Evolution of Enzymatic Activities in the Enolase Superfamily: Characterization of the (*D*)-Glucarate/Galactarate Catabolic Pathway in *Escherichia coli*†

Brian K. Hubbard,‡ Marjan Koch,§ David R. J. Palmer,§ Patricia C. Babbitt,|| and John A. Gerlt*,‡,§

Departments of Biochemistry and Chemistry, University of Illinois at Urbana−Champaign, 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 genes encoding the enzymes in the (*D*)-glucarate/galactarate catabolic pathway have been identified in the *Escherichia coli* genome. These encode, in three transcriptional units, (*D*)-glucarate dehydratase (GlucD), galactarate dehydratase, 5-keto-4-deoxy-(*D*)-glucarate aldolase, tartronate semialdehyde reductase, a glycerate kinase that generates 2-phosphoglycerate as product, and two hexaric acid transporters. We also have identified a gene proximal to that encoding GlucD that encodes a protein that is 72% identical in primary sequence to GlucD (GlucD-*r*elated *p*rotein or GlucDRP). However, whereas GlucD catalyzes the efficient dehydration of both (*D*)-glucarate and (*L*)-idarate as well as their epimerization, GlucDRP is significantly impaired in both reactions. Perhaps GlucDRP is an example of gene duplication and evolution in progress in the *E. coli* chromosome.

Complete genome sequences currently are available for 12 prokaryotes [eight eubacteria [*Haemophilus influenzae* Rd (*1*), *Mycoplasma genitalium* (*2*), *Mycoplasma pneumoniae* (*3*), *Synechocystis* sp. strain PCC6803 (*4*), *Helicobacter pylori* (*5*), *Escherichia coli* K-12 (*6*), *Bacillus subtilis* (*7*), and *Borrelia burgdorferi* (*8*)] and four archaebacteria [*Methanococcus jannaschii* (*9*), *Methanobacterium thermoautotrophicum* ΔH (*10*), *Archaeglobus fulgidus* (*11*), and *Aquifex aeolicus* (*12*)]] and 1 eukaryote [*Saccharomyces cerevisiae* (*13*)]. Approximately 40% of the identified orfs have no known function: their primary sequences are either unrelated or sufficiently dissimilar to proteins of known function to allow prediction of function. A major problem in biochemistry/enzymology is to develop an efficient strategy for assigning functions to these proteins.

Many enzymes necessarily belong to either families (>50% sequence identity; the members likely catalyze the same reaction on the same or similar substrates) or superfamilies (<50% sequence identity; the members likely catalyze the same *or different* overall reactions). In at least several superfamilies whose distantly related members catalyze different overall reactions, a common partial reaction is used to accomplish the unfavorable formation of an intermediate whose fate depends upon the identity of the overall reaction (*14*). Accordingly, we expect that significant restrictions on the functions of many unknown orfs can be achieved by (1) relating their sequences to distantly related proteins and (2) categorizing the range of chemistries that can be catalyzed by a given superfamily.

For enzymes in microbial pathways encoded by operons, assignment of a protein of unknown function to a superfamily may allow the function to be specified if the reactions or chemistries catalyzed by the other genes in the operon can be specified. This laboratory has used this approach in identifying (*D*)-galactonate dehydratase in *Escherichia coli* as a member of the enolase superfamily (*15*).

In this paper, we use these strategies to define the catabolic pathway for (*D*)-glucarate and galactarate[1] in *E. coli*. Early studies of this pathway (Figure 1) by Blumenthal and Fish (*16*) reported that (*D*)-glucarate is dehydrated to 5-keto-4-deoxy-(*D*)-glucarate (5-KDGluc)[2] by a (*D*)-glucarate dehydratase (GlucD). 5-KDGluc is converted by 5-KDGluc aldolase (KDGlucA) to tartronate semialdehyde and pyruvate; tartronate semialdehyde is reduced to (*D*)-glycerate by a tartronate semialdehyde reductase (TSAR); and, finally, (*D*)-glycerate was reported to be phosphorylated by a glycerate kinase (GK) to yield 3-phosphoglycerate (3-PGA). So, pyruvate and 3-PGA were reported to be the final products of the pathway. In addition, the early studies revealed that galactarate catabolism also was induced by (*D*)-glucarate. Since GlucD was reported not to utilize the 4-epimeric galactarate as substrate, a different dehydratase [galactarate dehydratase GalcD)] was implicated in the catabolism of galactarate. Presumably, GalcD also would catalyze the formation of 5-KDGluc (Figure 1), since this would allow the same aldolase, reductase, and kinase used for (*D*)-glucarate catabolism to complete the galactarate catabolic pathway. However, none of the enzymes constitut-

† This research was supported by Grants GM-40570 and GM-52594 from the National Institutes of Health.

* Address correspondence to this author (e-mail: j-gerlt@uiuc.edu).
‡ Department of Chemistry, University of Illinois.
§ Department of Biochemistry, University of Illinois.
|| Department of Pharmaceutical Chemistry, University of California.

[1] Galactarate is achiral because it is a meso compound.

[2] Abbreviations: GalcD, galactarate dehydratase; GlucD, (*D*)-glucarate dehydratase; GK, (*D*)-glycerate kinase; 5-KDGluc, 3-deoxy-(*L*)-*threo*-2-hexulosarate or 5-keto-4-deoxy-(*D*)-glucarate; KDGlucA, 5-KDGluc aldolase; LDH, lactate dehydrogenase; orf, open reading frame; TSAR, tartronate semialdehyde reductase.

S0006-2960(98)01124-6 CCC: $15.00
Published on Web 09/25/1998

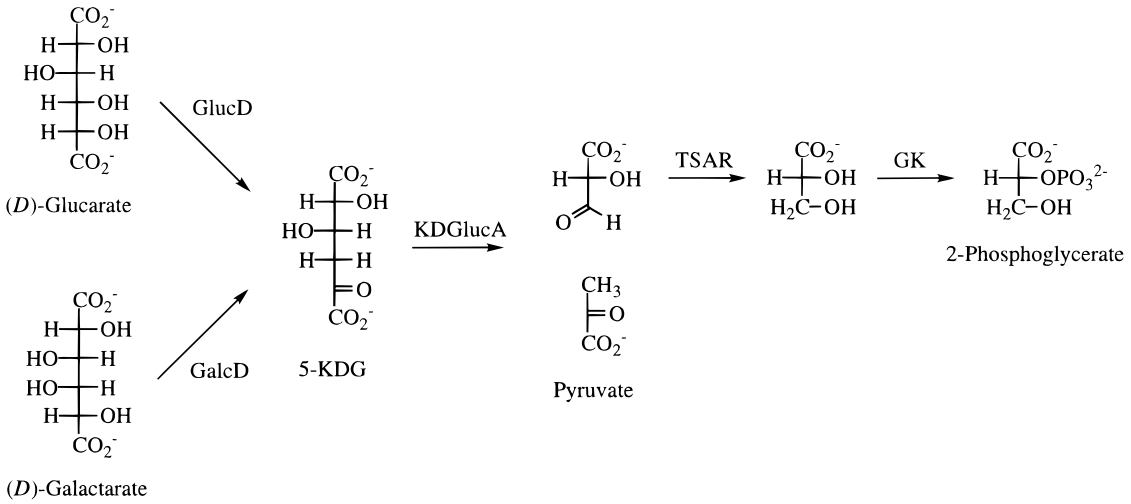


FIGURE 1: Pathway for (*D*)-glucarate/galactarate utilization in *E. coli*. In the original report of this pathway (*16*), 3-phosphoglycerate, rather than 2-phosphoglycerate, was reported to be the product of the GK-catalyzed reaction.

ing the pathway(s) for (*D*)-glucarate and galactarate catabolism were isolated to homogeneity and characterized.

We have used the primary sequence of the GlucD from *Pseudomonas putida* to "mine" the proteins encoded by the *E. coli* chromosome. We discovered two contiguous homologues of the GlucD, *only one of which* catalyzes the efficient dehydration of both (*D*)-glucarate and (*L*)-idarate to form 5-KDGluc as well as their epimerization. The aldolase responsible for the cleavage of 5-KDGluc (KDGlucA) also was purified from cell extracts, and its N-terminal sequence was used to locate two opposing transcriptional units that encode the remaining enzymes in the catabolism of both (*D*)-glucarate and galactarate. We expect that this characterization of the catabolic pathway will be instrumental in studies in which the catalytic functions of GlucD and its contiguous homologue are altered by directed evolution.

MATERIALS AND METHODS

All restriction enzymes were purchased from Gibco BRL. T4 DNA ligase and T7 polynucleotide kinase were purchased from New England BioLabs. Oligonucleotide primers were synthesized by Bio-Synthesis (Lewisville, TX), and DNA sequencing was performed by Ana-Gen Technologies Inc. (Palo Alto, CA) and the Genetic Engineering Facility, University of Illinois Biotechnology Center. N−Terminal sequencing of proteins was performed by the Genetic Engineering Facility. All other reagents were the highest quality grade commercially available and purchased from Sigma/Aldrich.

^{1}H NMR spectra were recorded with a Varian Unity INOVA 500 MHz NMR spectrometer. A Bio-Pilot FPLC (Pharmacia Biotech) was used for purification of proteins and acid sugars. Resins for column chromatographies were purchased from either Bio-Rad or Pharmacia Biotech. Spectrophotometric assays were performed with a Perkin-Elmer Lambda 14 spectrometer. One unit of enzyme activity is the amount that converts 1 μmol of substrate to product per minute.

Isolation of GlucD and KDGlucA. *E. coli* strain MG1655 was grown in 10 L of minimal media with (*D*)-glucarate as the sole carbon source according to Blumenthal and Fish (*16*). Cells were harvested by centrifugation (5000 rpm at 4 °C) and lysed by sonication at 80 W power for 10 cycles of 1 min duration. Protease inhibitor "Complete pellets" (Boehringer Mannheim) were added to the cell lysate. The lysate was fractionated with a DEAE-Sephacel (Pharmacia) ion-exchange column (2.6 cm × 70 cm) using a linear gradient (1 L) from 0 to 1 M NaCl in 10 mM Tris-HCl buffer, pH 7.8, containing 5 mM $MgCl_2$; 10 mL fractions were collected. Each fraction was assayed for both GlucD and 5-KDGluc aldolase (KDGlucA) activities. Those fractions that contained the desired activity were combined, concentrated to a volume of <2 mL, and further fractionated with a Superdex 200 (Pharmacia) gel filtration column (16 cm × 100 cm) eluted with 10 mM Tris-HCl buffer, pH 7.8, containing 5 mM $MgCl_2$; 10 mL fractions were collected. Fractions containing the desired activity were purified to homogeneity with a Resource Q (Pharmacia) column (16 mm × 30 mm) using a linear gradient (1 L) from 0 to 1 M NaCl in 10 mM Tris-HCl buffer, pH 7.8, containing 5 mM $MgCl_2$; 10 mL fractions were collected. The resulting proteins were assayed, and the N-terminal sequences were determined using an automated Edman sequencer (Applied Biosystems Model 477A).

Cloning, Expression, and Purification of Catabolic Enzymes. A *Stu*I restriction site was introduced into the pET16b expression vector (Invitrogen) following the factor Xa recognition/cleavage sequence using the Quick Change kit (Stratagene). The pET16b vector contains a bacteriophage T7 polymerase promoter followed by an initiation codon, an N-terminal histidine tag containing 10 histidine residues, and, finally, a factor Xa cleavage site. Each enzyme in the (*D*)-glucarate/galactarate catabolic pathway was purified to homogeneity by analogous procedures utilizing this expression vector.

Oligonucleotide primers containing the first and last 24 bases of each gene were used to PCR-amplify various genes from *E. coli* strain MG1655 genomic DNA using Pfu DNA polymerase (Stratagene); a protocol of 40 cycles of 95 °C denaturation, 60 °C annealing, and 72 °C extension was used for all genes. The blunted-ended PCR products were isolated using DNAzol (Molecular Research Center Inc.), 5′-phosphorylated with polynucleotide kinase, and ligated into the modified pET16b that had been linearized with *Stu*I. The

various proteins were expressed in *E. coli* strain BL21. Transformed cells were grown at 37 °C in LB to an OD at 600 nm = 2.0 and harvested by centrifugation (5000 rpm at 4 °C) in a Sorvall JA-10 rotor; IPTG was not added to the medium to induce the proteins. Each His-tagged protein was purified with Chelating Sepharose Fast Flow (Pharmacia) columns (16 cm × 40 cm). Each cell lysate was applied to the column in binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9), washed with wash buffer (60 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9), and eluted with 50% binding buffer and 50% strip buffer (100 mM EDTA, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9).

Assay of GlucD Activity. The formation of 5-KDGluc catalyzed by GlucD was assayed using semicarbazide to detect the α-keto acid product. For routine assays, an aliquot containing GlucD was added to 1 mL of 10 mM (*D*)-glucarate in 50 mM Tris-HCl buffer, pH 8.0, containing 5 mM $MgCl_2$, at 22 °C. The amount of 5-KDGluc formed after 0, 1, 2, and 3 min was quantitated by detection of its semicarbazone at 250 nm (*17*). Determinations of k_{cat} and K_m were performed by varying the concentration of (*D*)-glucarate or (*L*)-idarate.

Assay for 5-KDGluc Aldolase (KDGlucA) Activity. The pyruvate product of the KDGlucA-catalyzed reaction was assayed spectrophotometrically using lactate dehydrogenase (LDH) and NADH. For routine assays, an aliquot of KDGlucA was added to 1 mL of 10 mM 5-KDGluc in 75 mM potassium phosphate buffer, pH 7.5, containing 15 mM $MgSO_4$, 100 μM NADH, and 2 units of LDH (rabbit muscle) at 22 °C. The conversion of NADH to NAD^+ was monitored at 340 nm. Determinations of k_{cat} and K_m were performed by varying the concentration of 5-KDGluc.

Assay for Tartronate Semialdehyde Reductase (TSAR) Activity. Tartronate semialdehyde was prepared in situ by incubating 10 mM 5-KDGluc with 10 units of KDGlucA in 1 mL of 75 mM potassium phosphate buffer, pH 7.5, containing 15 mM $MgSO_4$ for 10 min at 22 °C. An aliquot of TSAR and NADH (100 μM) were added to the reaction mixture, and the TSAR-catalyzed conversion of NADH to NAD^+ was monitored at 340 nm.

Assay for Glycerate Kinase (GK) Activity. An aliquot of GK was added to 1 mL of 10 mM (*D*)-glycerate in 75 mM potassium phosphate buffer, pH 7.5, containing 15 mM $MgSO_4$, 100 μM NADH, 0.08 mM ATP, 0.08 mM PEP, 1 unit of pyruvate kinase, and 2 units of LDH at 22 °C. The conversion of NADH to NAD^+ was monitored at 340 nm. Determinations of k_{cat} and K_m were performed by varying the concentration of (*D*)-glycerate.

Assay for GalcD Activity. The formation of 5-KDGluc catalyzed by GalcD was assayed using semicarbazide to detect the α-keto acid product. For routine assays, an aliquot containing GalcD was added to 1 mL of 10 mM galactarate in 100 mM Tris-HCl buffer, pH 7.5, containing 170 mM β-mercaptoethanol and 0.5 mM $FeSO_4$, at 22 °C. The amount of 5-KDGluc formed after 0, 1, 2, and 3 min was quantitated by detection of its semicarbazone at 250 nm (*17*). Determinations of k_{cat} and K_m were performed by varying the concentration of galactarate.

^{1}H NMR Analyses. All ^{1}H NMR experiments were recorded with a Varian Unity INOVA 500 MHz NMR spectrometer with an inverse detection probe. Samples contained 100 mM substrates in 50 mM Tris-HCl buffer, pH 7.5. Spectra were acquired in 10% D_2O/90% H_2O using a presaturation pulse sequence (Varian).

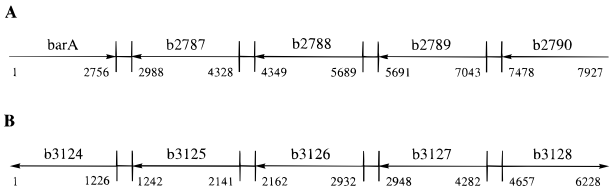


FIGURE 2: Panel A: Region of the *E. coli* chromosome that encodes GlucD (orf b2787), GlucDRP (b2788), and a (*D*)-glucarate transporter (b2789). Panel B: Region of the *E. coli* chromosome that encodes GK (b3124), TSAR (b3125), KDGlucA (b3126), (*D*)-glucarate/galactarate transporter (b3127), and GalcD (b3128). The numbers refer to the relative base pair numbers of the initiation and termination codons.

RESULTS AND DISCUSSION

Identification of an Operon Encoding GlucD and a Homologue in E. coli. The primary sequence of GlucD from *P. putida* (GenBank M69160) was used as the query sequence for a BLASTP search of the *E. coli* genome (*6*) using the NCBI BLAST server. Two contiguous open reading frames (orfs) having unassigned functions were identified; a diagram of this region of the chromosome is presented in Figure 2, panel A.

One of these orfs, designated b2787, contains 446 amino acids and has 80% sequence identity with the query GlucD sequence. The second orf, designated b2788, is transcribed in the same direction (counterclockwise on the chromosome, right to left in the figure), also contains 446 amino acids, and has 72% sequence identity with the query GlucD sequence. Orfs b2787 and b2788 have 72% sequence identity with one another. The initiation codon for b2787 is located 20 bp downstream of the termination codon for b2788, so these appear to be located in an operon. The termination codon of a third orf of unknown function transcribed in the same direction as b2787 and b2788, b2789, is located 2 bp upstream of the initiation codon for b2788, so this gene appears to be located in the same operon.

A BLASTP search using the sequence of the protein encoded by orf b2789 as query reveals that the protein is homologous to a large number of transporter proteins, including probable (*D*)-glucarate transporters in *Bacillus subtilis* and *P. putida*, a probable (*D*)-galactonate transporter in *E. coli*, and the known hexuronate transporter encoded by the *exu*T gene in *E. coli*. Thus, we expect that b2789 encodes a (*D*)-glucarate transporter in *E. coli*.

Proceeding clockwise on the chromosome (left to right in the figure), the next orf, b2790, encodes a protein of unknown function and is transcribed in the same direction as the b2789-b2788-b2787 operon, but its termination codon is 435 bp upstream of the initiation codon for b2789, suggesting that it is not cotranscribed with the operon. Proceeding counterclockwise on the chromosome, the next orf encodes a known protein, barA (b2786), and is transcribed in the opposite direction. So, we deduce that the transcriptional unit containing the two homologues of the GlucD from *P. putida* contains three genes. Genes that likely encode KDGlucA, TSAR, GK, and GalcD are neither contained in this operon nor appear to be proximal on the chromosome.

Identification of GlucD. The identification of two homologues of the GlucD from *P. putida* was unexpected. Accordingly, we purified the GlucD activity from *E. coli* strain MG1655 grown on (*D*)-glucarate as the sole carbon source. A single GlucD activity was detected in cell extracts and purified to homogeneity in an overall yield of 70%. The N-terminal sequence of the purified GlucD was determined to be X-S-Q-F-T-T-P-V-V-T-E-M-Q-V, where X represents an amino acid that could not be identified unequivocally. The predicted N-terminal sequence of b2787 is M-S-S-Q-F-T-T-P-V-V-T-E-M-Q-V and that of b2788 is M-A-T-Q-S-S-P-V-I-T-D-M-K-V-I. So, the single detectable GlucD is encoded by b2787. The enzymatic functions of b2788 or GlucDRP (GlucD-related protein) are described in a following section.

Identification of KDGlucA and the Remainder of the Catabolic Genes. The sequence of a known diacid sugar aldolase from *E. coli* strain C, 2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase, was used as the query sequence for a BLASTP search for the *E. coli* genome (GenBank Z47799). Two sequences were identified as likely Schiff base-dependent aldolases that utilize diacid sugars as substrates, b2245 (57% sequence identity), encoding a protein containing 267 amino acids, and b3126 (49% sequence identity), encoding a protein containing 256 amino acids. The function of neither orf was assigned previously.

We purified the KDGlucA activity from *E. coli* strain MG1655 grown on (*D*)-glucarate as the sole carbon source. A single KDGlucA activity was detected in cell extracts and purified to homogeneity in an overall yield of 55%. The N-terminal sequence of the purified KDGlucA was determined to be M-N-N-D-V-F-P-N-K-F-K-A-A-L-A-A, identical to the N-terminal sequence predicted for the protein encoded by b3126. The N-terminal sequence encoded by orf b2245 is M-N-A-L-L-S-N-P-F-K-E-R-L-R-K-G.

Orf b3126 (KDGlucA) is located within a cluster of four genes, b3127, b3126, b3125, and b3124, in the direction of transcription (counterclockwise on the chromosome), where the intergenic spacings, 16, 21, and 16 bp, respectively, are consistent with their being located within an operon (Figure 2, panel B). Another orf, b3128, is transcribed divergently (clockwise) from the operon.

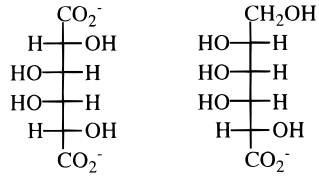
BLASTP searches were performed using the sequences of b3124, b3125, b3127, and b3128 as queries to determine whether these proteins together with b3127 (KDGlucA) could be the remaining enzymes in the putative (*D*)-glucarate/galactarate catabolic pathway (Figure 1).

Orf b3127 (444 amino acids) was found to be homologous to a large number of transporter proteins, including b2789 (54% sequence identity) that is located in the same operon as the GlucD we identified (previous section). On this basis, we hypothesized that b3127 also encodes a (*D*)-glucarate (or galactarate) transporter.

Orf b3125 (299 amino acids) was found to be homologous to 3-hydroxybutyrate (~30% sequence identity) and 6-phosphogluconate (~20% sequence identity) dehydrogenases, enzymes that catalyze the reduction/oxidation of β-carbonyl/hydroxyl groups of carboxylic acids. On this basis, we hypothesized that b3125 encodes TSAR.

Orf b3124 (408 amino acids) was found to be homologous (ranging from 33% to 55% sequence identity) to several orfs encoding proteins of unknown functions in several prokaryotes. However, a PROSITE analysis revealed that b3124 contains a consensus ATP/GTP-binding site motif (GISGDGKT in residues 105−112). Therefore, assuming that b3127-b3126-b3125-b3124 encode enzymes in the (*D*)-glucarate/galactarate catabolic pathway, by process of elimination we hypothesized that b3124 encodes GK, the final enzyme in the pathway described by Blumenthal and Fish (*16*).

Finally, orf b3128 (523 amino acids) was found to be homologous to (*D*)-altronate dehydratase from *E. coli* (UxaA, 32% sequence identity) and several prokaryotic orfs encoding proteins of unknown function. One of these is a protein located in a cluster of genes in the *B. subtilis* chromosome that appears to encode enzymes in a catabolic pathway for (*D*)-glucarate that yields α-ketoglutarate as product. These enzymes include a homologue of the *E. coli* and *P. putida* GlucDs, a 5-KDGluc dehydratase/decarboxylase that converts 5-KDGluc to 2,5-dioxopentanoate, and a dehydrogenase that oxidizes 2,5-dioxopentanoate to α-ketoglutarate. The absolute configurations of carbons-2, -3, and -4 of (*D*)-altronate and of carbons-5, -4, and -3 of galactarate are identical, respectively, a plausible situation if (*D*)-altronate dehydratase and GalcD are related by divergent evolution.



On these bases, we hypothesized that b3128 encodes galactarate dehydratase.

Characterization of KDGlucA (b3126), TSAR (b3125), and GK (b3124). Orfs b3126, b3125, b3124, and b3128 were amplified via PCR methodology and cloned downstream of a sequence encoding a polyhistidine tag in the pET16b vector as described in the Materials and Methods section. The encoded proteins were expressed and purified to apparent homogeneity.

The purified samples of KDGlucA and the putative TSAR and GK were added sequentially to a cuvette containing 1 equiv of 5-KDGluc, 3 equiv of NADH, excess ATP and PEP, lactate dehydrogenase, and pyruvate kinase. As shown in Figure 3, after addition of KDGlucA, 1 equiv of NADH was consumed due to the formation of pyruvate and subsequent reduction to lactate. After addition of the putative TSAR, a second equivalent of NADH was consumed due to the apparent reduction of tartronate semialdehyde. And, after addition of the putative GK, a third equivalent of NADH was consumed due to the utilization of 1 equiv of ATP coupled to the conversion of PEP to pyruvate and ultimately lactate. This experiment provides evidence that we correctly deduced the identities of b3126 (KDGlucA), b3125 (TSAR), and b3124 (GK) and that this operon encodes the remaining enzymes in the catabolic pathway for (*D*)-glucarate.

The purified samples of the putative KDGlucA, TSAR, and GK were subjected to kinetic analyses, and the values of the kinetic constants are summarized in Table 1. Tartronate semialdehyde is unstable under the conditions of the assays, so values of k_{cat}, K_m, and k_{cat}/K_m could not be

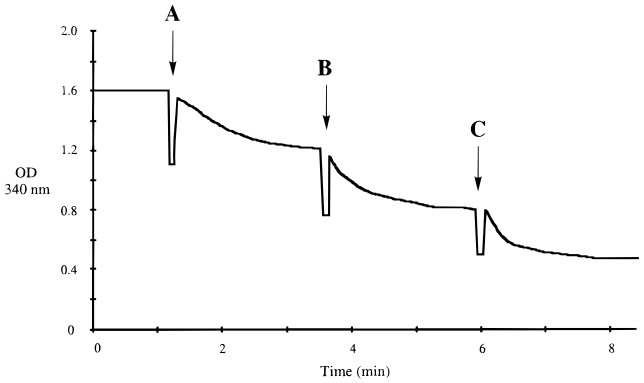


FIGURE 3: Consumption of NADH in the presence of excess ATP and PEP, lactate dehydrogenase, and pyruvate kinase. Points: A, addition of KDGlucA; B, addition of TSAR; C, addition of GK.

Table 1: Kinetic Constants for Enzymes in the (*D*)-Glucarate/Galactarate Catabolic Pathway

	k_{cat} (s^{-1})	K_m (10^6 M)	k_{cat}/K_m (M^{-1} s^{-1})
GlucD (glucarate)	20 ± 1	60 ± 2	3.3×10^5
GlucD (idarate)	20 ± 2	16 ± 2	1.2×10^5
GlucDRP (glucarate)	$5.6 \times 10^{-2} \pm 0.3 \times 10^{-2}$	320 ± 8	1.7×10^2
GlucDRP (idarate)	$1.3 \times 10^{-2} \pm 0.2 \times 10^{-2}$	170 ± 7	78
KDGlucA	27.3 ± 0.4	65 ± 3	4.2×10^5
TSAR	23 ± 4		
GK (glycerate)	2.5 ± 0.1	51 ± 1	5.0×10^4
GK (ATP)	2.5 ± 0.4	61 ± 2	4.1×10^4
GalcD	22.0 ± 0.7	800 ± 9	2.7×10^4

quantitated for this enzyme. Instead, only the k_{cat} was estimated using 10 mM tartronate semialdehyde generated in situ by the action of KDGlucA on 5-KDGluc and used immediately. The values for the kinetic constants confirm that the identities of the proteins encoded by b3126, b3125, and b3124, KDGlucA, TSAR, and GK, respectively, had been correctly deduced.

E. coli had been reported to contain two glycerate kinases (EC 2.7.1.31) designated GK1 and GK2 that catalyze the ATP-dependent formation of 3-phosphoglycerate (*18*). GK1 was expressed in cells grown on either (*D*)-glycerate or (*D*)-glucarate; GK2 was expressed in cells grown on glycolate. The spectrophotometric assay we used in measurement of the kinetic constants is based on the production of ADP from ATP and not the identity of the phosphoglycerate product, i.e., the previously reported 3-phosphoglycerate or its glycolytic precursor 2-phosphoglycerate. On the basis of the ^{1}H NMR spectrum of the reaction after it had proceeded to completion, the GK we identified (presumably GK1) catalyzes the formation of 2-phosphoglycerate (Figure 4), not 3-phosphoglycerate as reported by Blumenthal and Fish (*16*).

Characterization of GalcD (b3128). GalcD previously had been isolated from *E. coli* grown on either (*D*)-glucarate or galactarate as the sole carbon source (*16*). However, once the enzyme was isolated, it lost activity within a few hours and could not be reactivated. Using the knowledge that the sequence of GalcD is homologous to that of (*D*)-altronate dehydratase, an enzyme which requires Fe^{2+} and Mn^{2+} under reducing conditions for maximal activity, GalcD was both reactivated and assayed using conditions based on those described for (*D*)-altronate dehydratase (*19*). As purified, GalcD is catalytically inactive in the absence of added Fe^{2+},

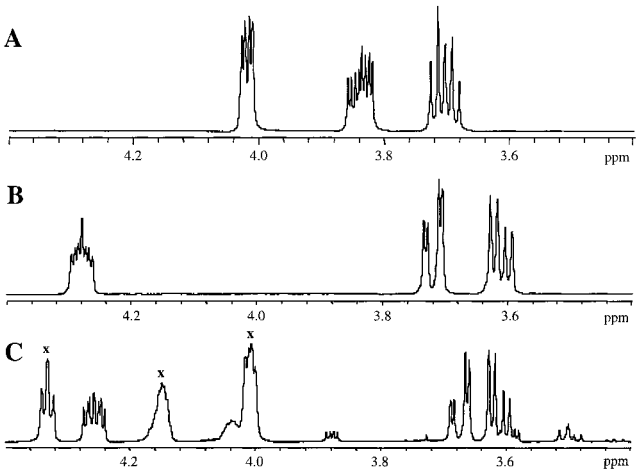


FIGURE 4: 500 MHz ^{1}H NMR spectra of (panel A) authentic 3-PGA, (panel B) authentic 2-PGA, and (panel C) the product obtained with GK. In panel C, the resonances associated with ADP/ATP are marked with x's.

Table 2: Relative Rates of Epimerization and Dehydration Catalyzed by GlucD and GlucDRP

epimerization/dehydration	GlucD	GlucDRP
glucarate to idarate	0.67 ± 0.03	1.0 ± 0.1
idarate to glucarate	0.83 ± 0.04	12 ± 1

Mn^{2+}, and β-mercaptoethanol. Like (*D*)-altronate dehydratase, GalcD was found to have maximal activity at pH 7.5 in the presence of 0.5 mM $FeSO_4$ and 170 mM β-mercaptoethanol; however, under these conditions GalcD is inhibited by Mn^{2+}. The values of k_{cat}, K_m, and k_{cat}/K_m determined using these conditions are presented in Table 1. These values confirm the identification of b3128 as GalcD.

Characterization of GlucD and GlucDRP. In a previous section we described the isolation of GlucD from extracts of *E. coli* MG1655 and found that it was encoded by orf b2787. We also isolated GlucDRP, encoded by b2788, and compared its kinetic properties with those of GlucD.

The kinetic constants for dehydration of both (*D*)-glucarate and (*L*)-idarate catalyzed by GlucD and GlucDRP also are presented in Table 1. In contrast to its homologue, GlucDRP is an inefficient catalyst of the dehydration of either (*D*)-glucarate or (*L*)-idarate. Despite systematic attempts to identify alternate tetraric, pentaric, and hexaric acid substrates for GlucDRP that might be dehydrated with k_{cat}s larger than those measured for (*D*)-glucarate and (*L*)-idarate, none could be identified (data not shown).

A ^{1}H NMR-based assay was used to quantitate the relative rates of the competing dehydration and epimerization reactions catalyzed by both GlucD and GlucDRP (Table 2) (*20*). The rates of GlucD-catalyzed dehydration of both (*D*)-glucarate and (*L*)-idarate are nearly equivalent and slightly exceed those for epimerization (using 100 mM substrates). In the case of GlucDRP, the k_{cat} for dehydration of (*D*)-glucarate exceeds that for (*L*)-idarate by a factor of 5, but the rate of epimerization of (*L*)-idarate exceeds that of (*D*)-glucarate by a factor of 2. That GlucDRP is deficient in both dehydration and epimerization suggests that neither of these hexaric acid substrates can bind in the optimal orientation for proton abstraction from carbon-5.

A possible structural explanation for the decreased rates of dehydration and epimerization catalyzed by GlucDRP is

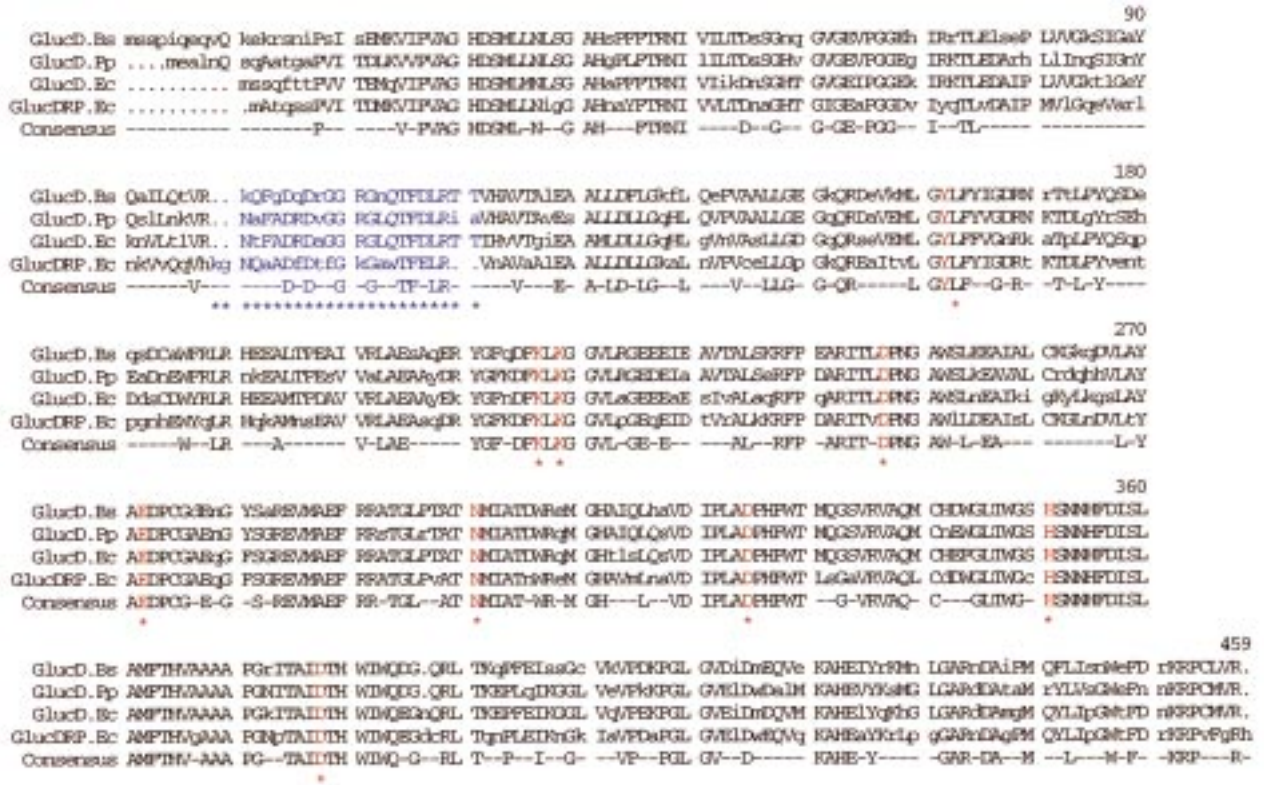


FIGURE 5: Alignment of the primary sequences of GlucDRP and the GlucDs from *E. coli*, *P. putida*, and *B. subtilis*. The conserved active site residues are marked in red, and the sequences associated with the undefined loop in the GlucD from *P. putida* are marked in blue. The active site residues and loop sequences also are marked with asterisks.

suggested by the alignment of the primary sequences of GlucD and GlucDRP (72% sequence identity) as well as the GlucDs from *P. putida* and *B. subtilis* (Figure 5). The alignment of these sequences together with the high-resolution structural information that is now available for both mandelate racemase (*21*, *22*) and the GlucD from *P. putida* (*23*) allows the prediction that the active sites of GlucDRP as well as all three GlucDs each contain *all* of the functional groups required for abstraction of the 5-proton. However, several differences between GlucDRP and the GlucDs are noted in the primary sequences of a structurally undefined loop that likely closes over the active site (residues 95−119 in GlucDRP), suggesting that this structural element is important in both substrate binding and the resulting attainment of a catalytically productive active site geometry.

Perhaps GlucDRP is the result of a gene duplication event that lead to a new substrate specificity in the enolase superfamily, e.g., dehydration of an unidentified sugar diacid that initiates its utilization as the carbon source by *E. coli*; however, as noted previously, a systematic study of hexaric, pentaric, and tetraric acids as carbon sources yielded no efficient substrates for GlucDRP. Alternatively, perhaps GlucDRP has not yet evolved to utilize an alternate diastereomeric pair of sugar diacids. In any event, for our future purposes we prefer to regard the GlucDRP as the starting point for the directed evolution of enhanced catalytic activity for utilization of (*D*)-glucarate [or (*L*)-idarate] that may provide functional identification of those structural elements that are involved in binding and aligning the substrate with active site functional groups in the active site of GlucD.

CONCLUSIONS

Taking advantage of the recently completed genome sequence for the *E. coli* chromosome, we have identified, isolated, and kinetically characterized the enzymes involved in the catabolism of both (*D*)-glucarate and galactarate by *E. coli*. The reactions catalyzed by these proteins were predicted by sequence relationships to previously identified enzymes catalyzing the same or similar reactions. We note, however, that the identification of their substrate specificities was accomplished only as a result of specific biochemical information; i.e., sequence relationships alone are sufficient to identify function but not substrate specificity. We anticipate that the determination of substrate specificities will be the most challenging problem in specifying the functions of the large number of proteins of unknown functions that are resulting from genome sequencing projects.

REFERENCES

1. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C. A., Gocayne, J. D., Scott, J. D., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., and Venter, J. C. (1995) *Science 269*, 296−512.
2. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J. L., Nguyen, D.

T., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J.-F., Dougherty, B. A., Bott, K. F., Hu, P.-C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison, C. A., III, and Venter, J. C. (1995) *Science 270*, 297−403.

3. Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B. C., and Herrmann, R. (1996) *Nucleic Acids Res. 24*, 4420−4449.
4. Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Mirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Maruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., and Tabata, S. (1996) *DNA Res. 3*, 109−136.
5. Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L. Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzgerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weldman, J. M., Fujil, Cl., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. M., Smith, H. O., Fraser, C. M., and Venter, J. C. (1997) *Nature 388*, 539−547.
6. Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) *Science 277*, 1453−1474.
7. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denizot, F., Devine, K. M., Dusterhoft, A., Ehrlich, S. D., Emmerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S. Y., Glaser, P., Goffeau, A., Golightly, E. J., Grandi, G., Guiseppi, G., Guy, B. J., Haga, K., Haiech, J., Harwood, C. R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M. F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S. M., Levine, A., Liu, H., Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S. H., Parro, V., Pohl, T. M., Portetelle, D., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S. J., Serror, P., Shin, B. S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H. F., Zumstein, E., Yoshikawa, H., and Danchin, A. (1997) *Nature 390*, 249−256.
8. Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R. A., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., Gwinn, M., Dougherty, B., Tomb, J.-F., Fleischmann, R. D., Richardson, D., Peterson, J., Kerlavage, A. R., Quackenbush, J., Salzberg, S., Hanson, M., van-Vugt, R., Palmer, N., Adams, M. D., Gocayne, J. D., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M. D., Horst, K., Roberts, K., Hatch, B., Smith, H. O., and Venter, J. C. (1997) *Nature 390*, 580−586.
9. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., Fitzgerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., and Venter, J. C. (1996) *Science 273*, 1058−1073.
10. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qui, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrookovsi, S., Church, G. M., Daniels, C. J., Mao, J.-I., Rice, P., Nölling, J., and Reeve, J. N. (1997) *J. Bacteriol. 179*, 7135−7155.
11. Klenk, H. P., Clayton, R. A., Tomb, J., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenney, K., Adams, M. D., Loftus, B., Peterson, S., Reich, C. I., McNeil, L. K., Badger, J. H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J. D., Weidman, J. F., McDonald, L., Utterback, T., Cotton, M. D., Spriggs, T., Artiach, P., Kaine, B. P., Sykes, S. M., Sadow, P. W., D'Andrea, K. P., Bowman, C., Fujii, C., Garland, S. A., Mason, T. M., Olsen, G. J., Fraser, C. M., Smith, H. O., Woese, C. R., and Venter, J. C. (1997) *Nature 390*, 364−378.
12. Deckert, G., Warren, P. V., Gassterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J., and Swanson, R. V. (1998) *Nature 392*, 353−358.
13. Mewes, H. W., Albermann, K., Bahr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., Kleine, K., Maierl, A., Oliver, S. G., Pfeiffer, F., and Zollner, A. (1997) *Nature 387*, 7−65.
14. Babbitt, P. C., and Gerlt, J. A. (1997) *J. Biol. Chem. 272*, 30591−30594.
15. 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*, 1159−1162.
16. Blumenthal, H. J., and Fish, D. C. (1963) *Biochem. Biophys. Res. Commun. 11*, 239−243.
17. MacGee, J., and Doudoroff, M. (1954) *J. Biol. Chem. 210*, 617−624.
18. Ornston, M. K., and Ornston, L. N. (1969) *J. Bacteriol. 97*, 1227−1233.
19. Dreyer, J. L. (1987) *Eur. J. Biochem. 166*, 623−630.
20. Palmer, D. R. J., Hubbard, B. K., and Gerlt, J. A. (1998) *Biochemistry 37*, 14350−14357.
21. 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.
22. 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.
23. Gulick, A. M., Palmer, D. R. J., Babbitt, P. C., Gerlt, J. A., and Rayment, I. (1998) *Biochemistry 37*, 14358−14368.

BI981124F