THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 280, No. 10, Issue of March 11, pp. 9662–9670, 2005
Printed in U.S.A.

Molecular Structure of Human Galactokinase

IMPLICATIONS FOR TYPE II GALACTOSEMIA*

Received for publication, November 15, 2004
Published, JBC Papers in Press, December 7, 2004, DOI 10.1074/jbc.M412916200

James B. Thoden‡, David J. Timson§, Richard J. Reece¶‖, and Hazel M. Holden‡**

From the ‡Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, the §School of Biology and Biochemistry, Queen's University Belfast, Medical Biology Centre, 97 Lisburn Road, Belfast BT9 7BL, United Kingdom, and the ¶Faculty of Life Sciences, The University of Manchester, The Michael Smith Building, Oxford Road, Manchester M13 9PT, United Kingdom

Galactokinase functions in the Leloir pathway for galactose metabolism by catalyzing the MgATP-dependent phosphorylation of the C-1 hydroxyl group of α-D-galactose. The enzyme is known to belong to the GHMP superfamily of small molecule kinases and has attracted significant research attention for well over 40 years. Approximately 20 mutations have now been identified in human galactokinase, which result in the diseased state referred to as Type II galactosemia. Here we report the three-dimensional architecture of human galactokinase with bound α-D-galactose and Mg-AMPPNP. The overall fold of the molecule can be described in terms of two domains with the active site wedged between them. The N-terminal domain is dominated by a six-stranded mixed β-sheet whereas the C-terminal motif contains six α-helices and two layers of anti-parallel β-sheet. Those residues specifically involved in sugar binding include Arg^{37}, Glu^{43}, His^{44}, Asp^{46}, Gly^{183}, Asp^{186}, and Tyr^{236}. The C-1 hydroxyl group of α-D-galactose sits within 3.3 Å of the γ-phosphorus of the nucleotide and 3.4 Å of the guanidinium group of Arg^{37}. The carboxylate side chain of Asp^{186} lies within ~3.2 Å of the C-2 hydroxyl group of α-D-galactose and the guanidinium group of Arg^{37}. Both Arg^{37} and Asp^{186} are strictly conserved among both prokaryotic and eukaryotic galactokinases. In addition to providing molecular insight into the active site geometry of the enzyme, the model also provides a structural framework upon which to more fully understand the consequences of the those mutations known to give rise to Type II galactosemia.

In most organisms, the conversion of β-D-galactose to glucose-1-phosphate is accomplished by the action of four enzymes that constitute the Leloir pathway as indicated in Scheme 1 (1–4). Galactokinase (EC 2.7.1.6), the focus of this investigation, catalyzes the second step in the pathway, namely the MgATP-dependent phosphorylation of the C-1 hydroxyl group of α-D-galactose to yield galactose-1-phosphate (5, 6). Thus far, ~20 mutations (including base substitutions, base deletions, and larger deletions) have been identified in human galactokinase, which result in the diseased state referred to as Type II galactosemia (MIM 230200). This disease, the main symptom of which is early onset cataracts, is less severe than either Type I or Type III galactosemia, which are caused by defects in the galactose-1-phosphate uridylyltransferase or UDP-galactose 4-epimerase genes, respectively (7–10).

The molecular mechanism of cataract formation is unclear but it is thought to be similar to that sometimes suffered by diabetics. High blood concentrations of galactose (or glucose in the case of diabetics) result in high concentrations of the sugar being transported into the lens cells of the eye. These cells have unusually high levels of the enzyme aldose reductase, which catalyzes the conversion of the sugar to its corresponding sugar alcohol galactitol (dulcitol) in the case of galactosemics and sorbitol in the case of diabetics (11, 12). Although the sugar can be transported across the cell membrane, the sugar alcohol cannot (13–15). Therefore, it accumulates and increases the osmotic potential drawing in water. It appears that this uptake of water leads to damage and ultimately to the formation of cataracts.

Within the last year, the crystal structures of two bacterial galactokinases were reported (16, 17). The structure of the enzyme from *Lactococcus lactis* was solved in the presence of galactose and a phosphate ion whereas that from *Pyrococcus furiosus* was determined with galactose and MgADP bound in the active site (16, 17). Many of the equivalent residues to the sites of disease-causing mutations in the human enzyme could be located in both models and speculations were made about the reason for structural or kinetic dysfunction (18). However, not all the disease-causing mutations could be mapped unambiguously onto these bacterial enzyme structures.

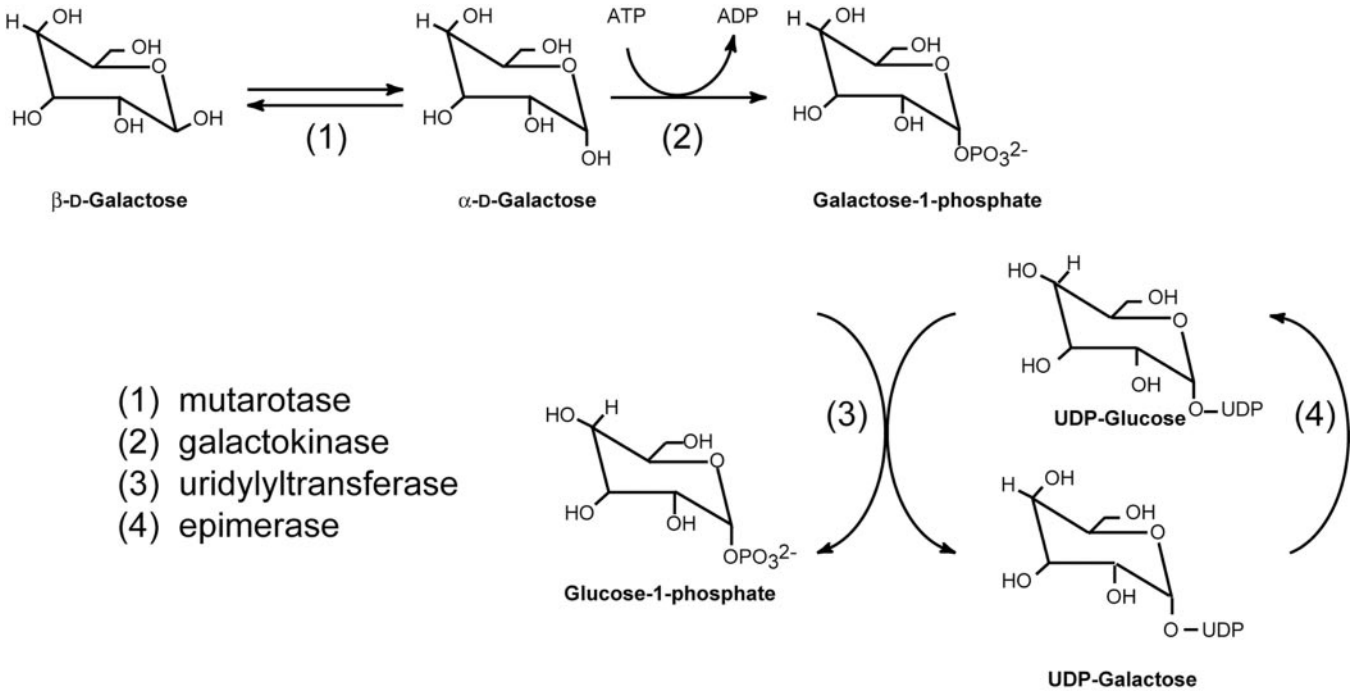
Galactokinase is known to be a member of the GHMP superfamily of small molecule kinases, which includes homoserine kinase, mevalonate kinase, and phosphomevalonate kinase, among others (reviewed in Ref. 18). Indeed, this family of proteins was originally identified on the basis of amino acid sequence identity (19). According to all presently available structural and biochemical data, members of this family can be subdivided into two groups depending upon their catalytic mechanisms. The crystal structure of rat mevalonate kinase in complex with ATP shows an aspartate residue (Asp^{204}) in the active site which is positioned to act as a catalytic base and a lysine residue (Lys^{13}) that is thought to influence the pK_a of the substrate hydroxyl group (20). In contrast, in *Escherichia coli* homoserine kinase, the equivalent residue to the aspartate is an asparagine (Asn^{141}), and there appears to be no other suitable residue close enough to act as a catalytic base (21, 22).

* This research was supported in part by Grant DK47814 from the National Institutes of Health (to H. M. H.) and from the Wellcome Trust and BBSRC (to R. J. R). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1WUU) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

‖ To whom correspondence may be addressed. E-mail: Richard.Reece@man.ac.uk.

** To whom correspondence may be addressed. Tel.: 608-262-4988; Fax: 608-262-1319; E-mail: Hazel_Holden@biochem.wisc.edu.



SCHEME I

Furthermore, there is no adjacent positively charged residue. Consequently, it has been hypothesized that homoserine kinase achieves catalysis through transition state stabilization alone (22). Which of the two subgroups galactokinase falls into is presently unclear. The crystal structures of the enzyme from both *L. lactis* and *P. furiosus* reveal an aspartate residue in an approximately equivalent position to that found in mevalonate kinase and an arginine occupying the same position as Lys[13] (16, 17). However, enzymological studies on the *Saccharomyces cerevisiae* enzyme (which also has an equivalent aspartate residue) showed little variation in the turnover number and specificity constants with respect to pH and negligible effects on addition of increasing mol fractions of deuterium oxide in place of water (23). Similarly, the human enzyme showed no solvent deuterium isotope effect (24). Both experiments would have been expected to show some effect on these kinetic constants if an ionizable residue was involved in a reaction step reflected in the measured parameters (25).

Here we report the three-dimensional structure of human galactokinase in complex with Mg-AMPPNP[1] and α-D-galactose, determined to a nominal resolution of 2.5 Å. This investigation has allowed for a more complete description of the active site geometry of galactokinase and has unambiguously revealed the locations of the mutations known to cause Type II galactosemia.

EXPERIMENTAL PROCEDURES

Expression of the Selenomethionine-labeled Protein—Cultures of *E. coli* HMS174(DE3) cells harboring the plasmid encoding the human *GALK1* gene, as described in Ref. 24, were grown overnight in M9 minimal medium at 37 °C. Subsequently, 15 ml of the overnight culture were used to inoculate each of 12 × 2-liter baffled flasks containing 500 ml of M9 minimal medium supplemented with 5 mg/liter thiamine. Cultures were grown at 37 °C to an OD_{600} of ~0.9 and then cooled for 10 min in an ice/water bath. They were then transferred to an incubator at 16 °C, and each flask was supplemented with 50 mg each of lysine, threonine, and phenylalanine and 25 mg each of leucine, isoleucine, valine, and selenomethionine. Following 20 min of additional growth, the cultures were induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside and allowed to grow for an additional 18 h. Subsequently, the cells were harvested by centrifugation at 10,000 × *g* for 10 min, and the cell paste frozen in liquid nitrogen for storage at −80 °C. Note that the recombinant plasmid employed in this investigation results in a His_6 tag at the N terminus of the protein.

Protein Purification—All protein purification steps were conducted on ice or at 4 °C. The cell paste was thawed in three volumes of lysis buffer containing 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, and 100 mM galactose (pH 8). Cells were lysed with four cycles of sonication and cooling (45 s of sonication separated by 5 min of cooling). Cellular debris was removed by centrifugation at 20,000 × *g* for 30 min. The clarified lysate was loaded onto a 10 ml Ni-NTA agarose column (Qiagen) pre-equilibrated in lysis buffer. After loading, the column was washed with 100 ml of lysis buffer, followed by 100 ml of wash buffer, which contained an additional 10 mM imidazole. The protein was recovered by a linear gradient elution with 20–300 mM imidazole in the lysis buffer. Protein-containing fractions were pooled on the basis of purity as visualized by SDS-PAGE and dialyzed against 10 mM Hepes, 200 mM NaCl, and 100 mM D-galactose (pH 8). The dialyzed protein was concentrated to ~7 mg/ml based on a calculated extinction coefficient of 2.05 ml/(mg·cm) (DNAstar Inc., Madison, WI). Galactose-dependent ATPase activity was verified using the standard pyruvate kinase/lactate dehydrogenase-coupled assay.

Crystallization—Large single crystals were grown at room temperature via batch methods using 6–8% polyethylene glycol 8000, 2% Me_2SO, 10 mM Mg-AMPPNP, and 100 mM HEPPS (pH 8). These crystals achieved typical dimensions of 0.5 mm × 0.1 mm × 0.1 mm in 2–3 weeks and belonged to the space group $P2_1$ with unit cell dimensions of a = 73.2, b = 109.6, and c = 115.8 and β = 95.9°. The asymmetric unit contained four polypeptide chains.

High Resolution X-ray Data Collection—Prior to flash-cooling, the crystals were harvested from the batch experiments and equilibrated in a synthetic mother liquor composed of 15% polyethylene glycol 8000, 150 mM NaCl, 100 mM D-galactose, 10 mM Mg-AMPPNP, 2% Me_2SO, and 100 mM HEPPS (pH 8). The crystals were then serially transferred to a cryoprotectant solution composed of 25% polyethylene glycol 8000, 250 mM NaCl, 100 mM D-galactose, 2% Me_2SO, 15% ethylene glycol, and 100 mM HEPPS (pH 8).

X-ray data were collected on a CCD detector at the COMCAT 32-ID beamline (Advanced Photon Source, Argonne National Laboratory, Argonne, IL). All x-ray data were processed and scaled with HKL2000 (26). Relevant x-ray data collection statistics are presented in Table I.

X-ray Structural Analyses—The structure of human galactokinase was solved via MAD phasing with crystals of the selenomethionine-substituted protein. The software package SOLVE was utilized to determine the positions of the 32 selenium atoms in the asymmetric unit

[1] The abbreviation used is: AMPPNP, adenosine 5′-(β,γ-imino)-triphosphate.

TABLE I
X-ray data collection statistics

	Wavelength	Resolution	No. independent reflections	Completeness	Redundancy	Avg I/Avg $\sigma(I)$	R_{sym}[a]
	Å	Å		%			%
Peak	0.97916	30.0–2.4	66449	94.3	3.5	18.9	7.1
Inflection	0.97931	30.0–2.4	66703	94.1	3.5	19.0	7.7
Remote	0.96408	30.0–2.4	68009	95.7	3.5	20.5	7.6
Native	0.97000	30.0–2.15	92692	93.7	3.3	21.5	5.4

[a] $R_{sym} = (\Sigma|I - \bar{I}|/\Sigma I) \times 100$.

and to generate initial protein phases (figure-of-merit = 0.5) (27). Solvent flattening and molecular averaging of the four polypeptide chains in the asymmetric unit with RESOLVE (figure-of-merit = 0.65) resulted in an interpretable electron density map calculated to 2.5-Å resolution (28). The "averaged" map allowed for a complete tracing of the structure (392 amino acids) with the exception of several residues at the N terminus and a surface loop near His^{229} to Ser^{233}. Four copies of the model were placed back into the asymmetric unit and alternate cycles of manual model building and least-squares refinement with the software package TNT reduced the R-factor to 20.1% for all measured x-ray data from 20.0- to 2.5-Å resolution (29). The crystals diffracted to ~2.1-Å resolution, but because of anisotropy in the diffraction pattern, the least-squares refinement was truncated at 2.5-Å resolution. Relevant refinement statistics are presented in Table II. A Ramachandran plot analysis indicates that 85.1, 14.6, and 0.3% of the residues are in the most favored, the additional allowed, and the generously allowed regions, respectively.

TABLE II
Least squares refinement statistics

Resolution limits (Å)	20.0–2.5
R-factor[a] (overall) %/ no. reflections	20.5/61550
R-factor (working) %/ no. reflections	20.1/55451
R-factor (free) %/ no. reflections	25.8/6099
No. protein atoms[b]	12367
No. heteroatoms[c]	560
Average B values (Å^2)	
Protein atoms	45.9
α-D-Galactose ligands	33.6
AMPPNP ligands	40.5
Magnesium ions	33.5
Waters	45.5
Weighted root mean square deviations from ideality	
Bond lengths (Å)	0.013
Bond angles (degree)	2.13
Trigonal planes (Å)	0.007
General planes (Å)	0.015
Torsional angles (deg)[d]	18.8

[a] R-factor = $(\Sigma|F_o - F_c|/\Sigma|F_o|) \times 100$, where F_o is the observed structure-factor amplitude, and F_c is the calculated structure-factor amplitude.
[b] This value includes multiple conformations for Glu^{70} in chain A and Arg^{297} in chain D.
[c] These include 4 galactose molecules, 4 AMPPNP, 4 Mg^{2+}, and 384 waters.
[d] The torsional angles were not restrained during the refinement.

RESULTS AND DISCUSSION

Structure of Human Galactokinase—The crystals of human galactokinase employed in this investigation contained four polypeptide chains in the asymmetric unit. In that the electron density was best defined for the fourth molecule, the following discussion will refer specifically to it unless otherwise stated. Note, however, that the backbone α-carbon atoms for the individual polypeptide chains in the asymmetric unit superimpose upon one another with root mean-square deviations of between 0.31 and 0.42 Å.

As indicated in Fig. 1*a*, the structure of human galactokinase, with overall dimensions of ~44 Å × 56 Å × 63 Å, can be envisioned as two domains. The N-terminal region initiates with a long α-helix of 15 residues, which flanks one side of a six-stranded mixed β-sheet. A four helical bundle surrounds the other side of this β-sheet. The C-terminal domain is dominated by six α-helices and two layers of anti-parallel β-sheet, each containing four strands. The location of the three structural motifs characteristic of the GHMP superfamily are indicated in Fig. 1*a*. Motif I, having the sequence ^{38}Val-Asn-Leu-Ile-Gly-Glu-His^{44}, abuts one side of the galactose moiety where the side chain of Glu^{43} and the backbone peptidic nitrogen of His^{44} participate in hydrogen bonding interactions with the C-6 hydroxyl group of the sugar. In human galactokinase, Motif II is formed by ^{136}Gly-Gly-Gly-Leu-Ser-Ser-Ser-Ala-Ser^{144} and envelopes the phosphoryl moieties of the nucleotide. Finally, Motif III, as indicated in Fig. 1*a*, is formed by ^{343}Met-Thr-Gly-Gly-Gly-Phe-Gly-Gly^{350}. The backbone peptidic nitrogen of Gly^{346} lies within 3 Å of a γ-phosphoryl oxygen of AMPPNP. There are three cis-prolines in human galactokinase at positions 85, 104, and 121.

Earlier reports suggested that the quaternary structures of the galactokinases from *E. coli*, *L. lactis*, *P. furiosus*, yeast, and human are monomeric, whereas that isolated from *Vicia faba* seeds is dimeric (16, 17, 24, 30–32). Analysis of the crystalline packing arrangement for the four molecules in the asymmetric unit reveals that, in this study, they associate as dimers through similar secondary structural elements. One such dimer, formed between molecules 3 and 4 (as listed in the deposited coordinate file) is displayed in Fig. 1*b*. The total surface area buried upon dimer formation, as calculated with a probe radius of 1.4 Å is ~1750 Å^2 (33). The subunit:subunit interface is formed primarily by two loops (Ser^{160}-Ile^{163} and Gly^{193}-Lys^{195}), and two β-strands (Glu^{207}-Pro^{212} and Lys^{388}-Cys^{391}) contributed by each subunit. Indeed, the β-strand defined by Glu^{207} to Pro^{212} extends the four-stranded anti-parallel β-sheet in the C-domain of one subunit into an eight-stranded anti-parallel β-sheet across the dimer interface. The residues involved in the formation of the subunit:subunit interface are not conserved among eukaryotic galactokinases. As indicated in Fig. 1*b*, there is a disulfide bridge formed by the side chains of Cys^{391} (one from each subunit). This occurs in both dimers in the asymmetric unit and is most likely a function of the crystallization conditions. Whether the dimeric quaternary structure of human galactokinase observed in this investigation is merely a function of crystalline packing or represents the state of the enzyme *in vivo* is unknown.

Active Site of Human Galactokinase—Electron density corresponding to the bound ligands, α-D-galactose and Mg-AMPPNP, is presented in Fig. 2*a*. Those amino acid residues lying within ~3.5 Å of the sugar or nucleotide are displayed in Fig. 2*b*. The adenine base of AMPPNP adopts the anti-conformation and participates in a stacking interaction with the indole ring of Trp^{106}. The nucleotide ribose assumes the $C_{3'}$-endo pucker. Tyr^{109} sits near the nucleotide ribose with its side chain pointing toward O6R as indicated in Fig. 2*b*. Immediately following Motif II, there is a rather long α-helix of 16 residues positioned with the positive end of its helix dipole moment directed toward the phosphoryl groups of the nucleotide. The three serine residues in Motif II aid in anchoring the phosphoryl groups of the nucleotide within the active site cleft. Specifically, O^{γ} of Ser^{142} sits at 3.2 Å and 2.7 Å, respectively, from an α- and a β-phosphoryl oxygen while O^{γ} of Ser^{141} is located within 2.8 Å of a

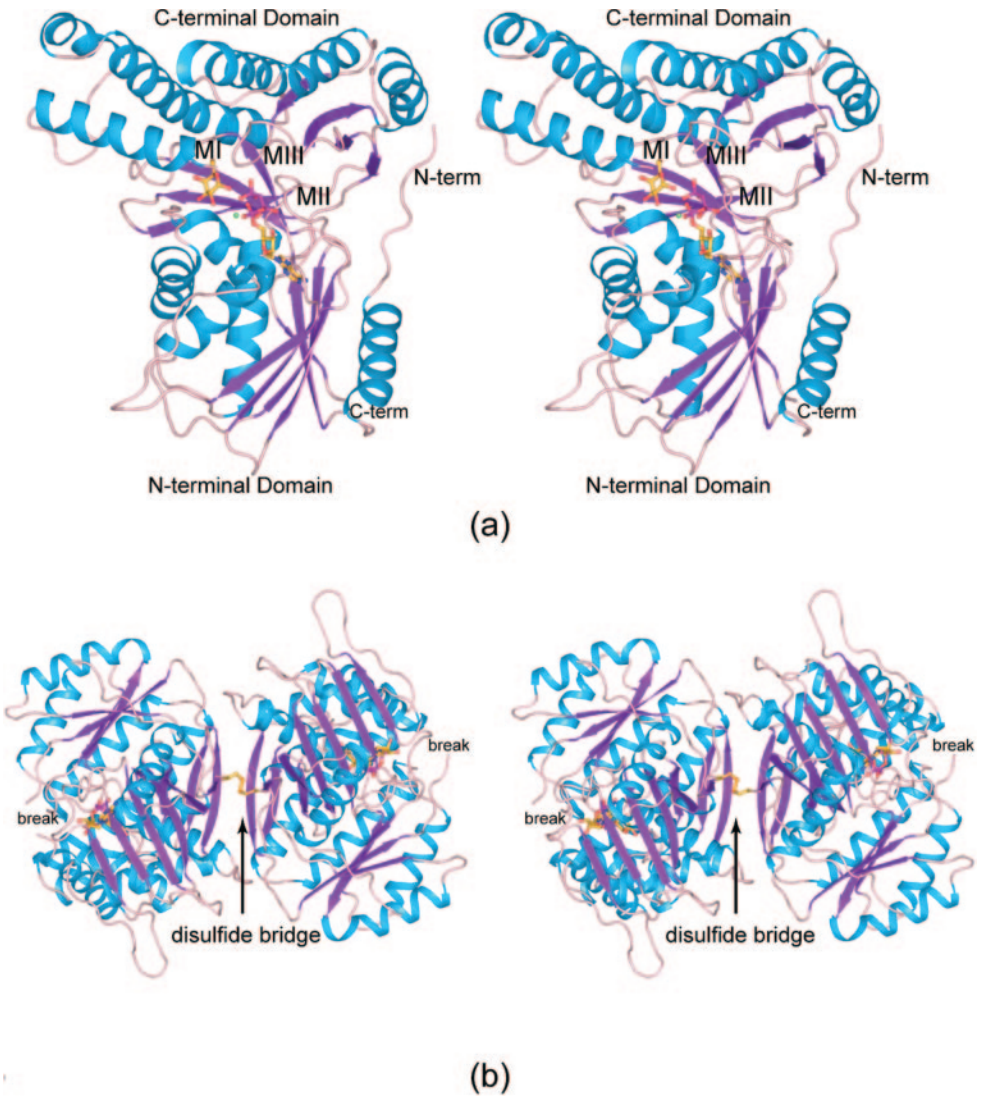


FIG. 1. **Molecular architecture of human galactokinase.** A *ribbon representation* of one subunit of the enzyme is shown in *a* with the bound α-D-galactose and Mg-AMPPNP ligands displayed in *ball-and-stick* representations. The *green sphere* indicates the position of the bound magnesium ion. The locations of the conserved structural motifs in the GHMP superfamily are indicated by the labels *MI*, *MII*, and *MIII*. The four molecules contained within the crystalline asymmetric unit pack as dimers with local 2-fold rotational axes. One of these dimers, viewed down the local dyad, is shown in *b*. In both dimers a disulfide bond is formed between neighboring Cys^{391} side chains.

β-phosphoryl oxygen. Additionally, the backbone peptidic nitrogen of Ser^{140} is positioned within 2.8 Å of a γ-phosphoryl oxygen and the backbone peptidic nitrogen of Ser^{141} lies within 2.9 Å a β-phosphoryl oxygen. The magnesium ion is wedged between α-, β-, and γ-phosphoryl oxygens with metal:oxygen distances of 2.3, 2.6, and 2.3 Å, respectively. At a resolution of 2.5 Å, it is not possible to completely define the coordination geometry surrounding the metal ion.

A closeup stereo view of potential interactions between the protein and both the sugar and the γ-phosphoryl group of AMPPNP is presented in Fig. 3. The C-1 hydroxyl oxygen, which is phosphorylated during the reaction, sits between two of the γ-phosphoryl oxygens of the nucleotide and lies within 3.3 Å of the γ-phosphorus. The side chain nitrogen of Arg^{37}, namely $N^{\eta 2}$, is within 3.4 Å of the galactose C-1 hydroxyl group and 3.1 Å of the third γ-phosphoryl oxygen. In the case of mevalonate kinase, for example, it has been postulated that a similarly positioned lysine residue serves both to lower the pK_a of the C-5 hydroxyl group of mevalonate and to aid in stabilization of the presumed pentacoordinated γ-phosphate transition state (20). As indicated in Fig. 3, the carboxylate group of Asp^{186} interacts with the C-2 hydroxyl of galactose (3.2 Å) and sits at 3.8 Å from the galactose C-1 hydroxyl group. Both Arg^{37} and Asp^{186} are strictly conserved among galactokinase sequences reported to date. Other side chains involved in anchoring the sugar substrate to the protein include Asp^{46} which bridges O-3 and O-4 of galactose (2.7 and 2.6 Å), Tyr^{236}, which interacts with O-4 (2.7 Å), and Glu^{43} which hydrogen bonds to O-6 (2.3 Å). In addition to side chain/sugar interactions, the backbone peptidic groups of Gly^{183} and His^{44} form hydrogen bonds with O-3 and O-6 of galactose, respectively.

A recent investigation of sugar recognition by human galactokinase demonstrated that 2-deoxy-D-galactose also functions as a substrate for the enzyme but *N*-acetyl-D-galactosamine, L-arabinose, D-fucose, and D-glucose are not phosphorylated (34). Assuming that for catalysis to occur, all of the above-mentioned sugars must bind to the protein in a similar orientation as that observed for galactose, it is now possible to provide a structural explanation for these results. In the case of *N*-acetyl-D-galactosamine, it is not possible to position the sugar into the active site without steric clashes arising between its *N*-acetyl group and the side chains of Glu^{174}, Met^{180}, Cys^{182}, Asp^{186}, and/or the phosphoryl groups of AMPPNP. Both L-arabinose (a five carbon analog of galactose) and D-fucose (6-

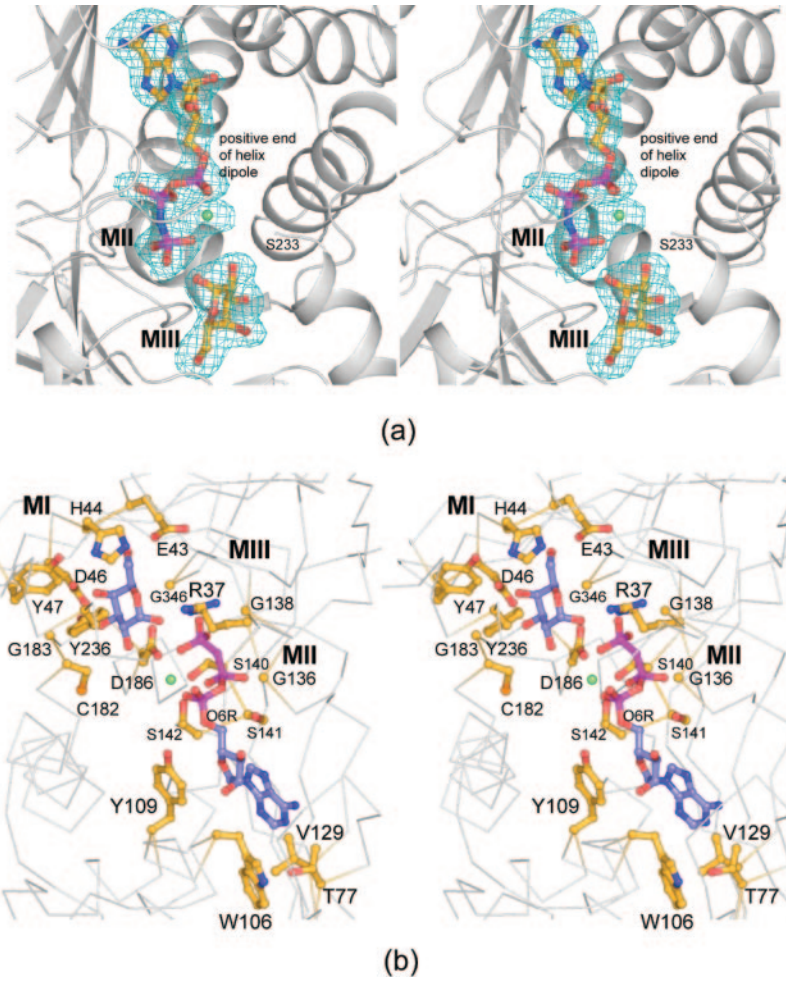


FIG. 2. **The active site.** Electron density corresponding to the bound ligands is displayed in *a*. The map was contoured at 3σ and calculated with coefficients of the form ($F_o - F_c$) where F_o and F_c were the native and calculated structure factor amplitudes, respectively. The coordinates for the ligands were not included in the map calculations. Those amino acid residues located within ~3.5 Å of the ligands are shown in *b*. The magnesium ion is depicted as a *green sphere*.

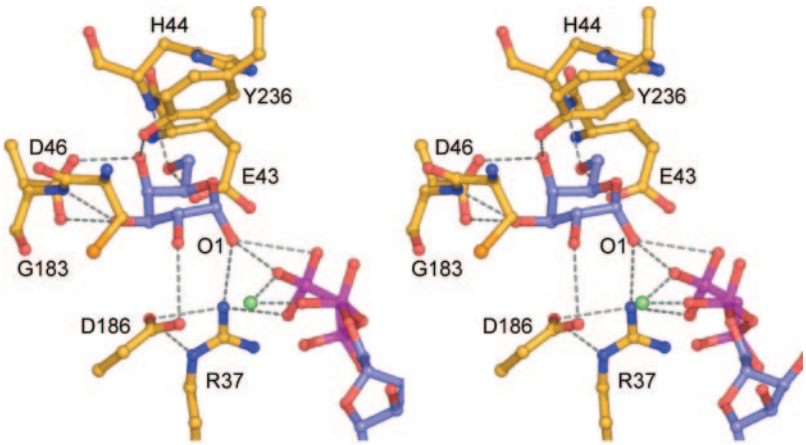


FIG. 3. **Potential interactions between the ligands and the protein.** The *dashed lines* indicate distances of between 2.3 and 3.2 Å. At a resolution of 2.5 Å; however, caution must be applied in the assignment of hydrogen bonds.

deoxygalactose) are missing the C-6 hydroxyl group. As shown in Fig. 3, both the side chain of Glu[43] and the peptidic nitrogen of His[44] participate in hydrogen bonding interactions with O-6 of galactose. The loss of these two electrostatic interactions between the protein and either L-arabinose or D-fucose most likely results in the inability of the respective sugar to bind to the enzyme. This is further supported by biochemical data demonstrating that these sugars do not function as inhibitors of the galactokinase reaction thereby ruling out an explanation of non-productive binding (34). Interestingly, in the recent study of the sugar specificity exhibited by human galactokinase, it was shown that replacing Glu[43] with an alanine residue resulted in little effect on the steady state kinetics of the enzyme with its natural substrate, α-D-galactose. Hence, while the abolition of the C-6 hydroxyl group in arabinose and fucose results in no catalytic activity, deleting the side chain carboxylate of Glu[43] causes little change in the steady state kinetics of the enzyme. This contradiction can now be rationalized on the basis of the structure presented here. Replacement of Glu[43] with an alanine results in the loss of only one of the two hydrogen bonding interactions that occur between the protein and the normal sugar substrate. The hydrogen bond formed between the C-6 hydroxyl group of galactose and the peptidic NH group of His[44] most likely remains intact and thus still

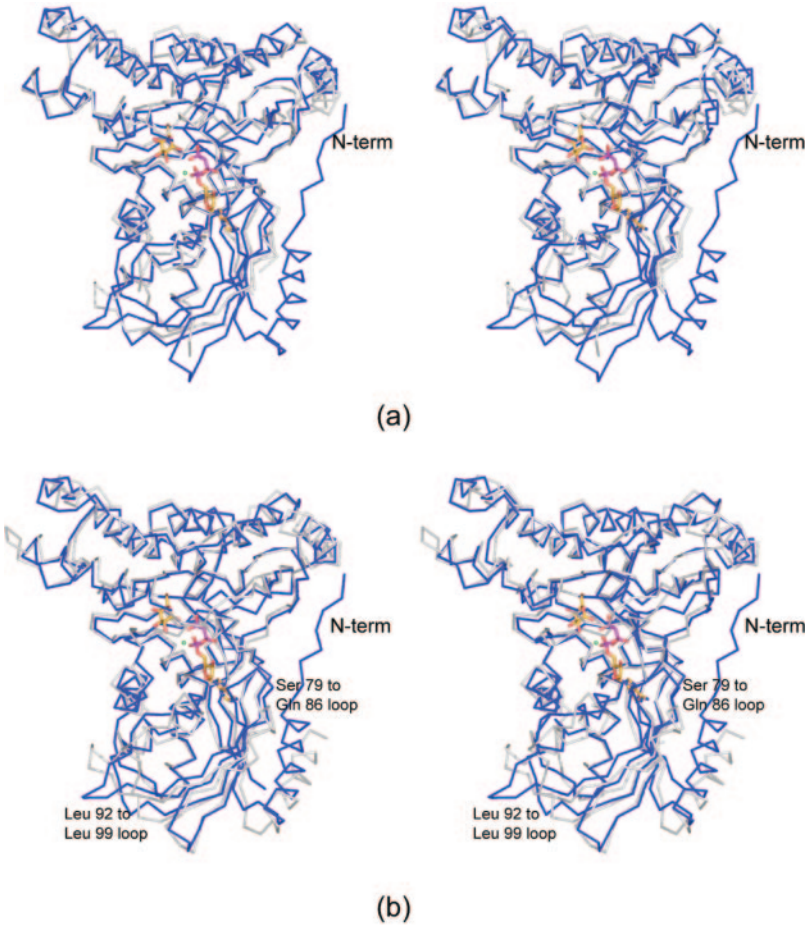


FIG. 4. **Comparisons of the human and bacterial galactokinase models.** A superposition of the structures for human (*blue*) and *P. furiosus* (*white*) galactokinases is given in *a*. Models for the human (*blue*) and *L. lactis* (*white*) enzymes are superimposed in *b*.

serves to properly orient the sugar within the active site. Finally, D-glucose and D-galactose differ simply by the hydroxyl configuration about C-4. Due to this difference in configuration, however, the hydroxyl group in glucose is no longer in the correct position to participate in hydrogen bonding interactions with the side chains of Asp^{46} and Tyr^{236}. Again, D-glucose does not function as a competitive inhibitor (34).

Implications for the Catalytic Mechanism of Galactokinase—The catalytic mechanisms of both homoserine kinase and mevalonate kinase, both members of the GHMP superfamily, have been the subject of several recent investigations (20, 22, 35). X-ray studies of ternary complexes of homoserine kinase with homoserine and non-hydrolyzable ATP analogs have revealed the lack of a catalytic base positioned near the δ-hydroxyl group of homoserine that is ultimately phosphorylated (22). The apparent absence of an active site base has led to the proposal that the catalytic mechanism of homoserine kinase proceeds through stabilization of the transition state. Specifically, the close positioning of the substrate and the γ-phosphate of the nucleotide within the active site is thought to result in the direct transfer of the proton from the δ-OH group of homoserine to the γ-phosphorus of ATP and attack of the δ-oxygen on the γ-phosphorus. The side chain hydroxyl group of Thr^{183} has been shown to interact with the β-phosphate of the nucleotide, and thus may help to stabilize the transition state.

In the case of mevalonate kinase, the catalytic mechanism is believed to proceed via acid/base chemistry. It has been proposed that the conserved aspartate (Asp^{204} in the rat enzyme or Asp^{155} in the *Methanococcus jannaschii* protein) serves to abstract the proton from the C-5 hydroxyl group of mevalonate for its subsequent attack on the γ-phosphorus of ATP. Additionally, there is a conserved lysine residue, (Lys^{13} in the rat enzyme or Lys^{8} in the *M. jannaschii* protein), which is hypothesized to lower the pK_a of the substrate C-5 hydroxyl group. It is important to note, however, that in the case of the *M. jannaschii* protein, the crystal structure reported is only that of the uncomplexed enzyme (35). Likewise, with the rat enzyme, the reported structure is that only of the binary complex with bound ATP (20).

With respect to galactokinase, initial kinetic analyses of the human enzyme have suggested that its catalytic mechanism may be more similar to that proposed for homoserine kinase (24). As can be seen in Fig. 3, however, the carboxylate group of Asp^{186} lies in an appropriate position to serve as a catalytic base. Likewise, the guanidinium group of Arg^{37} is positioned closely to the carboxylate of Asp^{186} (~3 Å), the C-1 hydroxyl group of galactose (3.4 Å), and the γ-phosphoryl moiety of the nucleotide (3.1 Å). Indeed, the structural results presented here suggest that, perhaps, the reaction mechanism of galactokinase proceeds via a similar mechanism as that proposed for mevalonate kinase. It is, however, intriguing that members of the GHMP superfamily are thought to function by quite distinct catalytic mechanisms. One of the obvious questions in the case of galactokinase and also in mevalonate kinase is whether the conserved aspartate serves as the catalytic base or rather is simply involved in substrate binding. Perhaps all members of the GHMP superfamily proceed via a similar mechanism which is not completely characterized yet and that the conserved aspartates in mevalonate kinase, phosphomevalonate kinase, and galactokinase function only in substrate binding. From the presently available biochemical and x-ray crystallographic data, it is not possible to rule out this possibility.

Similarities to the Bacterial Galactokinases—Human galactokinase shares amino acid sequence identities of 38 and 34%

TABLE III
Human galactokinase mutations associated with Type II galactosemia

Mutation	Disease phenotype[a]	Functional consequences	Refs.
M1I	Reduced blood enzyme activity	Not tested. Presumed to abolish initiation codon.	39
P28T	Very low or no detectable blood enzyme activity; formation of cataracts if galactose is not omitted from the diet	Protein insoluble on expression in *E. coli*	24, 37, 39, 42, 43
V32M	No detectable blood enzyme activity; formation of cataracts if galactose is not omitted from the diet	Protein insoluble on expression in *E. coli*	24, 36, 44
G36R	No detectable blood enzyme activity (heterozygote with frameshift)	Protein insoluble on expression in *E. coli*	24, 39
H44Y	No detectable blood enzyme activity (heterozygote with G349S)	Reduction in k_{cat}; increase in K_m for both substrates	24, 39
R68C	Low blood enzyme activity (heterozygote with A384P)	Small reductions in k_{cat} and K_m for galactose; increase in K_m for ATP	24, 40
S131I	Low blood enzyme activity (double mutation with G137C)	Not tested	42
G137C	Low blood enzyme activity (double mutation with S131I)	Not tested	42
A198V	Greater probability of cataracts in late middle age (50+)	No change greater than 1.5-fold in any parameter	24, 41
R239Q	Reduced red blood cell enzyme activity	Reduced activity and lower thermal stability	44, 45
R256W	Reduced blood enzyme activity (heterozygote with T344M)	Not tested	38
T288M	Reduced blood cell enzyme activity (heterozygote with frameshift)	Protein insoluble on expression in *E. coli*	24, 40
T344M	Reduced blood enzyme activity (heterozygote with R256W and homozygotes)	Not tested	38
G346S	No detectable blood enzyme activity (heterozygote with deletion)	k_{cat} and K_m for ATP reduced; K_m for galactose unaffected	24, 39
G349S	Reduced blood enzyme activity (homozygote); no detectable blood enzyme activity (heterozygote with H44Y)	k_{cat} reduced; modest increase in K_m for galactose; no change in K_m for ATP	24, 38, 39
A384P	Low blood enzyme activity (heterozygote with R68C)	Protein insoluble on expression in *E. coli*	24, 40

[a] Disease phenotype can be difficult to assess fully as patients who are identified as having low blood enzyme activity are placed on galactose-reduced diet.

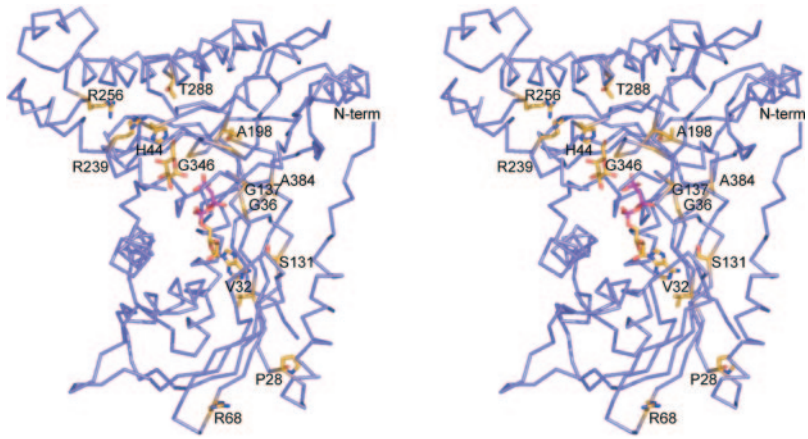


FIG. 5. **Mutations leading to Type II galactosemia.** Those positions where natural mutations in the human population have been shown to give rise to Type II galactosemia are indicated.

with the *P. furiosus* and the *L. lactis* enzymes, respectively. Shown in Fig. 4, *a* and *b* are superpositions of the models for these two bacterial enzymes onto the human galactokinase structure. The galactokinases from human and *P. furiosus* correspond with a root mean-square deviation of 1.4 Å for 310 structurally equivalent α-carbon atoms. The major difference is the absence of the first α-helix in the *P. furiosus* protein such that the human and bacterial proteins begin to superimpose at Ala^{31} and Thr^{5}, respectively. Additionally, the *P. furiosus* enzyme lacks the last β-strand at the C terminus and as a result the mixed β-sheet in its N-terminal domain contains five rather than six strands. The human and *L. lactis* galactokinases are more similar such that 353 equivalent α-carbons superimpose with a root mean-square deviation of 1.3 Å. The human and *L. lactis* polypeptide chains begin to match at Glu^{11} and Thr^{9}, respectively and the human enzyme contains two somewhat longer loop regions formed by Ser^{79}-Gln^{86} and Leu^{92}-Leu^{99}.

Implications for Type II Galactosemia—The lack of a functional galactokinase in humans results in Type II galactosemia (MIM 230200) (7–10). This relatively rare genetic disease (the incidence of homozygotes in Caucasians is estimated to be 1 in 1,000,000 (7)) causes an increased risk of early onset cataracts and, in more severe cases, results in kidney, liver and/or brain damage. A number of different mutations in the galactokinase gene (*GALK1*), as listed in Table III, have been shown to give rise to galactosemia including both deletions and point mutations (36–44). The locations of these mutations in human galactokinase are shown in Fig. 5. There is an apparent correlation between the biochemical consequence of the point mutation and the severity of the symptoms: mutations which give rise to the most severe symptoms, and near-zero blood galactokinase levels tend to be insoluble upon expression in *E. coli* suggesting improper folding (24, 44). Mutations that cause reduced blood galactokinase levels and less severe symptoms can be expressed in bacteria but have impaired kinetic parameters compared with the wild-type protein.

Those mutations that give rise to insoluble versions of the human protein on expression in *E. coli* include P28T, V32M,

G36R, T288M, and A384P. In wild-type galactokinase, the pyrrolidine ring of Pro^{28} abuts the side chain of Phe^{20} and is further surrounded by the side chains of Val^{64}, Leu^{392}, and the aliphatic portion of Arg^{17}. While it might be expected that substitution of a proline to a threonine would not cause considerable conformational changes given their similarity in size, most likely the introduction of the hydroxyl group of threonine into this quite hydrophobic pocket causes improper folding in this region. Val^{32}, in wild-type galactokinase, resides at the beginning of the first β-strand where its side chain projects into a pocket containing the side chains of Leu^{63} and Phe^{152}. There simply is not enough room in this region to accommodate the larger methionine side chain without major shifts in both backbone and side chain atoms. Likewise, Gly^{36} is situated at the end of the first β-strand where its α-carbon sits within 3.9 Å of the side chain of Val^{133} and projects into the pocket formed by the bulky side chains of Leu^{57}, Leu^{59}, and Leu^{139}. Substitution of an arginine for this glycine residue would have profound structural consequences. Thr^{288} is located near the middle of the α-helix formed by Lys^{272} to Arg^{297} where its side chain points toward a layer of β-sheet and is surrounded by the side chains of Ile^{41}, Leu^{51}, and Leu^{200}. As in the other cases described above, there is not enough room to accommodate the substitution to a methionine residue without major alterations in the protein backbone and/or side chains. The mutation of Ala^{384} to a proline is especially interesting. In the wild-type enzyme, Ala^{384} is positioned in a loop connecting the last two β-strands of the structure as can be seen in Fig. 5. This loop is quite critical for structural integrity in that it connects the N- and C-terminal domains together. Examination of the model in this region reveals no space constraints issues such as those for the other mutations described above. However, the phi, psi angles adopted by Ala^{384} in the wild-type protein are $-110°$ and $-174°$, respectively. These are incompatible for a proline in this position where its phi value is restricted to $\sim -60°$ due to its unique pyrrolidine ring side chain and, as such, the mutation of Ala^{384} to a proline would disrupt this loop region to a significant extent.

As listed in Table III, six naturally occurring mutations in galactokinase have been studied with regard to their kinetic parameters: H44Y, R68C, A198V, R239Q, G346S, and G349S. One of these mutant proteins, A198V, is soluble on expression in *E. coli* and has kinetic parameters, which are essentially indistinguishable from the wild-type (24). In contrast with the other mutations, where cataract formation is a near certainty within the first few years of life, this mutation merely results in a higher incidence of cataracts in later life (41). Ala^{198} resides in a β-strand and is $\sim$17 Å from the active site. Its C^{β} projects into a quite hydrophobic pocket formed by the side chains of Leu^{40}, Leu^{53}, Val^{211}, Leu^{213}, and Leu^{295}. Replacement of an alanine in this position with a valine could easily be tolerated with slight changes in the side chain orientations of Leu^{40}, Leu^{53}, and Leu^{213} while still maintaining the overall hydrophobic character of this region. It is thus not unexpected that this particular mutant protein demonstrates normal wild-type kinetic parameters. Of the remaining five mutations, those located near the active site include H44Y, G346S, and G349S. Both Gly^{346} and Gly^{349} belong to Motif III of the GHMP superfamily and it would be expected that any change would produce altered kinetic parameters. Indeed, in the model of human galactokinase presented here, the backbone peptidic NH group of Gly^{346} is perfectly positioned to form a hydrogen bond with a γ-phosphoryl oxygen of the AMPPNP ligand. With regard to His^{44}, this highly conserved residue forms part of the galactose binding site and the main chain carbonyl group is within hydrogen bonding distance of O-6 of the sugar (Fig. 3). Replacement of this residue with the bulkier tyrosine is likely to disrupt the local structure and may well result in steric hindrance in the active site, consequences of which would be entirely consistent with the kinetic changes observed in this mutant protein.

The remaining mutations presented in Table III, for which kinetic parameters have been measured, are R68C and R239Q. In the case of the R68C mutation, there are small reductions in both the k_{cat} and K_m for galactose, and an increase in K_m for MgATP. This arginine is located at the end of the second β-strand with its side chain solvent-accessible and is more than 22 Å from the active site. With the R239Q mutation, biochemical analyses indicate reduced activity and lower thermal stability. This residue is positioned at $\sim$11 Å from the galactose moiety in an α-helix defined by Glu^{235} to Ala^{249} (Fig. 5). It is in this region that there is an observed break in the polypeptide chain backbone (Ser^{230}-Ser^{233}) thus suggesting conformational flexibility in this area.

In conclusion, the three-dimensional structure of the ternary complex of human galactokinase with bound α-D-galactose and AMPPNP described here provides an unambiguous molecular scaffold upon which to more fully address the effects of mutations known to give rise to Type II galactosemia. Additionally, the model yields further insight into the active site geometry of this enzyme and poses new questions as to whether the catalytic mechanism of galactokinase is more similar to that proposed for homoserine kinase or mevalonate kinase. Further biochemical and structural analyses will be required to understand the intricacies of this fascinating small molecule kinase. These studies are currently in progress.

Acknowledgment—We thank Dr. Kit-Yin Ling for determining the initial crystallization conditions.

REFERENCES

1. Caputto, R., Leloir, L. F., Trucco, R. E., Cardini, C. E., and Paladini, A. C. (1949) *J. Biol. Chem.* **179,** 497–498
2. Caputto, R., Leloir, L. F., E., C. C., and Paladini, A. C. (1950) *J. Biol. Chem.* **184,** 333–350
3. Frey, P. A. (1996) *FASEB J.* **10,** 461–470
4. Holden, H. M., Rayment, I., and Thoden, J. B. (2003) *J. Biol. Chem.* **278,** 43885–43888
5. Wilkinson, J. F. (1949) *Biochem. J.* **44,** 460–467
6. Howard, S. M., and Heinrich, M. R. (1965) *Arch. Biochem. Biophys.* **110,** 395–400
7. Segal, S., and Berry, G. T. (1995) in *The Metabolic and Molecular Basis of Inherited Disease* (Scriver, C. R., Beaudet, A. L., Sly, W. S., and sValle, D., eds) pp. 967–1000, McGraw-Hill, New York
8. Petry, K. G., and Reichardt, J. K. (1998) *Trends Genet* **14,** 98–102
9. Novelli, G., and Reichardt, J. K. (2000) *Mol. Genet. Metab.* **71,** 62–65
10. Bosch, A. M., Bakker, H. D., van Gennip, A. H., van Kempen, J. V., Wanders, R. J., and Wijburg, F. A. (2002) *J. Inherit. Metab. Dis.* **25,** 629–634
11. Yabe-Nishimura, C. (1998) *Pharmacol. Rev.* **50,** 21–33
12. Ai, Y., Zheng, Z., O'Brien-Jenkins, A., Bernard, D. J., Wynshaw-Boris, T., Ning, C., Reynolds, R., Segal, S., Huang, K., and Stambolian, D. (2000) *Hum. Mol. Genet.* **9,** 1821–1827
13. Kinoshita, J. H., Merola, L. O., and Dikmak, E. (1962) *Biochim. Biophys. Acta* **62,** 176–178
14. Dvornik, E., Simard-Duquesne, N., Krami, M., Sestanj, K., Gabbay, K. H., Kinoshita, J. H., Varma, S. D., and Merola, L. O. (1973) *Science* **182,** 1146–1148
15. Lin, L. R., Reddy, V. N., Giblin, F. J., Kador, P. F., and Kinoshita, J. H. (1991) *Exp. Eye Res.* **52,** 93–100
16. Thoden, J. B., and Holden, H. M. (2003) *J. Biol. Chem.* **278,** 33305–33311
17. Hartley, A., Glynn, S. E., Barynin, V., Baker, P. J., Sedelnikova, S. E., Verhees, C., de Geus, D., van der Oost, J., Timson, D. J., Reece, R. J., and Rice, D. W. (2004) *J. Mol. Biol.* **337,** 387–398
18. Holden, H. M., Thoden, J. B., Timson, D. J., and Reece, R. J. (2004) *Cell. Mol. Life Sci.* **61,** 2471–2484
19. Bork, P., Sander, C., and Valencia, A. (1993) *Protein Sci.* **2,** 31–40
20. Fu, Z., Wang, M., Potter, D., Miziorko, H. M., and Kim, J. J. (2002) *J. Biol. Chem.* **277,** 18134–18142
21. Zhou, T., Daugherty, M., Grishin, N. V., Osterman, A. L., and Zhang, H. (2000) *Structure Fold Des.* **8,** 1247–1257
22. Krishna, S. S., Zhou, T., Daugherty, M., Osterman, A., and Zhang, H. (2001) *Biochemistry* **40,** 10810–10818
23. Timson, D. J., and Reece, R. J. (2002) *Biochimie* (*Paris*) **84,** 265–272
24. Timson, D. J., and Reece, R. J. (2003) *Eur. J. Biochem.* **270,** 1767–1774
25. Cornish-Bowden, A. (2004) *Fundamentals of Enzyme Kinetics*, pp. 209–229, Portland Press, London
26. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326

27. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–861
28. Terwilliger, T. C. (2000) *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972
29. Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987) *Acta Crystallogr. Sect. A.* **43,** 489–501
30. Wilson, D. B., and Hogness, D. S. (1969) *J. Biol. Chem.* **244,** 2137–2142
31. Timson, D. J., Ross, H. C., and Reece, R. J. (2002) *Biochem. J.* **363,** 515–520
32. Dey, P. M. (1983) *Eur. J. Biochem.* **136,** 155–159
33. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400
34. Timson, D. J., and Reece, R. J. (2003) *BMC Biochem.* **4,** 16
35. Yang, D., Shipman, L. W., Roessner, C. A., Scott, A. I., and Sacchettini, J. C. (2002) *J. Biol. Chem.* **277,** 9462–9467
36. Stambolian, D., Ai, Y., Sidjanin, D., Nesburn, K., Sathe, G., Rosenberg, M., and Bergsma, D. J. (1995) *Nat. Genet.* **10,** 307–312
37. Kalaydjieva, L., Perez-Lezaun, A., Angelicheva, D., Onengut, S., Dye, D., Bosshard, N. U., Jordanova, A., Savov, A., Yanakiev, P., Kremensky, I., Radeva, B., Hallmayer, J., Markov, A., Nedkova, V., Tournev, I., Aneva, L., and Gitzelmann, R. (1999) *Am. J. Hum. Genet.* **65,** 1299–1307
38. Asada, M., Okano, Y., Imamura, T., Suyama, I., Hase, Y., and Isshiki, G. (1999) *J. Hum. Genet.* **44,** 377–382
39. Kolosha, V., Anoia, E., de Cespedes, C., Gitzelmann, R., Shih, L., Casco, T., Saborio, M., Trejos, R., Buist, N., Tedesco, T., Skach, W., Mitelmann, O., Ledee, D., Huang, K., and Stambolian, D. (2000) *Hum. Mutat.* **15,** 447–453
40. Hunter, M., Angelicheva, D., Levy, H. L., Pueschel, S. M., and Kalaydjieva, L. (2001) *Hum. Mutat.* **17,** 77–78
41. Okano, Y., Asada, M., Fujimoto, A., Ohtake, A., Murayama, K., Hsiao, K. J., Choeh, K., Yang, Y., Cao, Q., Reichardt, J. K., Niihira, S., Imamura, T., and Yamano, T. (2001) *Am. J. Hum. Genet.* **68,** 1036–1042
42. Reich, S., Hennermann, J., Vetter, B., Neumann, L. M., Shin, Y. S., Soling, A., Monch, E., and Kulozik, A. E. (2002) *Pediatr. Res.* **51,** 598–601
43. Hunter, M., Heyer, E., Austerlitz, F., Angelicheva, D., Nedkova, V., Briones, P., Gata, A., de Pablo, R., Laszlo, A., Bosshard, N., Gitzelmann, R., Tordai, A., Kalmar, L., Szalai, C., Balogh, I., Lupu, C., Corches, A., Popa, G., Perez-Lezaun, A., and Kalaydjieva, L. V. (2002) *Pediatr. Res.* **51,** 602–606
44. Sangiuolo, F., Magnani, M., Stambolian, D., and Novelli, G. (2004) *Hum. Mutat.* **23,** 396
45. Magnani, M., Cucchiarini, L., Dacha, M., and Fornaini, G. (1982) *Hum. Hered.* **32,** 329–334