Homologous $(\beta/\alpha)_8$-Barrel Enzymes That Catalyze Unrelated Reactions: Orotidine 5′-Monophosphate Decarboxylase and 3-Keto-L-Gulonate 6-Phosphate Decarboxylase[†,‡]

Eric Wise,[§,||] Wen Shan Yew,[||,⊥] Patricia C. Babbitt,[#] John A. Gerlt,*[,⊥] and Ivan Rayment*[,§]

Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, University of California, San Francisco, California 94143, and Departments of Biochemistry and Chemistry, University of Illinois, Urbana, Illinois 61801

Received December 19, 2001; Revised Manuscript Received January 29, 2002

ABSTRACT: The 3-keto-L-gulonate 6-phosphate decarboxylase (KGPDC) encoded by the *ulaD* gene in the *Escherichia coli* genome [Yew, W. S., and Gerlt, J. A. (2002) *J. Bacteriol. 184*, 302−306] and orotidine 5′-monophosphate decarboxylase (OMPDC) are homologous (derived from a common ancestor) but catalyze different reactions. The metal-independent decarboxylation reaction catalyzed by OMPDC avoids the formation of a vinyl anion intermediate; the Mg^{2+}-dependent decarboxylation reaction catalyzed by KGPDC involves the formation of an enediolate anion intermediate. Based on the available structures of OMPDC, a sequence alignment allows the predictions that (1) KGPDC is a dimer of $(\beta/\alpha)_8$-barrels, with the active sites located at the dimer interface; (2) KGPDC and OMPDC share an aspartate residue at the end of the first β-strand and an Asp-x-Lys-x-x-Asp motif at the end of the third β-strand with OMPDC; but (3) KGPDC has a Glu instead of a Lys at the end of the second β-strand. The structure of KGPDC has been determined in the presence of Mg^{2+} and the substrate analogue L-gulonate 6-phosphate and confirms these predictions. The carboxylate functional groups at the ends of the first, second, and third β-strands in KGPDC are ligands of the Mg^{2+}; in OMPDC, the homologues of these residues participate in a hydrogen-bonded network that facilitates the decarboxylation reaction. The 3-OH group of the substrate analogue is coordinated to the Mg^{2+}, supporting the hypothesis that the mechanism of the decarboxylation catalyzed by KGPDC involves stabilization of an enediolate anion intermediate. These structural studies establish the existence of the OMPDC "suprafamily," in which members catalyze reactions that occur in different metabolic pathways and share no mechanistic relationship. The existence of this suprafamily demonstrates that divergent evolution can be opportunistic, conscripting active site features of a progenitor to catalyze unrelated functions. Accordingly, sequence or structure homology alone cannot be used to infer the functions of new proteins discovered in genome projects.

Understanding the relationships among protein sequence, structure, and function has long been a major focus of biochemistry. This problem now is even more acute with the results of numerous genome projects in which approximately 40% of the encoded proteins have unknown or uncertain functions. The constant value of this percentage is surprising, if, as estimated, the number of protein folds is modest (≤5000) (*1*−*3*). This suggests that some folds are reused and that extreme sequence divergence prevents recognition of proteins derived from a common ancestor and, therefore, possibly sharing functional features.

One such versatile fold is the $(\beta/\alpha)_8$-barrel which occurs in more enzymes than any other fold, although it is uncertain whether this fold arose once (followed by divergent evolution) or many times (convergent evolution) (*4*). This fold is well-suited for evolution of function: the catalytic groups are located on the loops that connect the C-terminal ends of the β-strands with the following α-helices. So placed, the active site is surrounded by groups that can be varied independently to catalyze new reactions, though the "rules" for how this might be accomplished in a rational manner are unknown. In the absence of rules, the only route to understanding those changes in sequence and structure that lead to new function is through careful functional and structural characterization of distantly related proteins sharing the $(\beta/\alpha)_8$-barrel. With sufficient information, the structural strategies that underlie the evolution of this and other folds can be defined so that the assignment of function to reading frames of unknown function may be facilitated.

[†] This research was supported by Grants GM-60595 (to P.C.B.), GM-40570 (to J.A.G.), GM-52594 (to J.A.G. and I.R.), and AR-35186 (to I.R.) from the National Institutes of Health. Use of the Argonne National Laboratory Structural Biology Center beamline at the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Energy Research, under Contract W-31-109-ENG-38.

[‡] The X-ray coordinates and structure factors for KGPDC have been deposited in the Protein DataBank (KGPDC·Mg^{2+}·phosphate and KGPDC·Mg^{2+}·L-gulonate 6-phosphate complexes: 1KV8 and 1KW1, respectively).

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

[§] University of Wisconsin, Madison.

[||] These authors contributed equally to this work.

[⊥] University of Illinois, Urbana-Champaign.

[#] University of California, San Francisco.

10.1021/bi012174e CCC: $22.00
Published on Web 03/03/2002

Two strategies for natural divergent evolution of enzyme function are widely accepted: retention of either chemical mechanism or ligand specificity. Each follows duplication of the gene encoding the progenitor so that the original function in metabolism needs to be retained. They differ in the essential feature of the progenitor that is conserved to provide the new function, although in both strategies the evolution of new activity is expected to occur within the context of a stable, folded protein (*4*, *5*).

In the first strategy, the progenitor provides the structural framework to catalyze a difficult partial reaction. This gives rise to mechanistically diverse superfamilies which are defined as homologous enzymes sharing generally less than 50% sequence identity and catalyzing different overall reactions but with a common mechanistic attribute (*6*). Examples of mechanistically diverse superfamilies that utilize the $(\beta/\alpha)_8$-barrel fold include the following: the enolase superfamily, whose members utilize divalent metal ion-mediated generation of an enolate anion derived from a carboxylate substrate (*7*); and the amidohydrolase/phosphotriesterase superfamily, whose members utilize a hydroxide ion coordinated to one or more divalent metal ions to mediate the hydrolyses of various amides and esters (*8*).

Functional diversity that retains a common mechanistic feature is also supported by other folds. In the crotonase superfamily [four turns of a ($\beta\alpha\beta$) superhelix], a conserved oxyanion hole stabilizes enolate anions usually (*9*), but not always (*10*), derived from coenzyme A esters; in the haloacid dehalogenase superfamily (β-sheet containing eight strands), a conserved Asp residue acts as the nucleophile in displacement reactions at either carbon or phosphorus (*11*−*13*); and in both the vicinal oxygen chelate superfamily (pairs of $\beta\alpha\beta\beta\beta$ motifs) (*14*−*16*) and the cupin superfamily (β-barrel) (*17*), a divalent metal ion acts as an electrophile. Although the substrates, mechanisms, and overall reactions for members of these superfamilies often are more diverse than those for members of the enolase superfamily, the active sites contain conserved functional features that are used in the same way in each reaction.

In the second strategy, the progenitor provides the template for binding a ligand or substrate, which can be modified to evolve the catalyst for the next reaction in the pathway. Examples include $(\beta/\alpha)_8$-barrel enzymes in the histidine and tryptophan biosynthetic pathways that are constructed from half-barrels (*18*).

Although these two strategies may appear sufficient for evolution of new functions, a third possibility is that evolution is opportunistic. In this case, the progenitor provides a structural framework and functional groups that can be used in different mechanistic and metabolic contexts in the evolutionary product. In this paper, we establish the first example of this strategy.

Orotidine 5′-monophosphate decarboxylase (OMPDC)[1] is the most proficient enzyme known (*19*). Recent reports of crystal structures of OMPDC's from four sources renewed attention on the mechanism of this reaction because it is

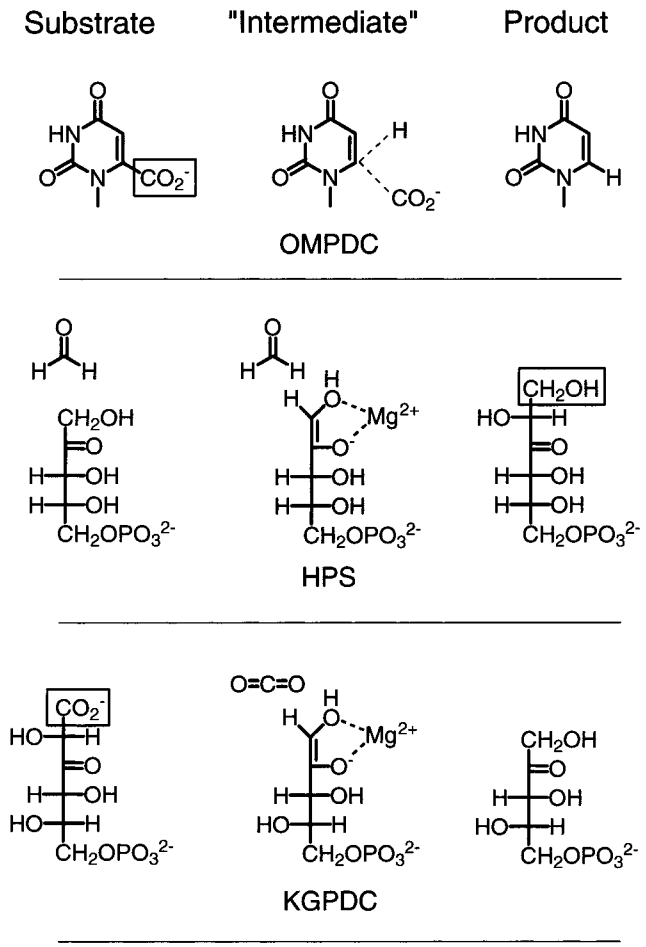

FIGURE 1: Substrates, intermediates, and products for the reactions catalyzed by OMPDC, HPS, and KGPDC. The sites of the chemical transformation in the substrates for OMPDC and KGPDC, and product in HPS are enclosed by a box.

metal ion-independent and avoids the formation of a discrete vinyl anion intermediate that cannot be stabilized by any obvious structural feature(s) of the active site and is much too unstable to be kinetically competent (Figure 1) (*20*−*23*). For this reason, one proposal for the mechanism is a concerted S_E2 reaction involving simultaneous decarboxylation and protonation of C6 (*20*, *22*).

The structures reveal that the enzyme is dimeric where each subunit consists of a $(\beta/\alpha)_8$-barrel, and the barrels are arranged in an antiparallel manner. The sequences of OMPDC's are divergent, with only eight active site residues strictly conserved in the $(\beta/\alpha)_8$-barrel fold. These include several residues thought to be functionally important: an Asp at the end of the first β-strand, a Lys at the end of the second β-strand, and an Asp-x-Lys-x-x-Asp motif at the end of the third β-strand. The active sites are located at the dimer interface: the first four conserved residues are contributed by one polypeptide, and the second Asp of the conserved Asp-x-Lys-x-x-Asp motif is provided by the second polypeptide. In the concerted S_E2 mechanism, the first conserved Asp of the Asp-x-Lys-x-x-Asp motif destabilizes the substrate carboxylate group, and the Lys of the same motif delivers a proton to C6 (*20*, *22*).

Searches of the protein databases identify divergent (<25% sequence identity) homologues of OMPDC with alternative or uncertain functions. Several are Mg^{2+}-dependent D-*ara*-

[1] Abbreviations: SeMet, selenomethionine; MAD, multiwavelength anomalous dispersion; APS, Advanced Photon Source at Argonne National Laboratory, Argonne, IL; KGPDC, 3-keto-L-gulonate 6-phosphate decarboxylase; OMPDC, orotidine 5′-monophosphate decarboxylase; HPS, D-*arabino*-hex-3-ulose 6-phosphate synthase; rms, root-mean-square; MePEG, methyl ether poly(ethylene glycol).

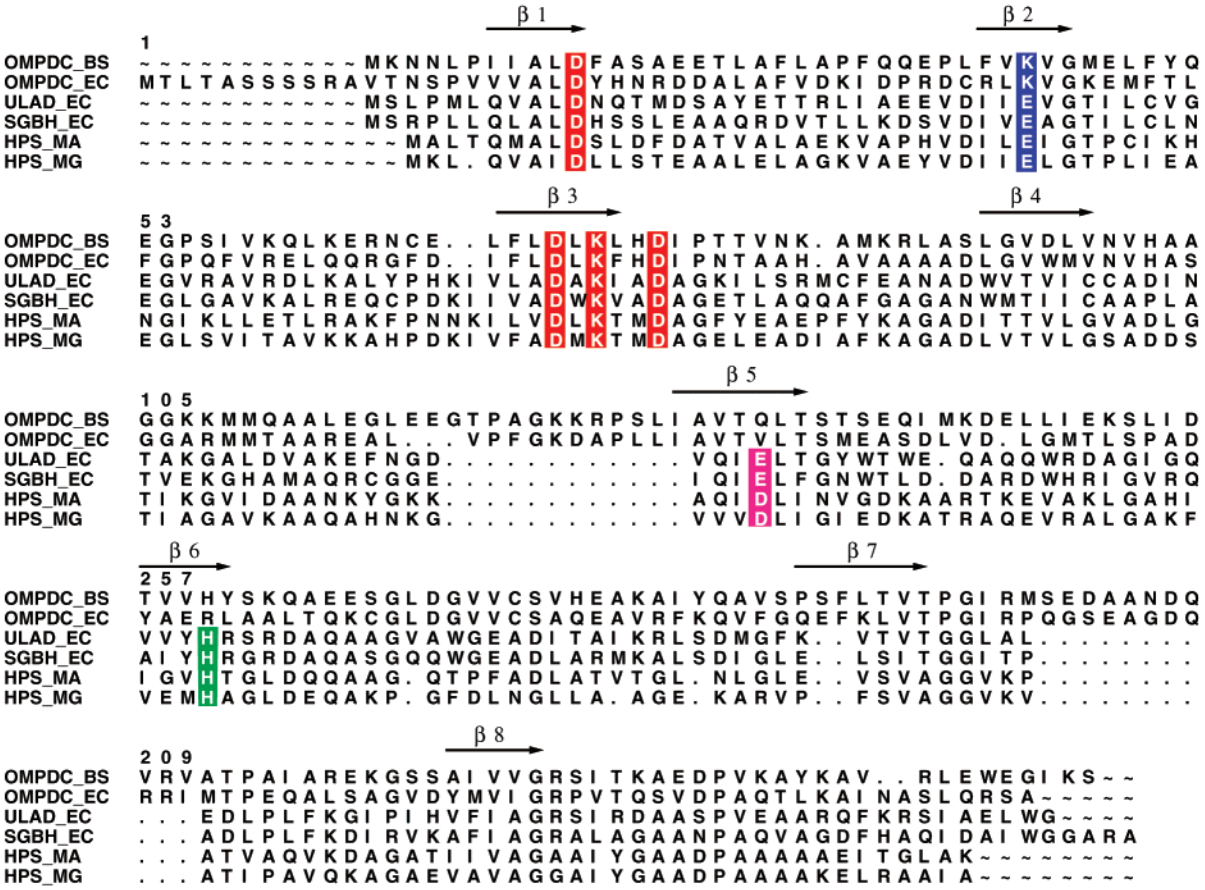

FIGURE 2: Alignment of the sequences of the OMPDC's from *E. coli* (OMPDC_EC) and *B. subtilis* (OMPDC_BS), two homologous KGPDC's from *E. coli* [ULAD_EC is the KGPDC characterized in this paper, and SGBH_EC is a homologue that catalyzes the same reaction (*30*)], and the HPS's from *Methylomonas aminofaciens* (HPS_MA) and *Mycobacterium gastri* (HPS_MG). The strictly conserved Asp^{11} and residues in the Asp-x-Lys-x-x-Asp motif are highlighted in red. Glu^{33} in KGPDC (Lys^{33} in OMPDC) is highlighted in blue. The position of His^{136} is highlighted in green, and the position of Glu^{112} is highlighted in magenta. The alignment was produced by the Pileup algorithm of the GCG software package. The pairwise sequence identity between ULAD_EC and OMPDC_BS is 18%, and that between HPS_METAM and OMPDC_BS is 29%.

bino-hex-3-ulose 6-phosphate synthases (HPS's) that catalyze the condensation of D-ribulose 5-phosphate with formaldehyde in microorganisms that "fix" formaldehyde (*24*−*29*). Others annotated as "probable" HPS's are Mg^{2+}-dependent 3-keto-L-gulonate 6-phosphate decarboxylases (KGPDC's) involved in L-ascorbate utilization by microorganisms, including *Escherichia coli* (*30*). The mechanisms of the HPS- and KGPDC-catalyzed reactions likely involve an enediolate anion intermediate stabilized by coordination to Mg^{2+} (Figure 1). While HPS and KGPDC form a mechanistically diverse superfamily, the mechanisms of their reactions share no features with the reaction catalyzed by OMPDC (*6*).

The sequences of KGPDC's and HPS's are similar in length to those of OMPDC's, consistent with a shared $(\beta/\alpha)_8$-barrel fold. Despite low levels of identity, a sequence alignment (Figure 2) suggests that KGPDC, HPS, and OMPDC share the Asp at the end of the first β-strand and the Asp-x-Lys-x-x-Asp motif at the end of the third β-strand (highlighted with red in Figure 2). However, the Lys at the end of the second β-strand in OMPDC is replaced with a Glu in the KGPDC and HPS (highlighted with blue in Figure 2). The expected presence of carboxylate ligands at the ends of the first three β-strands in the $(\beta/\alpha)_8$-barrels of KGPDC and HPS suggests that these form the binding site for the essential Mg^{2+}. The conserved Asp-x-Lys-x-x-Asp motif in KGPDC and HPS further suggests that their active sites also are located at the polypeptide interface in a dimer of antiparallel $(\beta/\alpha)_8$-barrels.

Herein we report structural studies of the KGPDC encoded by the *ulaD* gene from *E. coli*. KGPDC is a dimer of antiparallel $(\beta/\alpha)_8$-barrels in which the active sites are located at the polypeptide interface. Carboxylate groups located at the ends of the first, second, and third β-strands function as ligands for the essential Mg^{2+}. A substrate analogue is coordinated to the Mg^{2+} via its 3-OH group, consistent with a mechanism that involves stabilization of an enediolate anion intermediate. The details of the dimer interface are conserved with those in OMPDC, providing unequivocal evidence that KGPDC and OMPDC are derived from a common ancestor.

The group of homologous enzymes that includes KGPDC, HPS, and OMPDC is designated the OMPDC "suprafamily" to emphasize the absence of any discernible common strategy for forming and/or stabilizing chemically distinct unstable intermediates and/or transition states on the reaction coordinates (*6*). This is the first group of functionally distinct enzymes to be described that retains both sufficient sequence and structural homology to confirm that the members are related by divergent evolution. This study demonstrates that sequence or structural homology alone cannot be used reliably to infer functions of new proteins discovered in genome projects.

Table 1: Data Collection and Refinement Statistics

Data Collection Statistics					
	Se-Met KGPDC·phosphate			native KGPDC·phosphate	native KGPDC·L-gulonate 6-P
wavelength (Å)	0.97945 (peak)	0.97931 (edge)	1.0247 (remote)	0.97625	1.514
source	APS 19-ID	APS 19-ID	APS 19-ID	APS 19-BM	Cu K_
resolution (Å)	2.3	2.3	2.3	1.62	2.2
unique reflections	20955	20925	20965	58810	22847
total reflections	195027	194286	185534	480987	56895
completeness[a] (%)	99.7 (99.7)	99.6 (99.7)	98.0 (84.3)	97.2 (90.5)	91.2 (77.9)
average I/σ[a]	33.7 (23.4)	33.1 (23.4)	33.7 (23.7)	39.8 (5.8)	5.7 (3.4)
R-merge[a,b]	0.058 (0.098)	0.066 (0.104)	0.046 (0.079)	0.043 (0.181)	0.030 (0.099)

Refinement Statistics					
	KGPDC·phosphate	KGPDC·L-gulonate 6-P		KGPDC·phosphate	KGPDC·L-gulonate 6-P
resolution limits (Å)	500−1.62	30−2.10	no. of protein atoms	3273	3273
R-factor (%)[c]	18.2	18.0	no. of solvent atoms	417	231
R-free (%)[d]	19.9	21.2	average *B* factor (Å^2)	16.8	15.9
no. of reflections (working set)	54215	20446	rms bond lengths (Å)	0.004	0.006
no. of reflections (test set)	2870	1053	rms bond angles (deg)	1.303	1.200

[a] Numbers in parentheses represent completeness in highest resolution shell (KGPDC·phosphate, 1.68−1.62 Å; KGPDC·L-gulonate 6-P, 2.28−2.2 Å). [b] *R*-merge = $(\Sigma|I_{HKL} - I|/(\Sigma I_{HKL})$ where the average intensity I is taken over all symmetry equivalent measurements and I_{HKL} is the measured intensity for a given reflection. [c] $R = \Sigma||F_o| - k|F_c||/\Sigma|F_o|$. [d] *R*-free; *R*-factor for 5% of the data excluded from the refinement.

MATERIALS AND METHODS

Protein Purification. KGPDC, the product of the *ulaD* gene in *E. coli*, was expressed and purified with an N-terminal His-tag (*30*). The His-tag was removed with thrombin, leaving a G-S-H sequence appended to the N-terminal of the wild-type protein. Selenomethionine-labeled (SeMet) protein was similarly expressed in BL21-(DE3) cells with the following modifications: transformed cells were grown at 37 °C in minimal media [75 mM K_2HPO_4, 11 mM KH_2PO_4, 7.5 mM $(NH_4)_2SO_4$, 8 mM $MgSO_4$, 70 μM $CaCl_2$, 0.2 μM $FeSO_4$, 0.15 mM thiamin], supplemented with 50 mM glucose and 100 μg/mL ampicillin, to an $OD_{600} = 0.6$. Methionine-suppression solution (0.5 mM L-lysine, 0.7 mM L-threonine, 0.6 mM L-phenylalanine, 0.4 mM L-leucine, 0.4 mM L-isoleucine, 0.4 mM L-valine) was then added to the culture, and the cells were incubated for an additional 30 min at 37 °C. Selenomethionine (0.25 mM) and 0.5 mM IPTG were then added to the culture, and the cells were incubated for 16 h at 37 °C. Selenomethionine-labeled KGPDC was similarly purified as described before (*30*).

Crystallization. KGPDC was dialyzed against 50 mM HEPES, pH 7.5, 5 mM $MgCl_2$, 100 mM NaCl, concentrated to 15 mg/mL, frozen as small pellets in liquid nitrogen, and stored at −80 °C. Crystals were grown at room temperature by micro-batch experiments by combining 10 μL of protein solution and 10 μL of a solution containing 18% MePEG 5000, 50 mM NaH_2PO_4, 50 mM K_2HPO_4, and 50 mM Bis-Tris propane, pH 7.0. Crystals appeared in a few days after micro-seeding, grew to dimensions of 100 × 100 × 700 μm after approximately 1 week, and diffracted to beyond 2.0 Å resolution. They belong to space group *C2* ($a = 42$ Å, $b = 121$ Å, $c = 92$ Å, and $\beta = 97°$) with a dimer in the asymmetric unit. The SeMet-substituted protein crystallized under identical conditions. Crystals of KGPDC complexed with the inhibitor L-gulonate 6-phosphate [prepared from L-gulonate and ATP with the enzyme encoded by the *lyxK* gene from *E. coli* (*30*)] were grown similarly except that 20 mM inhibitor replaced the sodium/potassium phosphate.

Data Collection and Structure Determination of KGPDC with Bound Phosphate. Crystals grown in the presence of sodium/potassium phosphate were transferred to a cryo-protecting solution containing 20% MePEG 5000, 100 mM sodium/potassium phosphate, 100 mM PIPES, pH 7.0, 20% glycerol, and 200 mM NaCl. Crystals remained in this cryo-protecting solution for approximately 1 min and then were flash-frozen in liquid nitrogen.

A native data set was recorded at beamline 19BM of the Structural Biology Center-CAT of APS with a single 180° scan with 1° oscillations at 3 s per frame. The data were integrated and scaled with HKL2000 (*31*) (Table 1). The data from the SeMet-labeled crystal were collected on beamline 19ID utilizing a single 180° scan composed of 1° oscillations for each of three wavelengths (Table 1).

The structure was solved by multiple anomalous dispersion (MAD) with the program SOLVE (*32*) where 9 out of 10 selenium atoms were identified. These heavy atom sites were used to calculate initial phases which were refined by solvent flattening with the program DM in the CCP4 package (*33, 34*). Further solvent flattening with the program RESOLVE generated an interpretable map (*35*).

An initial model for residues Leu^3−Trp^{215} of one monomer of KGPDC was built into the experimental density map from the SeMet data set with the program TURBO FRODO (*36*). The second subunit in the asymmetric unit was generated by applying the noncrystallographic symmetry transformation. Subsequent refinement used the native data set. The model was improved by alternating rounds of model building and refinement with the program CNS (*37*). The final *R*-factor was 18.2% with an *R*-free of 19.8%. A Ramachandran plot indicates that 94.2% of the non-glycine and non-proline residues fall in the most favored regions with the remaining 5.8% falling in the additionally allowed regions. Final refinement statistics are given in Table 1.

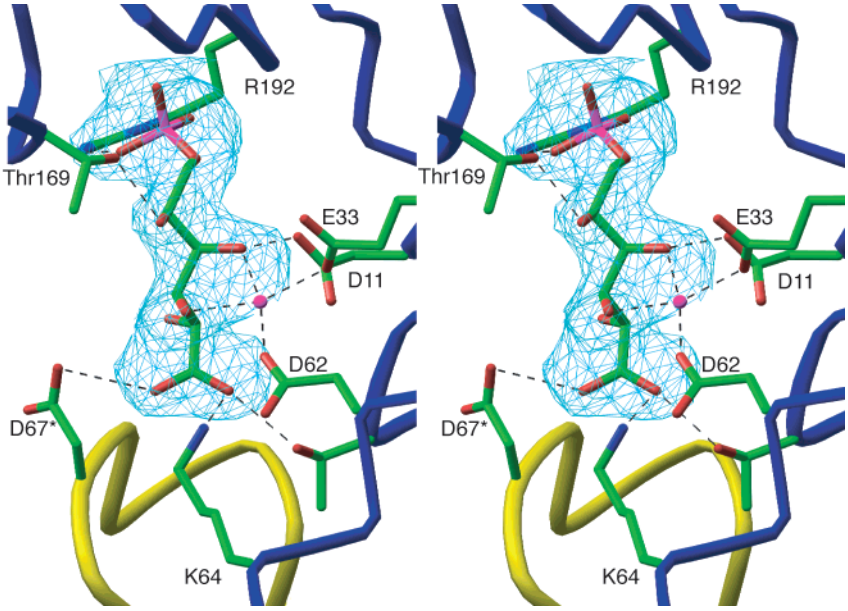

FIGURE 3: Stereoview of the electron density for L-gulonate 6-phosphate. This reveals the extensive hydrogen bonding network between the polar functional groups of the inhibitor and the enzyme. The Asp^{67} side chain, which is marked with an asterisk attached to a loop depicted in yellow, is contributed by the second subunit in the dimer. The map was calculated with coefficients of the form F_o-F_c where the inhibitor was omitted from the phase calculation. Figures 3−5 were prepared with the program Ribbons (*42*).

Structure Determination of KGPDC Complexed with L*-Gulonate 6-Phosphate.* Crystals grown in the presence of L-gulonate 6-phosphate were mounted in a glass capillary, and diffraction data were collected with a Siemens Hi-Star area detector at 4 °C. Data processing and scaling were done using the programs XDS and Xscalibre (*38*). The complex was solved by molecular replacement utilizing the structure of the KGPDC·Mg^{2+}·phosphate complex as a search model without the ions or water with the program CNS (*37*). Thereafter, the model was built and refined with the programs TURBO−Frodo and CNS. The final *R*-factor was 17.6% with a final *R*-free of 20.3% (Table 1). A Ramachandran plot indicates that 94.2% of the non-glycine and non-proline residues fall in the most favored regions with the remaining 5.8% falling in the additionally allowed regions. A representative section of electron density for the inhibitor is shown in Figure 3.

RESULTS AND DISCUSSION

Structure of KGPDC·Mg^{2+}·Phosphate. The asymmetric unit contains 2 subunits of KGPDC which both exhibit continuous electron density from Leu^3 to Trp^{215} where the genome-encoded KGPDC is 216 amino acid residues in length. The N-terminal extension derived from the His-tag is disordered. As anticipated, the two subunits of KGPDC exist as a dimer of antiparallel $(\beta/\alpha)_8$-barrels (Figure 4) in which the monomers are related by a 2-fold rotation axis. The monomers are virtually identical in conformation, with an rms deviation of 0.9 Å between equivalent α-carbons in each monomer. The dimer interface is formed by extensive contacts between the C-terminal ends of the third, fourth, and fifth β-strands; the buried surface area is approximately 8400 $Å^2$. A disulfide bond between Cys^{89} from each monomer was observed in the Se-Met structure used for MAD phasing but was reduced in the structure of the complex with L-gulonate 6-phosphate (next section).

The active sites, as indicated by the location of a phosphate ion in each subunit and by analogy to OMPDC, lie at the dimer interface. Each is built with residues primarily from one monomer, with the exception of the strictly conserved Asp^{67} which is contributed by the second subunit. A Mg^{2+} with octahedral coordination is liganded by the side chains of Glu^{33} and Asp^{62} and four ordered water molecules. The phosphate ion is coordinated by contacts with the backbone of Gly^{171}, Gly^{191}, and Arg^{192} and with the side chain of Arg^{192}.

*Structure of KGPDC·Mg^{2+}·*L*-Gulonate 6-Phosphate.* The structure of the KGPDC·Mg^{2+}·L-gulonate 6-phosphate complex is similar to that of the KGPDC·Mg^{2+}·phosphate complex except for small changes in the positions of active site residues. The Mg^{2+} is coordinated to the hydroxyl groups on C3 and C4 of the inhibitor, the carboxylate groups of Glu^{33} and Asp^{62}, and two ordered water molecules. These water molecules are, in turn, coordinated by the side chains of Asp^{11} and Thr^{36}.

The location of the phosphate moiety of the inhibitor in the KGPDC·Mg^{2+}·L-gulonate 6-phosphate complex is identical to that observed for the inorganic phosphate ion in the KGPDC·Mg^{2+}·phosphate complex. Extensive hydrogen-bonding contacts are formed between the polar functional groups of the inhibitor and the enzyme (Figure 3). The side chains of Asp^{11}, Thr^{36}, and Thr^{169} make hydrogen bonding contacts with the hydroxyl groups on the inhibitor.

Mechanism of the KGPDC-Catalyzed Reaction. The reaction catalyzed by KGPDC is dependent on the presence of Mg^{2+} (*30*). However, the presence of Lys^{64} in the active site could suggest a Schiff-base mechanism analogous to that used by acetoacetate decarboxylase (*39*). In unpublished experiments, we have determined that the value of k_{cat} for the K64A mutant is 20% that measured for wild-type enzyme, eliminating this mechanistic possibility (W. S. Yew and J. A. Gerlt). Furthermore, we also have observed that wild-type KGPDC catalyzes the exchange of the 1-*pro-S* hydrogen of the L-xylulose 5-phosphate product with solvent.

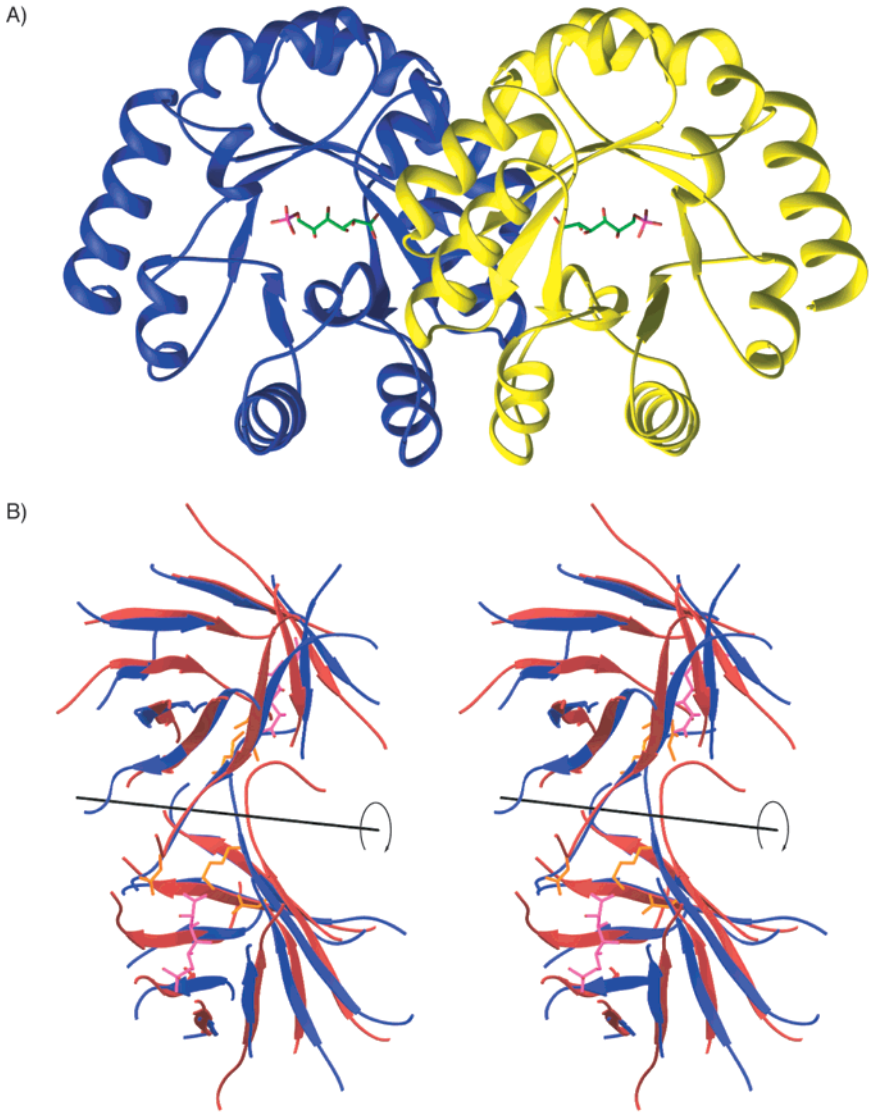

FIGURE 4: Ribbon representation of the KGPDC·L-gulonate 6-phosphate complex (A) and stereo-overlap between the KGPDC·L-gulonate 6-phosphate complex and OMPDC from *Bacillus subtilis* (B). The ribbon representation of the KGPDC·L-gulonate 6-phosphate complex reveals that the active sites, as indicated by the presence of the inhibitor, lie at the interface between the two subunits. In the overlap between KGPDC (shown in blue) and OMPDC (shown in red), only the β-strands and loops that form the interface between the two barrels are shown for clarity. Conserved active site residues from KGPDC and the L-gulonate 6-phosphate inhibitor are shown to indicate the position of the active site. The figure was prepared with the coordinates for the OMPDC from *B. subtilis* [RSCB accession number 1DBT (*20*)]. The superposition was done with the program ALIGN (*43*) modified at the University of Wisconsin by Dr. Gary Wesenberg to allow selection of the segments used for the alignment (available on request).

Taken together, the available structural and functional data provide compelling evidence for a mechanism involving a Mg^{2+}-stabilized enediolate anion intermediate (Figure 1).

The carboxylate group of L-gulonate 6-phosphate is hydrogen bonded to Lys^{64} and Asp^{67} of the second subunit of the dimer. Significantly, the observed conformation of the substrate analogue is not compatible with a productive enzyme·substrate complex, in which the carboxylate group must be orthogonal to the plane of the forming enediolate anion. A productive complex could involve coordination of the substrate to the Mg^{2+} via its 3-keto group and 2-hydroxyl group, not its 4-hydroxyl group. When such coordination is modeled, the carboxylate group is forced to be orthogonal to the plane of the forming enediolate anion and spatially proximal to His^{136}, located at the end of the fifth β-strand, and Asp^{67}, contributed by the second subunit; Lys^{64} is also proximal to the substrate, although not positioned to interact with the carboxylate group. These residues are candidates for the general acid catalyst that protonates the enediolate anion intermediate and likely catalyzes the observed exchange of the 1-*pro-S* hydrogen of the L-xylulose 5-phosphate product with solvent.

Determination of the stereochemical course of the decarboxylation reaction and detailed analyses of the reactions catalyzed by active site mutants are in progress.

Comparison of KGPDC and OMPDC. The structure of KGPDC is remarkably similar to the available structures of OMPDC despite low pairwise sequence identities (Figure 2), confirming that KGPDC and OMPDC are related by divergent evolution. Both KGPDC and OMPDC are dimers of $(\beta/\alpha)_8$-barrels arranged in an antiparallel fashion. The folds of the KGPDC and OMPDC monomers are essentially identical, with the flexible loops involved with substrate binding being the major differences. The rms deviation between α-carbons of OMPDC from *Bacillus subtilis* and KGPDC is 1.5 Å for 154 residue pairs.

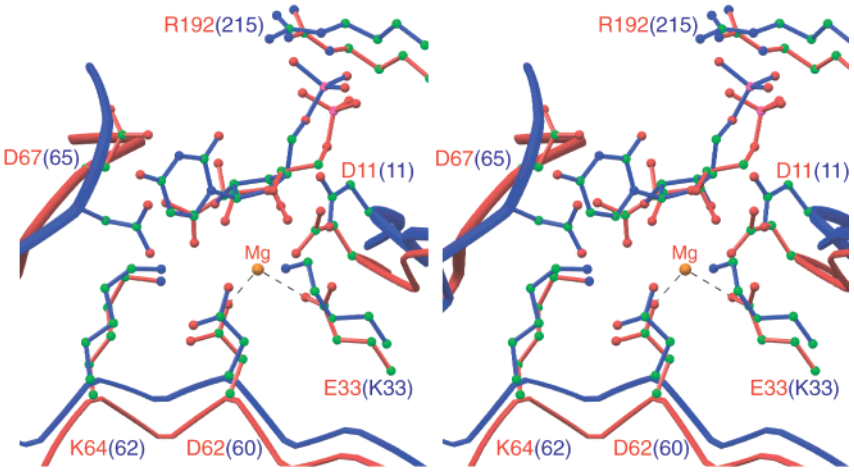

FIGURE 5: Close-up stereoview of the active sites of KGPDC·L-gulonate 6-phosphate complex and OMPDC·UMP. The side chains from conserved residues from KGPDC (shown in red) overlay almost exactly with those from OMPDC (shown in blue). The inhibitors, L-gulonate 6-phosphate (shown in red) and UMP (shown in blue), occupy nearly identical positions in the active sites. The phosphate groups of the inhibitors are bound analogously. The figure was prepared with the coordinates for the OMPDC from *B. subtilis* [RSCB accession number 1DBT (*20*)]. The superposition was done with the program ALIGN (*43*).

The similarity in the arrangement of the monomers within the dimer is perhaps more striking than the conservation in the structures of the monomers (Figure 4B). Superposition of the α-carbons of the dimers of KGPDC and of OMPDC from *B. subtilis* (18% sequence identity) yields an rms deviation of 1.7 Å for 237 residue pairs. Both the three-dimensional arrangements of the monomers relative to each other and the interactions in the dimer interface are almost completely conserved, further supporting the assertion that KGPDC and OMPDC are the products of divergent evolution.

An overlay of the active site of KGPDC with that of OMPDC shows that conserved residues occupy nearly identical positions (Figure 5). However, the mechanisms of the reactions catalyzed by KGPDC and OMPDC are distinct (Figure 1), so the conserved functional groups are used in different ways. The most striking differences in the use of side chains involve the ligands for the Mg^{2+} in KGPDC. In OMPDC, the side chain of Asp^{60} likely is positioned proximal to the carboxylate group of OMP (*20*, *22*), with the resulting repulsive interactions destabilizing the ground state. In KGPDC, the side chain of the homologous Asp^{62}, which occupies a nearly identical position to that in OMPDC, is coordinated to the Mg^{2+} and does not interact with the carboxyl leaving group of L-gulonate 6-phosphate. In OMPDC, the side chain of the conserved Lys^{33} participates in a hydrogen bonding network that facilitates a mechanism that must avoid a vinyl anion intermediate that is too unstable to be kinetically competent. In KGPDC, the structurally homologous Glu^{33} is coordinated to the Mg^{2+}. Accordingly, residues that mediate catalysis in OMPDC by direct interactions with the substrate have been conscripted to serve as ligands for the essential Mg^{2+} in KGPDC. The Mg^{2+}, in turn, interacts with the substrate to stabilize an enediolate anion intermediate.

The only mechanistic similarity supported by conserved amino acids might be the identity of the general acid catalyst that delivers a solvent-derived proton to the product. In OMPDC, Lys^{62} is proposed to donate a proton to C6 of OMP after decarboxylation. In KGPDC, the homologous Lys^{64} may function as the catalytic acid in protonation of the enediolate anion. However, these proton transfer reactions are highly exergonic, so this possible shared function may not be critical for understanding the structural strategies for divergent evolution. Instead, the distinct mechanisms for stabilizing (KGPDC) or avoiding (OMPDC) highly unstable intermediates appear to be the essential criterion for discriminating between the reactions catalyzed by these homologous active sites and specifying the underlying nature of the diversity of function.

Like OMPDC, KGPDC makes extensive contacts with its substrate through a network of hydrogen bonds in the active site. The side chains of Asp^{11}, Thr^{36}, and Thr^{169} make hydrogen bonding contacts with the hydroxyl groups on the acyclic inhibitor. Homologous, but not strictly conserved, residues in OMPDC make similar contacts with the hydroxyl groups of the furanoside ring in OMP.

The position of the inhibitor in the active site of KGPDC is very similar to that of substrate analogues or product in the active site of OMPDC (Figure 5). In both enzymes, the phosphate group of the substrate is bound by a conserved phosphate binding site. In OMPDC, hydrogen bonds are formed with the backbone amide hydrogens from Gly^{214} and Arg^{215} and the side chain of Arg^{215}; in KGPDC, hydrogen bonds are formed with the backbone amide hydrogens from Gly^{191} and Arg^{192} and the side chain of Arg^{192}. This binding site, remote from the position of chemistry, is the single active site functional feature that has been retained in the course of divergent evolution. Perhaps these conserved interactions are essential for mediating conformational changes that establish productive orientations of the carboxylate groups of the substrates with active site residues.

Relationship of KGPDC, HPS, and OMPDC. No structural information is yet available for HPS, although the sequence similarities between HPS and OMPDC have previously been noted (*40*). However, based on the sequence alignment (Figure 2), it is predicted that the both the quaternary structure and active site of HPS will be remarkably similar to those determined for KGPDC. In particular, it is expected that HPS will be a dimer of $(\beta/\alpha)_8$-barrels in which the active sites are located at the interface, and that the Mg^{2+} reported to be essential for activity in the HPS-catalyzed reaction will

be coordinated to an Asp residue at the end of the first β-strand, a Glu residue at the end of the second β-strand, and the first Asp of the Asp-x-Lys-x-x-Asp motif at the end of the third β-strand. Furthermore, it is predicted that the D-ribulose 5-phosphate substrate will coordinate to the Mg^{2+} via its 2-carbonyl group, thereby allowing stabilization of an enediolate anion intermediate (Figure 1).

The sequences of the KGPDC's and HPS's that can be identified in the databases contain homologues of His^{136} in KGPDC (highlighted with green in Figure 2), suggesting that this residue will mediate a shared partial reaction in both reactions. However, all KGPDC's contain homologues of Glu^{112} located at the end of the fifth β-stand whereas all HPS's are predicted to contain an Asp at this position (highlighted with magenta in Figure 2); the differing functional groups likely are involved in the divergent partial reactions catalyzed by KGPDC and HPS.

The mechanisms of the reactions catalyzed by KGPDC and HPS do not share any common strategy for forming and/or stabilizing chemically distinct intermediates and transition states with the reaction catalyzed by OMPDC (Figure 1). However, they do share a common partial reaction [formation and stabilization of enediolate anion intermediate via coordination to Mg^{2+} that either follows (KGPDC) or precedes (HPS) carbon−carbon bond breaking/forming reactions] and, therefore, constitute a mechanistically diverse superfamily within the OMPDC suprafamily.

Concluding Remarks. Comparison of the structures of KGPDC and OMPDC shows that Nature's strategies for divergent evolution are not restricted to conservation of either chemical mechanism or substrate specificity, attributes that might assist relating proteins of unknown or uncertain function to those of known function. Instead, evidence is presented here for a strategy in which evolution can be opportunistic so that any structural feature of an active site can be conscripted for the development of new enzymatic functions.

The functional plasticity of the $(\beta/\alpha)_8$-barrel fold suggests that many, if not all, $(\beta/\alpha)_8$-barrel enzymes may be derived from a limited number of $(\beta/\alpha)_8$-barrel progenitors (*4*, *41*). In the course of divergent evolution, discernible sequence similarity may vanish, obscuring the mutational events that lead to significantly altered function. The present studies expose remarkable changes in function that can occur via divergent evolution in response to selective pressure. They suggest that it is inappropriate to conclude function based on sequence comparisons alone, since this makes the simplest assumption that conservation of active site residues correlates with conservation of their functional roles. This calls into question ongoing efforts to assign function to genomic sequences.

These structural studies of KGPDC expand the membership of the "ribulose phosphate binding" superfamily as defined by the SCOP database (*http://scop.mrc-lmb.cam.ac.uk/scop/*) and raise the question of whether other mechanistically distinct reactions might be catalyzed by this same arrangement of functional groups. Structural and mechanistic studies of other enzymes whose sequences are homologous to OMPDC, KGPDC, and HPS are in progress to address this question.

REFERENCES

1. Govindarajan, S., Recabarren, R., and Goldstein, R. A. (1999) *Proteins: Struct., Funct., Genet. 35*, 408−414.
2. Wolf, Y. I., Grishin, N. V., and Koonin, E. V. (2000) *J. Mol. Biol. 299*, 897−905.
3. Zhang, C., and DeLisi, C. (1998) *J. Mol. Biol. 284*, 1301−1305.
4. Copley, R. R., and Bork, P. (2000) *J. Mol. Biol. 303*, 627−641.
5. Todd, A. E., Orengo, C. A., and Thornton, J. M. (2001) *J. Mol. Biol. 307*, 1113−1143.
6. Gerlt, J. A., and Babbitt, P. C. (2001) *Annu. Rev. Biochem. 70*, 209−246.
7. Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L., and Gerlt, J. A. (1996) *Biochemistry 35*, 16489−16501.
8. Holm, L., and Sander, C. (1997) *Proteins: Struct., Funct., Genet. 28*, 72−82.
9. Holden, H. M., Benning, M. M., Haller, T., and Gerlt, J. A. (2001) *Acc. Chem. Res. 34*, 145−157.
10. Grogan, G., Roberts, G. A., Bougioukou, D., Turner, N. J., and Flitsch, S. L. (2001) *J. Biol. Chem. 276*, 12565−12572.
11. Baker, A. S., Ciocci, M. J., Metcalf, W. W., Kim, J., Babbitt, P. C., Wanner, B. L., Martin, B. M., and Dunaway-Mariano, D. (1998) *Biochemistry 37*, 9305−9315.
12. Koonin, E. V., and Tatusov, R. L. (1994) *J. Mol. Biol. 244*, 125−132.
13. Morais, M. C., Zhang, W., Baker, A. S., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2000) *Biochemistry 39*, 10385−10396.
14. Armstrong, R. N. (2000) *Biochemistry 39*, 13625−13632.
15. Bergdoll, M., Eltis, L. D., Cameron, A. D., Dumas, P., and Bolin, J. T. (1998) *Protein Sci. 7*, 1661−1670.
16. McCarthy, A. A., Baker, H. M., Shewry, S. C., Patchett, M. L., and Baker, E. N. (2001) *Structure (London) 9*, 637−646.
17. Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R., and Goodenough, P. W. (2001) *Trends Biochem. Sci. 26*, 740−746.
18. Hocker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A., and Sterner, R. (2001) *Nat. Struct. Biol. 8*, 32−36.
19. Radzicka, A., and Wolfenden, R. (1995) *Science 267*, 90−93.
20. Appleby, T. C., Kinsland, C., Begley, T. P., and Ealick, S. E. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 2005−2010.
21. Miller, B. G., Hassell, A. M., Wolfenden, R., Milburn, M. V., and Short, S. A. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 2011−2016.
22. Wu, N., Mo, Y., Gao, J., and Pai, E. F. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 2017−2022.
23. Harris, P., Navarro Poulsen, J. C., Jensen, K. F., and Larsen, S. (2000) *Biochemistry 39*, 4217−4224.
24. Kato, N., Ohashi, H., Tani, Y., and Ogata, K. (1978) *Biochim. Biophys. Acta 523*, 236−244.
25. Yanase, H., Ikeyama, K., Mitsui, R., Ra, S., Kita, K., Sakai, Y., and Kato, N. (1996) *FEMS Microbiol. Lett. 135*, 201−205.
26. Kato, N. (1990) *Methods Enzymol. 188*, 397−401.
27. Mitsui, R., Sakai, Y., Yasueda, H., and Kato, N. (2000) *J. Bacteriol. 182*, 944−948.
28. Yasueda, H., Kawahara, Y., and Sugimoto, S. (1999) *J. Bacteriol. 181*, 7154−7160.
29. Kemp, M. B. (1974) *Biochem. J. 139*, 129−134.
30. Yew, W. S., and Gerlt, J. A. (2002) *J. Bacteriol. 184*, 302−306.
31. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C. W. J., Sweet, R. M., Abelson, J. N., and Simon, M. I., Eds.) pp 307−326, Academic Press, New York.
32. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D55*, 849−861.

33. CCP4 (1994) *Acta Crystallogr. D50*, 760−763.
34. Cowtan, K., and Main, P. (1998) *Acta Crystallogr. D54*, 487−493.
35. Terwilliger, T. C. (2000) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 56*, 965−972.
36. Roussel, A., and Cambillau, C. (1991) in *Silicon Graphics Geometry Partners Directory*, Silicon Graphics.
37. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.
38. Kabsch, W. (1988) *J. Appl. Crystallogr. 21*, 916−924.
39. Highbarger, L. A., Gerlt, J. A., and Kenyon, G. L. (1996) *Biochemistry 35*, 41−46.
40. Traut, T. W., and Temple, B. R. (2000) *J. Biol. Chem. 275*, 28675−28681.
41. Farber, G. K., and Petsko, G. A. (1990) *Trends Biochem. Sci. 15*, 228−234.
42. Carson, M. (1977) in *Methods in Enzymology*, p 493, Academic Press, New York.
43. Cohen, G. H. (1997) *J. Appl. Crystallogr. 30*, 1160−1161.

BI012174E