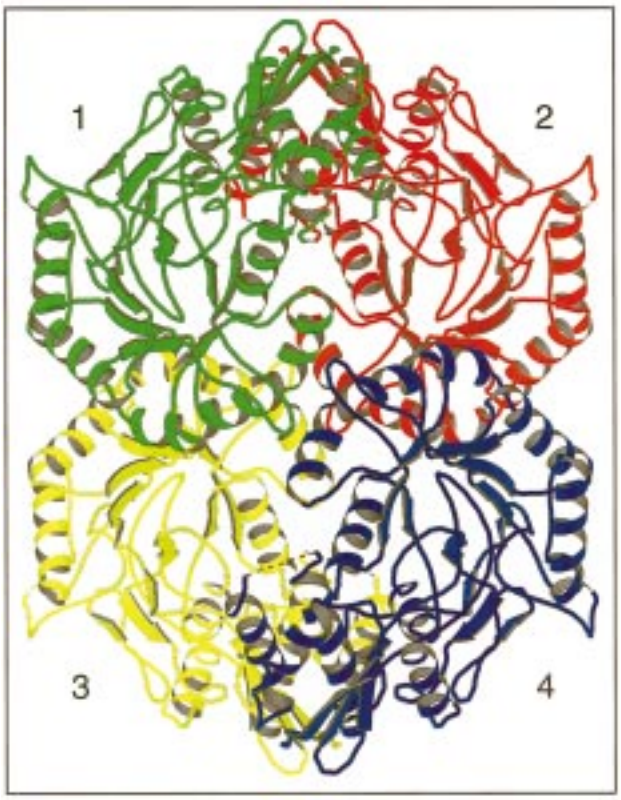


FIGURE 4: Ribbon diagram of the GlucD tetramer. The GlucD protein is a tetramer with 222 symmetry. The protein is represented as a ribbon diagram looking down one of the 2-fold axes. Subunit 1 is in roughly the same orientation as the subunit shown in Figure 3. The interactions with other subunits are made largely through the β-barrel domain.

Subunit Interactions of the GlucD Tetramer. The GlucD tetramer exhibits 222 symmetry, containing three orthogonal 2-fold rotation axes (Figure 4). Each subunit interacts extensively with one other subunit of the tetramer and weakly with the remaining two subunits. GlucD is thus a dimer of dimers. In isolation, each subunit has a molecular surface area of 15 936 Å^2. The interface between subunits in a pseudodimer consists of 1360 Å^2 of surface from each subunit; an additional 1090 Å^2 of surface from each subunit is utilized in forming the complete tetramer as determined with the program EDPDB using a probe size of radius 1.4 Å (*24*).

The interactions between subunits 1 and 2 (as labeled in Figure 4) result from two sets of interactions. These consist of a hydrophobic patch on helices α9 and α10 and a region of extensive hydrogen-bonding interactions between the N-terminal domain and the start of the β-barrel domain. The hydrophobic patch contains residues Trp 301, Arg 302, Met 304, Gly 305, Ile 308, Met 334, and Trp 338 from both subunits. The hydrogen-bonding interactions between subunits 1 and 2 include direct interactions between the side chain of Gln 87 and the carbonyl of Gly 143 and also between the side chain of Asn 91 and the carbonyl of Glu 144. Water-mediated interactions occur between the main chain carbonyls of Gly 84 and Ala 140, the side chains of Arg 94 and Glu 144, the main chain nitrogens of Ser 88 and Gln 146, the side chain of Asn 85 and the main chain nitrogen of Gln 146, and the side chain of Asn 91 and the backbone carbonyl of Gln 146. The helix α10 from one subunit runs nearly perpendicular to the α10 helix from the neighboring subunit, thereby allowing two salt links to form between Arg 330 of one subunit and Glu 337 of the neighbor.

Fewer interactions are observed between the pseudodimers. Interactions occur between residues Gly 274−Glu 283 from one subunit and residues Ile 297−His 306 from a subunit of the alternate pseudodimer (subunits 1 and 3 in Figure 3). In this region, the side chain of Tyr 275 from one subunit stacks against the side chain of Arg 302 from the other. The side chain of Glu 283 forms ionic interactions with the side chains of Arg 302 and His 306. Additionally, the side chain of Glu 279 interacts with the side chains of Ser 276 and Arg 278 from the neighboring subunit. Interactions between subunits 1 and 4 occur between the side chain of Arg 286 and Glu 311 of subunit 4 and between the carbonyl of Arg

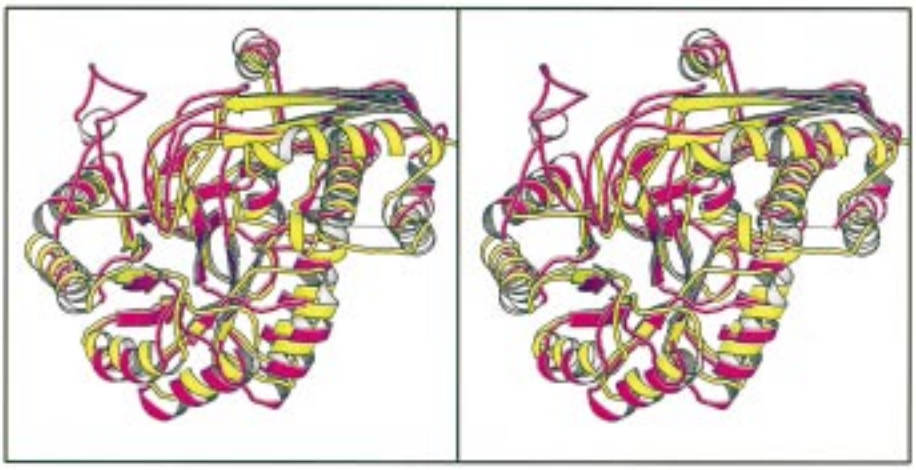


FIGURE 5: Superposition of GlucD and MR (2MNR). The two structures were rotated into the same orientation using ALIGN (*24*). The structure of GlucD is shown in magenta while that of MR is shown in yellow. The dashed line represents the break in the structure of GlucD. The 29 amino acid insertion in GlucD relative to MR can be seen in the upper left corner of the figure.

286 and the side chain of Arg 291. All four subunits come into close contact near the end of helix α9; in this region, Gln 309−Gln 311 from all four subunits interact.

Evolution of Structure and Function in the Enolase Superfamily. A superposition of the α-carbon backbones of GlucD and MR is shown in Figure 5. The two proteins were aligned with the program ALIGN (*24*). Additionally, the β-barrel domains of four proteins from the enolase superfamily are shown in Figure 6. From this superposition and the sequence and functional correlations described above, we conclude that the members of the superfamily are related by divergent evolution.

As noted in the introduction, the primary sequence alignment of MR with the homologous GlucDs (Figure 1) allowed the prediction of the functional groups in the active site of GlucD from *P. putida* that facilitate the epimerization of (*D*)-glucarate and (*L*)-idarate to be homologous to those in the active site of MR. Figure 6 shows the alignment of the active sites of GlucD, MR, MLE I, and enolase.

As predicted, the superposition reveals that Lys 211 and Lys 213 in GlucD are structurally homologous with Lys 164 and Lys 166 in MR. In MR, Lys 164 interacts with the carboxylate oxygen of the substrate that is coordinated to the essential Mg^{2+}, presumably assisting in the stabilization of the enolic intermediate. Although the present structure of GlucD was determined in the absence of a substrate or substrate analogue, we predict that Lys 211 will play an analogous role in the GlucD-catalyzed reaction. In MR, Lys 166 is the (*S*)-specific base that mediates proton transfers from and to (*S*)-mandelate; in stereochemical analogy, we predict that Lys 213 is the base that mediates proton transfers from and to (*D*)-glucarate [the 5-carbon of (*D*)-glucarate has the (*S*)-absolute configuration].

The superposition also confirms the prediction that a hydrogen-bonded dyad between His 345 and Asp 319 can be superimposed on the dyad between His 297 and Asp 270 in the active site of MR. The His 297−Asp 270 dyad functions as the (*R*)-specific base that mediates proton transfers from and to (*R*)-mandelate, so we predict that the His 345−Asp 319 dyad in GlucD is the base that mediates proton transfers from and to (*L*)-idarate [the 5-carbon of (*L*)-idarate has the (*R*)-absolute configuration].

A GlucD crystal was soaked in 1 mM $MnCl_2$ to identify the location of the metal ion in the active site (Figure 2A). In the native structure, the solvent molecule OW 571 approximates the location of the metal ion. The structure of GlucD reveals the identities of the three ligands for the conserved Mg^{2+} ion, the carboxylate groups of Asp 241 and Glu 266 and the carboxamide group of Asn 295. The ligands for the Mg^{2+} ion in the active site of MR are the carboxylate groups of Asp 195, Glu 221, and Glu 247. Thus, while the identities of the first two ligands for the Mg^{2+} ion are conserved in the two active sites, the third ligand for the Mg^{2+} ion in GlucD is not conserved. The catalytic implications of this differing charge in the inner coordination sphere of the Mg^{2+} ion are unknown, although the increase in the positive charge on the Mg^{2+} ion could be expected to provide additional electrostatic stabilization of the enolic intermediate. In this context, we note that the pK_a values of the α-protons of (*D*)-glucarate/(*L*)-idarate are expected to be approximately 3 units greater than those of the enantiomers of mandelate, 32 vs 29, respectively (*33*).

The superposition reveals that the carboxylic acid group of Asp 372 in the active site of GlucD is spatially homologous to the carboxylic acid group of Glu 317. The latter electrophilic catalyst contributes significantly to the catalytic proficiency of the MR-catalyzed reaction by hydrogen bonding to the carboxylate oxygen of the substrate that is not coordinated to the Mg^{2+} ion, so it is hypothesized that the carboxylic acid group of Asp 372 functions analogously in the GlucD-catalyzed reaction.

In addition to these conserved features in the active sites of both GlucD and MR, the structure reveals a new functional group that has no homologue in the active site of MR but which is likely important for catalysis in GlucD: the phenolic hydroxyl group of Tyr 156 is observed in close proximity to the conserved features and may be appropriately positioned to catalyze either the β-elimination of the 4-hydroxyl group (Scheme 2, reaction 2) and/or the stereospecific protonation of the resulting intermediate, reactions not catalyzed by MR (Scheme 2, reaction 3). Tyr 156 is located at the end of the first strand of β-sheet in the β-barrel. Given the previous observation that the functional groups in the active site of MR are located at the ends of all of the β-sheets with the

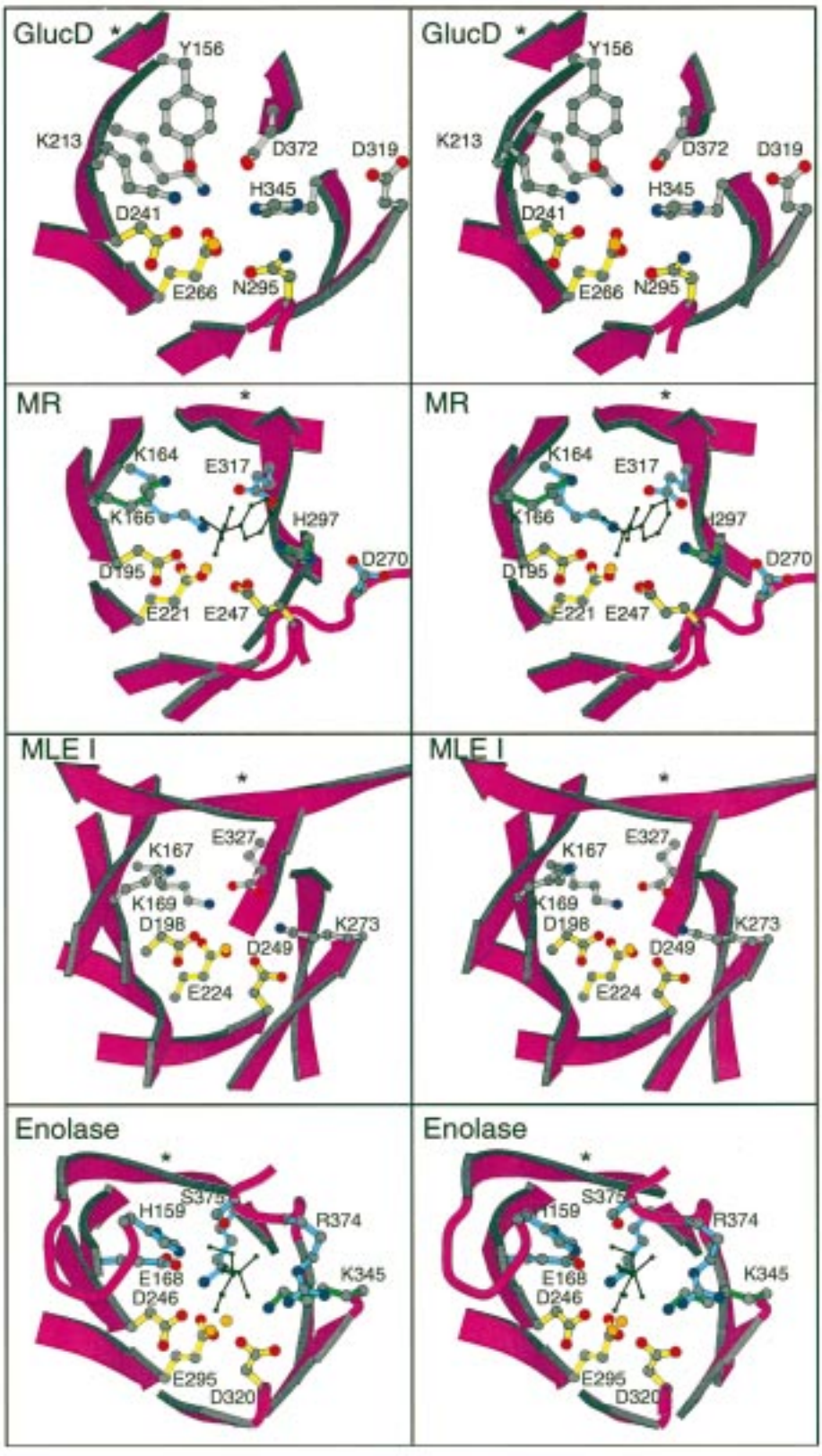


FIGURE 6: β-Barrel domains of four members of the enolase superfamily. The structures of the barrel domains of four proteins are shown: enolase (1EBH), GlucD (this paper), muconate lactonizing enzyme (1MUC), and mandelate racemase [1MDR; (*S*)-atrolactate bound in the active site]. The strands that form the β-barrel for each enzyme are shown, as are the subsequent loops that contain residues that are involved in the active site. The first strand of each β-barrel is indicated with an asterisk. The metal ions, Mg^{2+} in enolase and MR and Mn^{2+} in GlucD and MLE, are shown in orange. Enolase and mandelate racemase contain a molecule of 2-phosphoglycerate and (*S*)-atrolactate, respectively, shown in black. The active site residues are labeled and are colored according to their function in the active site. The residues that coordinate the metal ion are in yellow. The catalytic residues that function to remove the protein from the α-carbon are colored green. The residues that are involved in hydrogen-bonding or ionic interactions with the substrate are shown in blue, and residues of current uncertain function are shown in gray.

exception of the first, this structure suggests that the dehydration reactions catalyzed by GlucD (general acid-catalyzed departure of the hydroxide leaving group and/or stereospecific protonation of the resulting enol intermediate) require an extra functional group beyond those present in the active site of MR that can be accommodated by the structural scaffold of the β-barrel. The location of Tyr 156 is consistent with the hypothesis that evolution of new functions in enzymes that contain β-barrels can be accomplished by independent variation of the identities of the functional groups at the ends of each of the eight β-strands (*3*, *4*). Experiments to test the role of Tyr 156 in catalysis are underway.

Structure-Based Alignment of the Primary Sequences of MR and GlucD. As described in the previous section, the genetics-based alignment of the primary sequences of MR and the homologous GlucDs presented in Figure 1 (panel A; obtained with Clustal W) leads to the successful prediction of the identities of most of the functional groups that likely are essential to the reactions catalyzed by GlucD. However, this genetics-based alignment did not predict the presence of either Tyr 156 or Asp 372 in the active site of GlucD nor the identity of the third ligand for the metal ion.

In panel B of Figure 1 we present a structure-based sequence alignment of the primary sequences of MR with GlucD that utilizes the observed structural elements to relate the primary sequences. Tyr 156 is located in a loop at the end of the first strand of β-sheet in the β-barrel in GlucD and has no known functional homologue in the active site of MR. Asp 372 is located at the end of the eighth strand of β-sheet in GlucD and, as described in the previous section, likely is a functional homologue of Glu 317, an electrophilic catalyst in the active site of MR. The Glu to Asp change is functionally conservative, but without a bound substrate or substrate analogue we cannot comment on whether this difference may be correlated to a change in the geometry of the coordination of the carboxylate group of the substrate to the conserved Mg^{2+} or the nature of stabilization of the enolic intermediate.

The structure-based alignment does not correctly match the third residues that coordinate the metal ion: Asn 295 of GlucD is matched with Gly 246 rather than Glu 247 of MR. This difference results from a slightly different orientation of the two loops that contain these residues. While the structure-based alignment minimizes the distance between Cα's of the two proteins, the carboxylate side chain of Glu 247 is only 2.4 Å from the carboxamide of Asn 295 in the final alignment.

The apparent evolution of function that incorporates both Tyr 156 and Asp 372 in the active site at the C-terminal end of the β-barrel is characterized by changes in both the primary sequences and local structures of the (structurally adjacent) first and eighth β-strands in the β-barrel relative to those in MR, thereby preventing a functionally and structurally meaningful alignment of their primary sequences in the absence of three-dimensional structural information. Such changes are in accord with our hypothesis (*3*, *4*) that the β-barrel scaffold is well-suited for the evolution of new functions while retaining functional and structural features of the progenitor member of the superfamily.

CONCLUSIONS

The high-resolution X-ray structure of GlucD from *P. putida* reveals, as expected, that both domains of its three-dimensional structure are homologous to those in MR, MLE I, and enolase. The structure also fully confirms the earlier predictions regarding the identities of the active site functional groups that were based upon the alignment of its primary sequence with that of MR as well as provides new information concerning the structural requirements for catalysis by GlucD (e.g., Tyr 156). This structure justifies the approach of combining sequence alignment with a structural database to understand the evolution of enzyme diversity, in general, and expands our understanding of the enolase superfamily, in particular.

ACKNOWLEDGMENT

We thank Cheom-Gil Cheong for assisting in the growth of GlucD crystals containing metal ions.

REFERENCES

1. Horowitz, N. H. (1945) *Proc. Natl. Acad. Sci. U.S.A. 31*, 153−157.
2. Horowitz, N. H. (1965) in *Evolving Genes and Proteins* (Bryson, V., and Vogel, H. J., Eds.) pp 15−23, Academic Press, New York.
3. 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.
4. Babbitt, P. C., and Gerlt, J. A. (1997) *J. Biol. Chem. 272*, 30591−30594.
5. 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.
6. Neidhart, D. J., Howell, P. L., Petsko, G. A., Powers, V. M., Li, R., Kenyon, G. L., and Gerlt, J. A. (1991) *Biochemistry 30*, 9264−9273.
7. Helin, S., Kahn, P. C., Guha, B. L., Mallows, D. G., and Goldman, A. (1995) *J Mol. Biol. 254*, 918−941.
8. Larsen, T. M., Wedekind, J. E., Rayment, I., and Reed, G. H. (1996) *Biochemistry 35*, 4349−4358.
9. Wedekind, J. E., Poyner, R. R., Reed, G. H., and Rayment, I. (1994) *Biochemistry 33*, 9333−9342.
10. Kallarakal, A. T., Mitra, B., Kozarich, J. W., Gerlt, J. A., Clifton, J. G., Petsko, G. A., and Kenyon, G. L. (1995) *Biochemistry 34*, 2788−2797.
11. Mitra, B., Kallarakal, A. T., Kozarich, J. W., Gerlt, J. A., Clifton, J. G., Petsko, G. A., and Kenyon, G. L. (1995) *Biochemistry 34*, 2777−2787.
12. 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.
13. Palmer, D. R. J., and Gerlt, J. A. (1996) *J. Am. Chem. Soc. 118*, 10323−10324.
14. Palmer, D. R. J., Hubbard, B. K., and Gerlt, J. A. (1998) *Biochemistry 37*, 14350−14357.
15. 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.
16. Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
17. Kabsch, W. (1988a) *J. Appl. Crystallogr. 21*, 67−71.
18. Kabsch, W. (1988b) *J. Appl. Crystallogr. 21*, 916−924.
19. Wesenberg, G., and Rayment, I. (1998) (in preparation).
20. Terwilliger, T. C., and Eisenberg, D. (1983) *Acta Crystallogr. A39*, 813−817.
21. Wang, B. C. (1985) in *Methods in Enzymology* (Wyckoff, H. W., Hirs, C. H. W., and Timasheff, S. N., Eds.) pp 90−112, Academic Press, New York.

22. Jones, T. A. (1985) *Methods Enzymol.*, 157−171.
23. Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987) *Acta Crystallogr. A43*, 489−501.
24. Zhang, X.-J., and Matthews, B. M. (1995) *J. Appl. Crystallogr. 28*, 624−630.
25. Cohen, G. H. (1997) *J. Appl. Crystallogr. 30*, 1160−1161.
26. CCP4 (1994) *Acta Crystallogr. D50*, 760−763.
27. Kabsch, W. (1976) *Acta Crystallogr. A32*, 922−923.
28. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
29. Fisher, A. J. (1998) (in preparation).
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
31. Babbitt, P. C., Mrachko, G. T., Hasson, M. S., Huisman, G. W., Kolter, R., Ringe, D., Petsko, G. A., Kenyon, G. L., and Gerlt, J. A. (1995) *Science 267*, 1159−1161.
32. Farber, G. K., and Petsko, G. A. (1990) *Trends Biochem. Sci. 15*, 228−234.
33. Gerlt, J. A., Kozarich, J. W., Kenyon, G. L., and Gassman, P. G. (1991) *J. Am. Chem. Soc. 113*, 9667−9669.
34. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1978) *Arch. Biochem. Biophys. 185*, 584−591.
35. Chothia, C., and Lesk, A. M. (1986) *EMBO J. 5*, 823−826.

BI981123N