Molecular Structure of Galactokinase*

James B. Thoden and Hazel M. Holden‡

From the Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706

Galactokinase plays a key role in normal galactose metabolism by catalyzing the ATP-dependent phosphorylation of α-D-galactose to galactose 1-phosphate. In humans, mutations in the galactokinase gene can lead to the diseased state referred to as Type II galactosemia. Here we describe the three-dimensional structure of galactokinase from *Lactococcus lactis* determined to 2.1-Å resolution. As expected from amino acid sequence alignments, galactokinase adopts a similar topology to that observed for members of the GHMP superfamily. The N-terminal domain is characterized by a five-stranded mixed β-sheet while the C-terminal motif is dominated by two distinct four-stranded anti-parallel β-sheets. The structure was solved in the presence of α-D-galactose and inorganic phosphate. These ligands are wedged between the N- and C-terminal domains. Amino acid side chains responsible for anchoring the sugar ligand to the protein include Arg[36], Glu[42], Asp[45], Asp[183], and Tyr[233]. Both Arg[36] and Asp[183] are strictly conserved in the amino acid sequences available in the literature thus far for galactokinases. Interestingly, the carboxylate side chain of Asp[183] is positioned within 3.5 Å of the C-1 hydroxyl group of galactose, whereas the guanidinium group of Arg[36] is situated between both the C-1 hydroxyl group and the inorganic phosphate. Most likely these residues play key roles in catalysis. The structure of galactokinase described here serves as a model for understanding the functional consequences of point mutations known to result in Type II galactosemia in humans.

In most organisms, the conversion of galactose to glucose 1-phosphate is accomplished by a series of enzymes belonging to the Leloir pathway as indicated in Scheme 1. Deficiencies in three of these enzymes in humans can result in the diseased state referred to as galactosemia (1, 2). Although galactosemia is rare, it is a potentially lethal genetic disease with clinical manifestations, including intellectual retardation, liver dysfunction, and cataract formation, among others. At present, the high resolution x-ray crystallographic structures of three of the four enzymes in the pathway, namely galactose mutarotase, galactose-1-phosphate uridylyltransferase, and UDP-galactose 4-epimerase, are known through efforts by this laboratory (3–9). The remaining enzyme of unknown structure, galactokinase, catalyzes the second step of the pathway, the ATP-dependent conversion of α-D-galactose to galactose 1-phosphate.

As has been the case for all of the enzymes belonging to the Leloir pathway, galactokinase has been the subject of intensive biochemical investigations for many years. Interestingly, the kinetic properties of galactokinase seem to differ according to the source of the protein. In the enzyme isolated from *Escherichia coli*, it appears that the reaction mechanism is random with either ATP or galactose binding first (10). In the rat, yeast, and human enzymes, however, it appears that the reaction mechanism is ordered with ATP binding first (11–14). Strikingly, in plant galactokinases, the reaction mechanism is reported to be ordered but with galactose rather than ATP binding in the first step (15).

In recent years it has been shown on the basis of amino acid sequence alignments that galactokinase belongs to a unique class of ATP-dependent enzymes known as the GHMP[1] superfamily (the abbreviation of which refers to the original members: galactokinase, homoserine kinase, mevalonate kinase, and phosphomevalonate kinase) (16). Enzymes belonging to this superfamily contain three well-conserved motifs, the second of which has the typical sequence Pro-X-X-X-Gly-Leu-X-Ser-Ser-Ala and is involved in ATP binding (17). Within the last three years, the three-dimensional architectures of homoserine kinase, mevalonate kinase, and phosphomevalonate kinase have been elucidated (17–21). Homoserine kinase plays a critical role in bacterial threonine biosynthesis by catalyzing the phosphorylation of L-homoserine to O-phospho-L-homoserine. Mevalonate kinase catalyzes the phosphorylation of the C-5 hydroxyl oxygen of mevalonic acid to yield mevalonate 5-phosphate, a key intermediate employed for the synthesis of isoprenoids and sterols. Likewise, phosphomevalonate kinase is involved in isoprenoid biosynthesis by promoting the transfer of the γ-phosphoryl group of ATP to the phosphate oxygen of (R)-5-phosphomevalonate to yield (R)-5-pyrophosphomevalonate. All of these enzymes display a distinctly bilobal appearance with their N-terminal motifs dominated by a mixed β-sheet flanked on one side by α-helices and their C-terminal domains containing a four stranded anti-parallel β-sheet.

Here we describe the x-ray crystallographic structure of galactokinase determined to a nominal resolution of 2.1 Å. For this study, the gene encoding the enzyme from *Lactococcus lactis* was cloned, the protein was overexpressed and purified from *E. coli*, and crystals were grown in the presence of α-D-galactose and inorganic phosphate. The enzyme from *L. lactis* contains 399 amino acid residues. From this investigation, the

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

[1] The abbreviations used are: GHMP, superfamily of galactokinase, homoserine kinase, mevalonate kinase, and phosphomevalonate kinase; Ni-NTA, nickel-nitrilotriacetic acid; CHES, 2-(cyclohexylamino)-ethanesulfonic acid; AMPPNP, 5′-adenylyl-β,γ-imidodiphosphate.
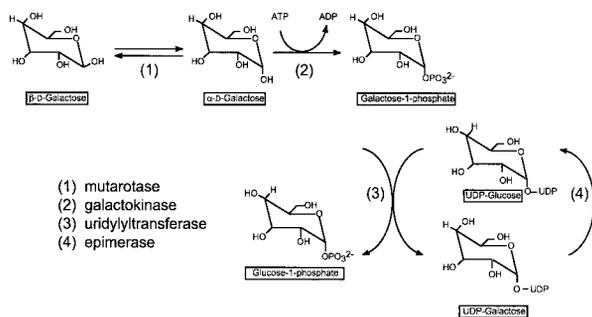
SCHEME 1

overall fold of galactokinase has now been defined, and the location of the active site was identified. Key residues involved in galactose binding include Arg[36], Glu[42], Asp[45], Asp[183], and Tyr[233]. The structure of the enzyme is discussed in context with other members of the GHMP superfamily. Additionally, the bacterial and human forms of galactokinase are 34% identical and 47% similar. As such, this structure provides insight into the molecular basis for Type II galactosemia in humans.

EXPERIMENTAL PROCEDURES

Cloning of the Galactokinase Gene—Genomic DNA from *L. lactis var. lactis* was prepared as previously reported (3). The galactokinase gene was PCR-amplified from genomic DNA such that the forward primer 5′-GGAATTCATATGTCTATAGTTGTAGAAAATAGTACCGTA-3′ and the reverse primer 5′-CCGCTCGAGTTATTCACCATCCAACTTT-GTTGAACCAGA-3′ added *Nde*I and *Xho*I cloning sites, respectively. The gene was PCR-amplified with Platinum *Pfx* DNA polymerase (Invitrogen) according to the manufacturer's instructions and standard cycling conditions. The PCR product was purified with the QIAquick PCR Purification kit (Qiagen), followed by A-tailing and ligation into pGEM-T Vector (Promega) and transformation into *E. coli* DH5α cells. The galactokinase gene was subsequently sequenced in the pGEM-T vector construct with the ABI Prism™ Big Dye Primer Cycle Sequencing kit (Applied Biosystems, Inc.) to confirm that no mutations were introduced during PCR amplification. The pGEM-T vector construct was then digested with *Nde*I and *Xho*I, and the gene was separated from digestion by-products on a 1.0% agarose gel. Subsequently, the galactokinase gene was excised from the gel, purified with a QIAquick Gel Purification kit (Qiagen), and ligated into the expression vector pET-28b(+) (Novagen) that was previously cut with *Nde*I and *Xho*I. Insertion of the galactokinase gene into the pET-28b(+) plasmid resulted in a construct containing an additional 20 amino acid residues at the N terminus with the sequence MGSSHHHHHHSSGLVPRGSH, thereby allowing for subsequent purification utilizing Ni-NTA affinity resin. *E. coli* DH5α cells were transformed with the ligation mixture and subsequently plated onto LB media supplemented with 30 μg/ml kanamycin. Individual colonies were selected and cultured overnight, and plasmid DNA was extracted with the QIAprep Spin Miniprep kit (Qiagen). Plasmids were tested for insertion of the galactokinase gene by digesting with *Nde*I and *Xho*I.

Protein Expression—For protein expression, *E. coli* Rosetta(DE3) (Novagen) cells were transformed with the pET28b-*galk* plasmid and plated onto LB agarose plates supplemented with 30 μg/ml kanamycin and 30 μg/ml chloramphenicol. Single colonies from the plates were selected and grown overnight in 500 ml of LB media (plus kanamycin and chloramphenicol) for the inoculation of 12- × 2-liter baffled flasks containing 500 ml of M9 minimal medium (30 μg/ml kanamycin and 30 μg/ml chloramphenicol, plus antibiotics). Cells were grown at 37 °C with aeration to an A_{600} of ~0.9, at which time they were cooled on ice for 5 min and transferred to a 16 °C shaker before inducing with 1 mM isopropyl-1-thio-β-D-galactopyranoside. The cells were allowed to grow for an additional 18 h before harvesting by centrifugation at 6000 × g for 8 min. Cell paste was frozen in liquid nitrogen and stored at −80 °C.

Expression of the Selenomethionine-labeled Protein—Cultures of *E. coli* Rosetta(DE3) cells containing the pET28b-*galk* plasmid were grown overnight in M9 minimal medium (plus kanamycin and chloramphenicol) at 37 °C. Subsequently 15 ml of the overnight culture was used to inoculate each of the 12- × 2-liter baffled flasks containing 500 ml M9 minimal medium supplemented with 5 μg/ml thiamine, 30 μg/ml kanamycin, and 30 μg/ml chloramphenicol. Cultures were grown at 37 °C to an A_{600} of ~0.9. Subsequently, the flasks were cooled on ice

for 5 min, and the temperature of the incubator was adjusted to 16 °C before each flask was supplemented with 50 mg each of L-lysine, L-threonine, and L-phenylalanine, and 25 mgs each of L-leucine, L-isoleucine, L-valine, and L-selenomethionine (22). After 20 min of additional growth, the cells were induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside and allowed to grow for 18 h. Cultures were harvested by centrifugation at 6000 × g for 8 min, and the cell paste was frozen in liquid nitrogen for storage at −80 °C.

Protein Purification—All protein purification steps were carried out on ice or at 4 °C. The cell paste was thawed in three volumes of cold Ni-NTA lysis buffer containing 50 mM NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole at pH 8.0. Cells were lysed on ice by four cycles of sonication (45 s) separated by 5 min of cooling. Cellular debris was removed by centrifugation at 4 °C for 25 min at 20,000 × g. The clarified lysate was loaded onto a 10-ml Ni-NTA-agarose (Qiagen) column pre-equilibrated with Ni-NTA lysis buffer. After loading, the column was washed with ~100 ml of Ni-NTA lysis buffer, followed by a wash with 100 ml of Ni-NTA wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, and 20 mM imidazole at pH 8.0). The protein was recovered by gradient elution with 20–300 mM imidazole in Ni-NTA lysis buffer. Protein-containing fractions were pooled based on SDS-PAGE and dialyzed against 10 mM Hepes (pH 7.5) containing 200 mM NaCl. The dialyzed protein was concentrated to ~15 mg/ml, based on an extinction coefficient of ~1.75 ml/(mg cm) as calculated by the program Protean (DNAstar, Inc., Madison, WI). Galactose-dependent ATPase activity was verified using the standard pyruvate kinase/lactate dehydrogenase-coupled assay.

Crystallization of Wild-type and Selenomethionine-labeled Galactokinase—Potential crystallization conditions were examined with a sparse matrix screen, designed in this laboratory and composed of 144 conditions. Crystallization trials were conducted at both room temperature and 4 °C via the hanging drop method of vapor diffusion. Protein samples containing 10 mM D-(+)-galactose were utilized for the crystallization experiments. Single crystals were observed growing at room temperature at pH 9 with 2.4 M Na⁺/K⁺ phosphate as the precipitant. Refinement of the crystallization conditions via macro-seeding into batch experiments resulted in crystals being routinely grown at room temperature with precipitant concentrations of 1.6–1.8 M Na⁺/K⁺ phosphate and 100 mM CHES (pH 9.0). These crystals achieved typical dimensions of ~0.6 × 0.6 × 0.6 mm in 3–6 weeks and belonged to the cubic space group I23 with unit cell dimensions of $a = b = c = 143.4$ Å and one molecule per asymmetric unit. Crystals of the selenomethionine-labeled protein were grown in the same manner, using seeds of the native enzyme to initiate crystallization.

High Resolution x-ray Data Collection—Both the wild-type and selenomethionine-substituted protein crystals were flash-frozen in the same manner. Briefly, crystals were harvested from the batch experiments and soaked for several hours in a synthetic mother liquor composed of 2.15 M Na⁺/K⁺ phosphate, 10 mM D-(+)-galactose, and 100 mM CHES (pH 9.0). These crystals were then serially transferred to a cryoprotectant solution composed of 3.0 M Na⁺/K⁺ phosphate, 10 mM D-(+)-galactose, 15% ethylene glycol, and 100 mM CHES (pH 9.0). Subsequently, the crystals were flash-cooled to −150 °C in a stream of nitrogen gas. Frozen crystals were stored under liquid nitrogen until synchrotron beam time became available. X-ray data from crystals of the selenomethionine-substituted protein were collected on the SBC3 CCD detector at the Structural Biology Center 19-BM beamline. X-ray data from crystals of the wild-type protein were collected with the SBC2 CCD detector on the19-ID beamline (Advanced Photon Source, Argonne National Laboratory, Argonne, IL). All x-ray data sets were processed with HKL2000 and scaled with SCALEPACK (23). Relevant x-ray data collection statistics are presented in Table I.

X-ray Structural Analyses—The structure of galactokinase was solved via multiple wavelength anomalous dispersion phasing with x-ray data collected from crystals of the selenomethionine-substituted protein. The software package SOLVE was utilized to determine and refine the positions of six selenium atoms (24). Solvent flattening with the software package DM resulted in an interpretable electron density map calculated to 2.6-Å resolution (25). This initial electron density map allowed for the tracing of residues Thr[9] to Ala[396]. The remaining three residues at the C terminus were not visible. Note that the electron density corresponding to the first 28 residues at the N terminus was not as well ordered as that observed for other portions of the map. Additionally, the His tag at the N terminus was not visible in the electron density map, and it is possible that this tag contributed to the observed conformational flexibility of the molecule in this region.

The initial model of galactokinase was subsequently refined against x-ray data collected from crystals of the wild-type protein. Alternate cycles of least-squares refinement with the software package TNT (26)

TABLE I
X-ray data collection statistics

	Wavelength	Resolution	No. independent reflections	Completeness	Redundancy	Avg. I/Avg. $\sigma(I)$	$R_{sym}{}^a$
	Å			%			%
Peak	0.97932	30.0–2.54	16126	99.9	14.4	64.0	7.8
		2.63–2.54^b	1606	99.8	10.7	7.6	28.0
Inflection	0.97951	30.0–2.54	16253	99.4	13.7	53.1	9.5
		2.63–2.54	1536	94.9	9.8	4.0	35.1
Remote	0.96411	30.0–2.54	16207	99.9	14.6	54.5	6.5
		2.63–2.54	1598	99.9	10.5	5.7	30.4
Native	0.65255	20.0–2.10	28616	98.7	14.0	75.4	6.7
		2.17–2.10	2804	99.5	9.7	9.3	21.4

^{a} $R_{sym} = (\Sigma|I - \bar{I}|\Sigma I) \times 100$.
^{b} Statistics for the highest resolution bin.

TABLE II
Least-squares refinement statistics

Resolution limits (Å)	20.0–2.10
R-factor (overall) %/no. reflections^a	20.4/28,536
R-factor (working) %/no. reflections	19.9/25,673
R-factor (free) %/no. reflections	26.6/2,896
No. protein atoms^b	2985
No. hetero-atoms^c	181
Average *B* values (Å^2)	
Protein atoms	50.4
Galactose	30.4
Phosphate	72.5
Waters	60.6
Weighted RMS deviations from ideality	
Bond lengths (Å)	0.013
Bond angles (deg)	2.59
Trigonal planes (Å)	0.006
General planes (Å)	0.012
Torsional angles (deg)^d	18.1

^{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 N^{155}.
^{c} These include 1 galactose molecule, 1 phosphate ion, and 164 waters.
^{d} The torsional angles were not restrained during the refinement.

and manual model-building with the graphics program TURBO (27) reduced the *R*-factor to 20.4% for all measured x-ray data from 20- to 2.1-Å resolution. Relevant least-squares refinement statistics are summarized in Table II. A Ramachandran plot for all non-glycinyl ϕ,ψ angles is given in Fig. 1*a*. There are only two residues that lie substantially outside of the allowed region of the Ramachandran plot, Arg^{213} ($\phi = 69.4°$, $\psi = -126.5°$) and Thr^{323} ($\phi = -90.3°$, $\psi = -94.0°$). The electron density for both of these residues is unambiguous. Both Arg^{213} and Thr^{323} are far removed from the active site cleft. A representative portion of the electron density map is displayed in Fig. 1*b*. From the electron density it is clear that α-D-galactose was trapped in the active site.

RESULTS AND DISCUSSION

Structure of Galactokinase—The quaternary structures of the galactokinases obtained from *E. coli*, yeast, and human are monomeric, whereas that isolated from *Vicia faba* seeds is dimeric (14, 15, 28, 29). Examination of the packing arrangement of the *L. lactis* enzyme in the crystalline lattice suggests that it is also monomeric. Shown in Fig. 2 is a ribbon representation of the *L. lactis* enzyme. The molecule is elongated with overall dimensions of ~70 × 54 × 57 Å and folds into two domains of roughly equal size. The N-terminal domain is delineated by Thr^9 to Gly^{190}, whereas the C-terminal domain is formed by Glu^{191} to Gly^{390}. In addition, the last amino acid residues, Ser^{391} to Ala^{396}, curl back onto the N-terminal domain. The N-terminal domain initiates with an α-helix formed by Thr^9 to Phe^{22}. Following this α-helical region, the polypeptide chain folds into four β-strands, referred to as A–D and defined by Glu^{29} to Ile^{40}, Gly^{49} to Leu^{66}, Lys^{71} to Thr^{75}, and

Ile^{85} to Phe^{87}. The A–B and C–D pairs run anti-parallel. After β-strand D, the polypeptide chain adopts an α-helical conformation from Ser^{101} to Leu^{111}, which is followed by the fifth β-strand of the N-terminal domain, β-strand E (Gly^{121} to Gly^{128}). This β-strand is situated between the A–B and C–D pairs such that it runs anti-parallel with β-strand B and parallel to β-strand C. There are an additional three α-helices (Ser^{137} to Asp^{151}, Pro^{159} to Ile^{175}, and Leu^{182} to Phe^{189}) that flank one side of this five-stranded mixed β-sheet. Amino acid residues involved in the binding of the inorganic phosphate are located in the loop connecting β-strand E to the following α-helix.

The C-terminal domain contains two regions of β-sheet. The first is a hairpin motif formed by β-strands F (Ala^{195} to Leu^{198}) and G (Ala^{204} to Val^{210}). These strands are connected by a Type III reverse turn (Asp^{199} to Thr^{202}). In combination with portions of β-strands A and B, this β-hairpin contributes in the formation of a four-stranded anti-parallel β-sheet. The second major tertiary element in the C-terminal domain is a four-stranded anti-parallel β-sheet formed by β-strands H–K (Asp^{216} to Asn^{221}, Val^{340} to Thr^{346}, Cys^{353} to Ala^{359}, and Ser^{383} to Ala^{387}). The β-strand H is connected to β-strand I via five α-helices (Ser^{231} to Arg^{250}, Asn^{261} to Ala^{266}, Glu^{274} to Ala^{299}, Leu^{302} to Asp^{319}, and Leu^{325} to Lys^{336}), whereas β-strands J and K are linked by an α-helix defined by Val^{363} to Val^{378}. The active site is wedged between the N- and C-terminal domains.

Active Site Geometry—A close-up view of the galactokinase active site is shown in Fig. 3. Note that Gly^{347} and Ala^{348} abut one side of the sugar ligand. These residues are located in Motif III of the GHMP superfamily. For the *L. lactis* enzyme, these residues link β-strands I and J and have the following sequence: Gly^{347}-Ala^{348}-Gly^{349}-Phe^{350}-Gly^{351}-Gly^{352}-Cys^{353}.

The galactose moiety is anchored to the protein via the guanidinium group of Arg^{36}, the carboxylate groups of Glu^{42}, Asp^{45}, and Asp^{183}, and O$^\eta$ of Tyr^{233}. There are two water molecules that lie within 3.2 Å of the 3- and 6-hydroxyl groups of the sugar ligand. The backbone peptidic groups of His^{43} and Gly^{180} provide hydrogen-bonding interactions with the 6- and 3-hydroxyl groups of galactose, respectively. The inorganic phosphate group is situated near Ser^{137}. This residue is part of the conserved Motif II of the GHMP superfamily which, for the galactokinases, has the typical sequence Pro-*X*-*X*-*X*-Gly-Leu-*X*-Ser-*X*-*X* (31). In the *L. lactis* enzyme, the sequence is Pro^{131}-Thr^{132}-Ala^{133}-Ser^{134}-Gly^{135}-Leu^{136}-Ser^{137}-Ser^{138}-Ser^{139}-Ala^{140}. Pro^{131} initiates a Type I turn (Pro^{131} to Ser^{134}) that connects β-strand E to the third α-helix of the N-terminal domain (Fig. 2).

Quite strikingly, the 1-hydroxyl group of galactose, which is phosphorylated during the reaction, is positioned within ~3.1 Å of two of the phosphoryl oxygens of the inorganic phosphate and 3.0 Å of N$^{\eta 2}$ of Arg^{36}. The carboxylate side chain of Asp^{183} is located at 3.5 Å from the galactose 1-hydroxyl group, 3.3 Å
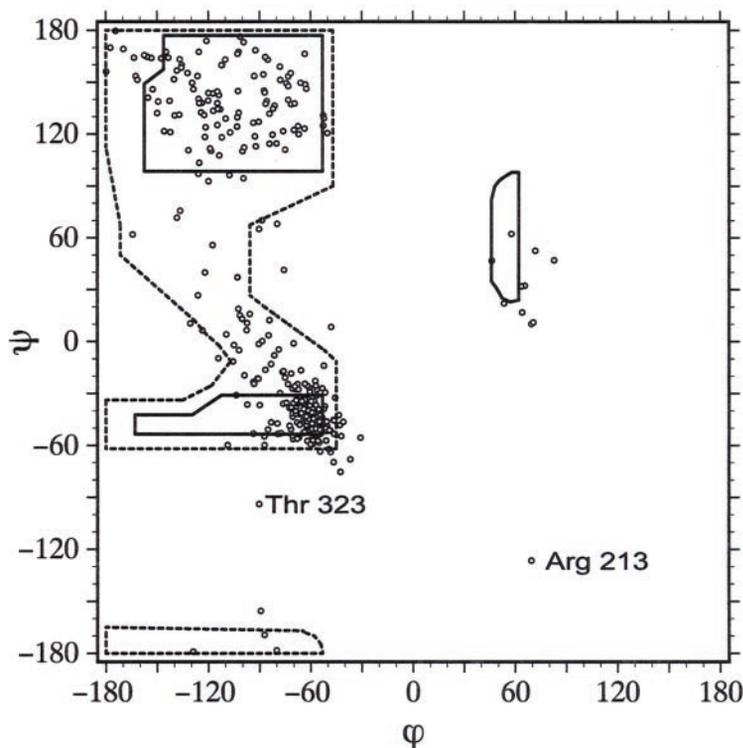
FIG. 1. **Quality of the x-ray model.** The structure of galactokinase presented here was solved to 2.1-Å resolution. Shown in *a* is a Ramachandran plot of all non-glycinyl main chain ϕ,ψ angles. Those regions that are fully or partially allowed are indicated by the *solid* and *dashed lines*, respectively. *b*, electron density corresponding to the region near the active site. The map was calculated with coefficients of the form $(F_o - F_c)$, where F_o was the native structure factor amplitude and F_c was the calculated structure factor amplitude from the model lacking the coordinates for the ligand. The map was contoured at 3σ. All figures were prepared with the software package MOL-SCRIPT (31).
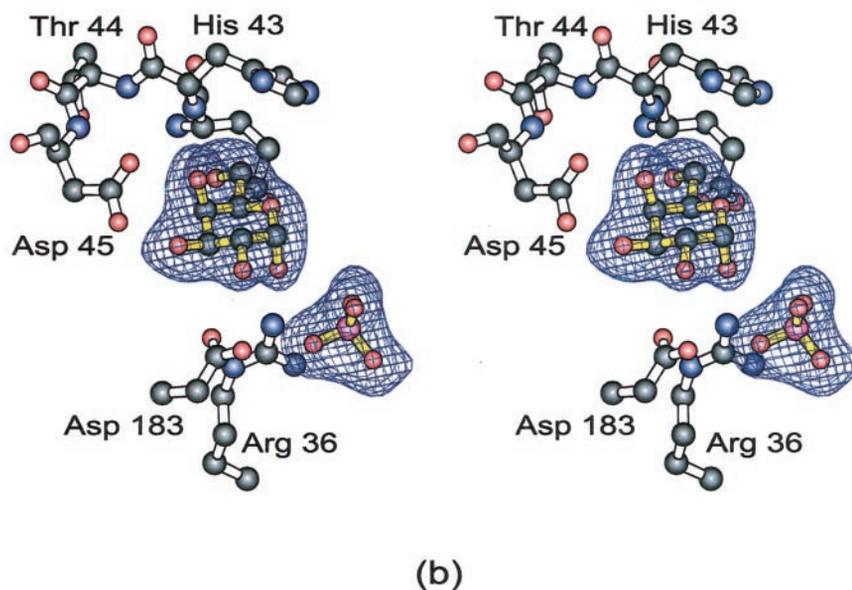
from the 2-hydroxyl group, and 2.7 Å from one of the inorganic phosphate oxygens. This residue is strictly conserved in 56 of the galactokinase primary structures deposited in the SwissProt data bank. Arg[36], as observed in the *L. lactis* galactokinase, is also strictly conserved in the 56-amino acid sequences deposited in the SwissProt data bank and is part of the conserved Motif I in the GMHP superfamily. In the *L. lactis* enzyme, Motif I has the following primary structure: Pro[34]-Gly[35]-Arg[36]-Ile[37]-Asn[38]-Leu[39]-Ile[40]-Gly[41] and these residues form β-strand A (Fig. 2). Based on the observed geometry within the active site and the fact that these residues are

strictly conserved, it can be speculated that Asp[183] and Arg[36] play key roles in catalysis.

Comparison with Another Member of the GHMP Kinase Superfamily—As expected from amino acid sequence analyses, the overall topology of galactokinase is similar to that observed for homoserine kinase, mevalonate kinase, and phosphomevalonate kinase. A superposition of galactokinase onto homoserine kinase (complexed with AMPPNP and homoserine) is displayed in Fig. 4a. These two enzymes superimpose with a root mean square deviation of 2.4 Å for 259 structurally equivalent α-carbon positions. Homoserine kinase from *Methano-*

FIG. 2. **Ribbon representation of galactokinase.** Shown is a ribbon representation of one subunit of galactokinase with the bound ligands, α-D-galactose and phosphate, depicted in *ball-and-stick* representations.
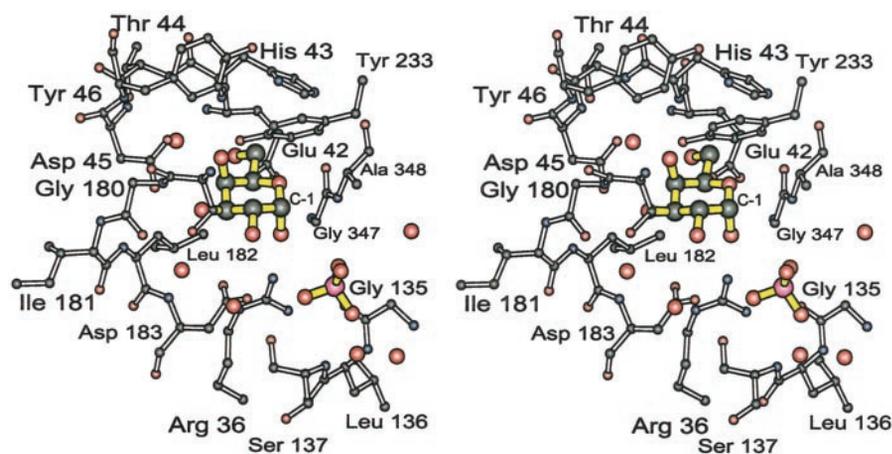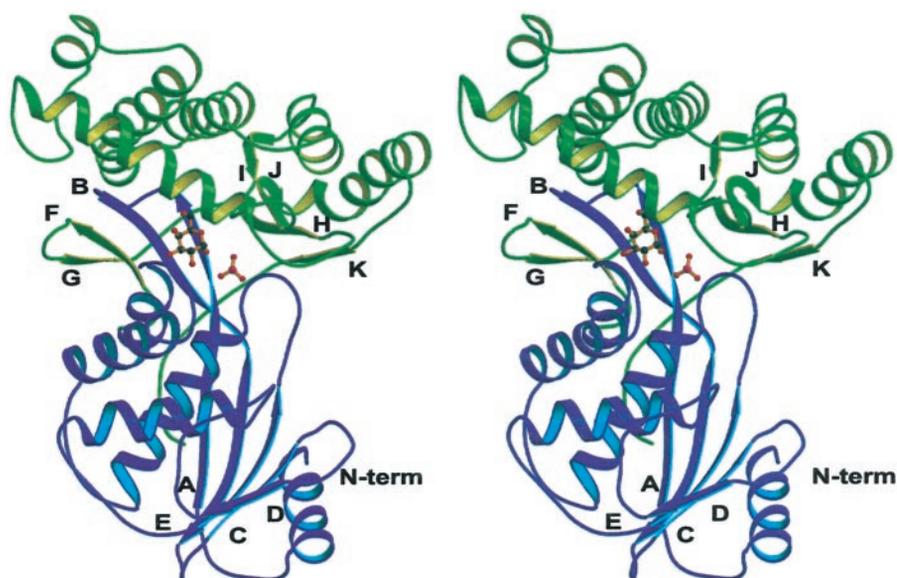

FIG. 3. **Close-up view of the galactokinase active site.** A close-up view of the active site, within ~3.2 Å of the galactose and inorganic phosphate molecules, is shown. The ligands are highlighted in *yellow-filled* bonds.

coccus jannaschii contains 300 amino acids compared to the *L. lactis* galactokinase with 399 residues. In galactokinase, the first 26 residues fold into an α-helical region and a short loop connecting to the first β-strand. This structural motif is missing in homoserine kinase. As such the two proteins begin to align at Glu[27] in galactokinase and Met[5] in homoserine kinase. Excluding this difference in the N-terminal regions, there are basically six areas where the polypeptide chains differ considerably between these two proteins. The first difference occurs at Asn[38] in galactokinase where there is a five-residue insertion, relative to homoserine kinase. This region connects β-strands A and B. The next difference occurs in galactokinase beginning at Asn[78] where there is a 14-residue insertion relative to that observed in homoserine kinase. This region lies between β-strands C and D (Fig. 4a). The third region where the polypeptide chains diverge occurs near the end of the N-terminal domain beginning at Phe[189] in galactokinase and Phe[147] in homoserine kinase. Although there is only a four-residue insertion in galactokinase, the polypeptide chains adopt quite different conformations in these two GHMP family members. A three-residue deletion occurs in galactokinase, relative to homoserine kinase, beginning at Cys[200]. This portion of the polypeptide chain connects β-strands F and G. The largest insertion in galactose kinase, relative to homoserine kinase, occurs at Asn[234] and contains 38 extra residues. Due to this insertion, there is an additional α-helix in galactokinase formed by Asn[261] to Ala[266]. Finally, there is a six-residue insertion in galactokinase near the C terminus beginning at

Tyr[374]. The polypeptide chain in this region provides the connection from the last β-strand to the last α-helix of the enzyme.

The inorganic phosphate observed in the galactokinase structure occupies a similar position to that of the γ-phosphate of AMPPNP in homoserine kinase. The hydroxyl groups that are phosphorylated by homoserine kinase and galactokinase occupy similar positions within the active site as well. Additionally, the distance between the hydroxyl group of homoserine and the γ-phosphorus of AMPPNP is 3.4 Å. This separation is comparable to the distance of 3.5 Å observed in galactokinase between the C-1 hydroxyl group of the ligand and the inorganic phosphate. It can thus be speculated that in galactokinase the γ-phosphate of ATP would occupy a similar position to that observed for the inorganic phosphate, and indeed, the C-1 hydroxyl group of galactose is in the proper position for a single in-line displacement reaction.

Implications for Galactosemia—Impairment of three of the enzymes participating in the Leloir pathway results in the diseased state, galactosemia (2, 30). Although relatively rare, this autosomal recessive disease can be life-threatening by causing high levels of galactose in the body thereby leading to such clinical manifestations as jaundice, liver enlargement, cataracts, and brain damage, among others. There are three types of galactosemia classified thus far, namely Types I, II, and III. The most common and severe form is galactosemia I or classic galactosemia, which results from impairment of the galactose-1-phosphate uridylyltransferase in the Leloir pathway (Scheme 1). At least 30 different known mutations have
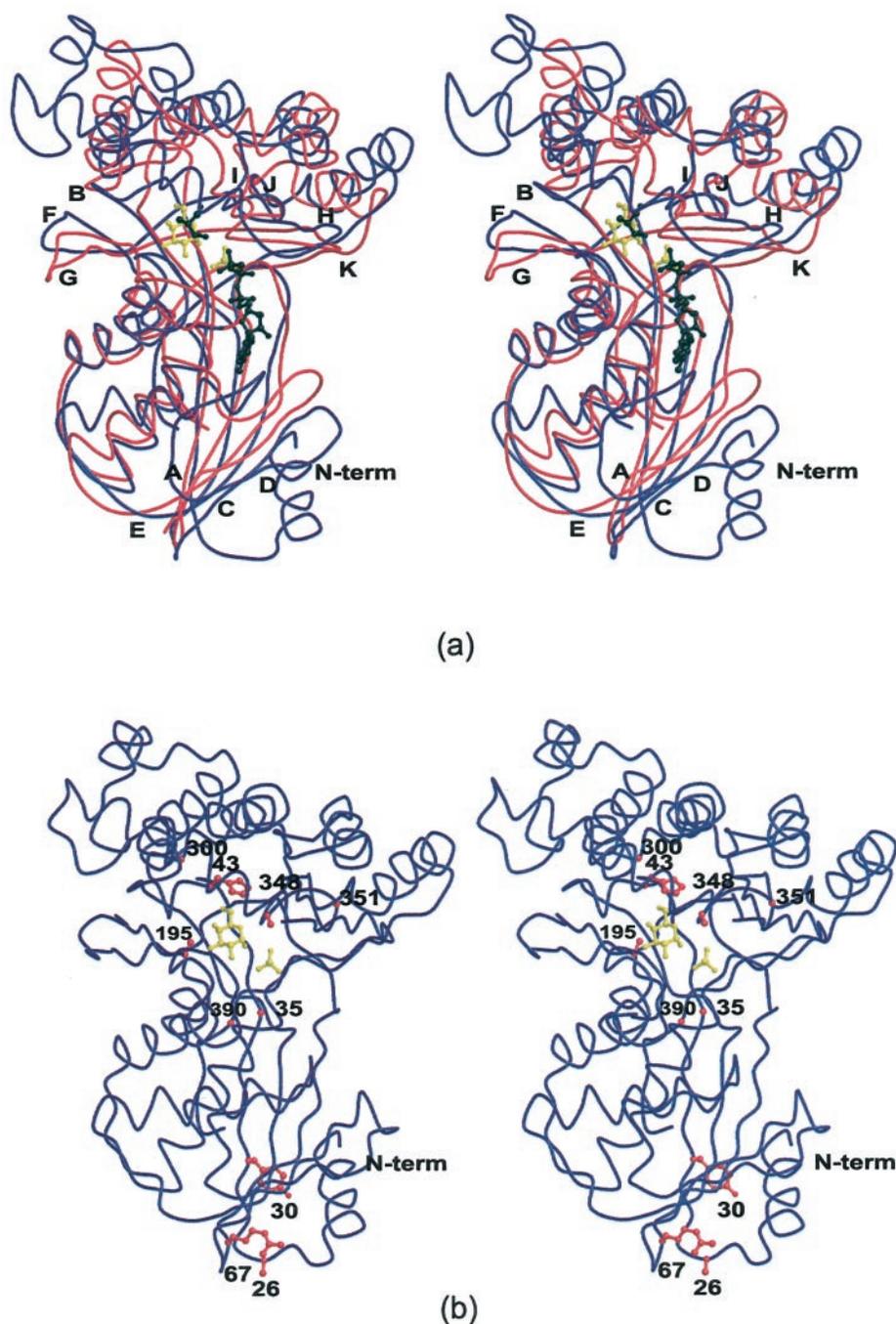
FIG. 4. **Superposition of the Alpha-carbon traces for galactokinase and homoserine kinase.** The α-carbon traces for galactokinase and homoserine kinase are drawn in *blue* and *red*, respectively in *a*. The bound galactose and inorganic phosphate in galactokinase are displayed in *yellow*, whereas the homoserine and AMPPNP moieties in homoserine kinase are depicted in *black*. Shown in *b* is an α-carbon trace of the *L. lactis* enzyme with the amino acid residue positions known to cause Type II galactosemia in humans indicated by the *ball-and-stick* representations.

been identified thus far for galactosemia I with symptoms, including liver and renal dysfunction, cataracts, and abnormal neurological development. Type II galactosemia results from galactokinase deficiency and is considered relatively benign. Patients with galactosemia II do not generally suffer from liver damage or neurological defects and, if treated with galactose-free diets, will not develop cataracts. Galactosemia III, resulting from impairment of the enzyme UDP-galactose 4-epimerase, is the most rare form of the disease. Mutations in the epimerase gene can lead to either a benign form of the disease where no treatment is required or a severe form with symptoms similar to those of galactosemia I, including seizures.

Recently ten mutations known to cause Type II galactosemia in humans were constructed, expressed, and characterized in *E. coli* (14). These included P28T, V32M, G36R, T288M, A384P, H44Y, R68C, G346S, G349S, and A198V. On the basis of amino

acid sequence alignments, the equivalent positions in the *L. lactis* galactokinase are at 26, 30, 35, 300, 390, 43, 67, 348, 351, and 195, respectively. Since the bacterial and human forms of galactokinase are 34% identical with respect to amino acid sequence, it is possible to speculate, with caution, as to where these residues might be located on the human enzyme. As can be seen in Fig. 4*b*, the ten human galactosemic mutations, mapped onto the *L. lactis* enzyme model, fall into three categories: three that are centered in the N-terminal region, four that are located directly within the active site region, and three that reside in the C-terminal domain. Of the ten mutations identified in patients with Type II galactosemia, the P28T, V32M, G36R, T288M, and A384P mutant proteins (which correspond to positions 26, 30, 35, 300, and 390 in Fig. 4*b*) were not soluble in the *E. coli* expression systems employed for the investigation thus suggesting that these proteins were

improperly folded *in vivo* (14). Four of the other mutant proteins studied, H44Y, R68C, G346S, and G349S (corresponding to positions 43, 67, 348, and 351 in Fig. 4*b*) were less active than the wild-type protein. From the present study, it can be predicted that in the human form of galactokinase, His[44] (His[43] in the *L. lactis* protein) is located near the sugar binding pocket. The measured K_m,gal for the H44Y mutant protein increased from 970 μM for the wild-type enzyme to 7700 μM, whereas the K_m,ATP increased from 34 to 130 μM (14). For the wild-type human galactokinase, the reported k_{cat} is 8.7 s^{-1}. In the H44Y mutant protein, the k_{cat} is reduced to 2.0 s^{-1} (14). These changes in the kinetic parameters are most likely due to structural perturbations in the active site due to the replacement of the side chain imidazole group with a more bulky aromatic moiety.

For the R68C mutant protein, the reported K_m,gal and K_m,ATP values are 430 and 110 μM, respectively, and the k_{cat} is 3.9 s^{-1}. On the basis of the *L. lactis* enzyme model, this residue is located at ~31 Å from the active site suggesting that the observed kinetic effects are due to changes in the overall structure of the enzyme and not to a direct protein/substrate interactions. Both the G346S and G349S mutant human galactokinases display altered kinetic properties as well (14). The structurally homologous residues in the bacterial enzyme, Ala[348] and Gly[351], form part of Motif III. Interestingly, the A198V mutation in the human enzyme demonstrates kinetic properties similar to the wild-type enzyme. In the *L. lactis* enzyme, the equivalent residue, Ala[195], is located at ~17 Å from the active site.

In conclusion, the structure of galactokinase from *L. lactis* has now been determined to 2.1-Å resolution. The polypeptide chain folds into two domains of approximately equal size, and the overall topology is virtually identical to that observed for members of the GHMP superfamily. Key amino acid side chains involved in galactose binding include Arg[36], Glu[42], Asp[45], Asp[183], and Tyr[233]. Both Arg[36] and Asp[183] are strictly conserved in the galactokinase amino acid sequences examined to date. On the basis of the observed binding positions for both the galactose and inorganic phosphate ligands, it can be speculated that Asp[183] serves as the catalytic base to deprotonate the 1-hydroxyl group of the sugar substrate and that Arg[36] aids in the reaction mechanism by lowering the pK_a of the 1-hydroxyl group. Experiments designed to further investigate the role of these amino acid residues in the reaction mechanism of galactokinase are in progress.

REFERENCES

1. Petry, K. G., and Reichardt, J. K. (1998) *Trends Genet.* **14,** 98–102
2. Novelli, G., and Reichardt, J. K. (2000) *Mol. Genet. Metab.* **71,** 62–65
3. Thoden, J. B., and Holden, H. M. (2002) *J. Biol. Chem.* **277,** 20854–20861
4. Thoden, J. B., Kim, J., Raushel, F. M., and Holden, H. M. (2002) *J. Biol. Chem.* **277,** 45458–45465
5. Thoden, J. B., Kim, J., Raushel, F. M., and Holden, H. M. (2002) *Protein Sci.* **12,** 1051–1059
6. Wedekind, J. E., Frey, P. A., and Rayment, I. (1995) *Biochemistry* **34,** 11049–11061
7. Thoden, J. B., Ruzicka, F. J., Frey, P. A., Rayment, I., and Holden, H. M. (1997) *Biochemistry* **36,** 1212–1222
8. Thoden, J. B., and Holden, H. M. (1998) *Biochemistry* **37,** 11469–11477
9. Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L., and Holden, H. M. (2000) *Biochemistry* **39,** 5691–5701
10. Gulbinsky, J. S., and Cleland, W. W. (1968) *Biochemistry* **7,** 566–575
11. Ballard, F. J. (1966) *Biochem. J.* **101,** 70–75
12. Walker, D. G., and Khan, H. H. (1968) *Biochem. J.* **108,** 169–175
13. Timson, D. J., and Reece, R. J. (2002) *Biochimie* (Paris) **84,** 265–272
14. Timson, D. J., and Reece, R. J. (2003) *Eur. J. Biochem.* **270,** 1767–1774
15. Dey, P. M. (1983) *Eur. J. Biochem.* **136,** 155–159
16. Bork, P., Sander, C., and Valencia, A. (1993) *Protein Sci.* **2,** 31–40
17. Zhou, T., Daugherty, M., Grishin, N. V., Osterman, A. L., and Zhang, H. (2000) *Structure* **8,** 1247–1257
18. Krishna, S. S., Zhou, T., Daugherty, M., Osterman, A., and Zhang, H. (2001) *Biochemistry* **40,** 10810–10818
19. Yang, D., Shipman, L. W., Roessner, C. A., Scott, A. I., and Sacchettini, J. C. (2002) *J. Biol. Chem.* **277,** 9462–9467
20. Fu, Z., Wang, M., Potter, D., Miziorko, H. M., and Kim, J. -J. (2002) *J. Biol. Chem.* **277,** 18134–18142
21. Romanowski, M. J., Bonanno, J. B., and Burley, S. K. (2002) *Proteins: Struct. Funct. Genet.* **47,** 568–571
22. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) *J. Mol. Biol.* **229,** 105–124
23. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
24. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 849–861
25. Cowtan, K., and Main, P. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 487–493
26. Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987) *Acta Crystallogr. Sect. A* **43,** 489–501
27. Roussel, A., Fontecilla-Camps, J. C., and Cambillau, C. (1990) *Acta Crystallogr. Sect. A* **4,** C66–C67
28. Wilson, D. B., and Hogness, D. S. (1969) *J. Biol. Chem.* **244,** 2137–2142
29. Timson, D. J., Ross, H. C., and Reece, R. J. (2002) *Biochem. J.* **363,** 515–520
30. 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
31. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950