

The Molecular Architecture of Human *N*-Acetylgalactosamine Kinase^{*[5]}

Received for publication, May 25, 2005, and in revised form, June 10, 2005 Published, JBC Papers in Press, July 8, 2005, DOI 10.1074/jbc.M505730200

James B. Thoden and Hazel M. Holden¹

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

Galactokinase plays a key role in normal galactose metabolism by catalyzing the conversion of α -D-galactose to galactose 1-phosphate. Within recent years, the three-dimensional structures of human galactokinase and two bacterial forms of the enzyme have been determined. Originally, the gene encoding galactokinase in humans was mapped to chromosome 17. An additional gene, encoding a protein with sequence similarity to galactokinase, was subsequently mapped to chromosome 15. Recent reports have shown that this second gene (*GALK2*) encodes an enzyme with greater activity against GalNAc than galactose. This enzyme, GalNAc kinase, has been implicated in a salvage pathway for the reutilization of free GalNAc derived from the degradation of complex carbohydrates. Here we report the first structural analysis of a GalNAc kinase. The structure of the human enzyme was solved in the presence of MnAMPPNP and GalNAc or MgATP and GalNAc (which resulted in bound products in the active site). The enzyme displays a distinctly bilobal appearance with its active site wedged between the two domains. The N-terminal region is dominated by a seven-stranded mixed β -sheet, whereas the C-terminal motif contains two layers of anti-parallel β -sheet. The overall topology displayed by GalNAc kinase places it into the GHMP superfamily of enzymes, which generally function as small molecule kinases. From this investigation, the geometry of the GalNAc kinase active site before and after catalysis has been revealed, and the determinants of substrate specificity have been defined on a molecular level.

β -D-Galactose is converted to the more metabolically useful glucose 1-phosphate through the actions of four enzymes that constitute the Leloir pathway (1). In the second step of this pathway, α -D-galactose is phosphorylated at the C-1 position by galactokinase, a MgATP-dependent small molecule kinase (Scheme 1). Defects in the *GK1* gene encoding human galactokinase have been shown to give rise to type II galactosemia, clinical symptoms of which include the formation of cataracts at an early age (2). Recently, the three-dimensional structures of the galactokinases from *Lactococcus lactis* (3), *Pyrococcus furiosus* (4), and humans (5) have been determined. These investigations have demonstrated that the galactokinases share similar molecular architectures

with members of the GHMP (galactose kinase, homoserine kinase, mevalonate kinase, and phosphomevalonate kinase) superfamily. The galactokinases fold into two distinct N- and C-terminal domains with their active sites wedged between these motifs. On the basis of the observed MgAMPPNP² and α -D-galactose binding modes to human galactokinase, it has been postulated that two strictly conserved residues, Arg³⁷ and Asp¹⁸⁶, may play critical roles in catalysis (5).

It has become increasingly apparent that the three-dimensional motif exhibited by the galactokinases also serves as a molecular scaffold in transcriptional regulation. In *Saccharomyces cerevisiae*, for example, expression of the genes encoding the enzymes of the Leloir pathway are tightly controlled through the action of three key proteins: Gal4p, Gal80p, and Gal3p (6). Gal4p functions as a transcriptional activator but is rendered inactive in the presence of Gal80p (7). For transcription to occur, Gal80p must interact with Gal3p in a manner that is still not entirely understood. What is known, however, is that Gal3p requires both galactose and ATP to interact with Gal80p and that it functions as a ligand sensor. Gal3p shows ~70 and ~50% amino acid sequence similarities to yeast and human galactokinases, respectively (8, 9).

The *GK1* gene coding for human galactokinase was originally mapped to chromosome 17 (10, 11). A second gene encoding a putative galactokinase, with ~35% amino acid sequence identity, was subsequently reported that mapped to chromosome 15, suggesting, perhaps, that humans contain two galactokinases (12). These apparently contradicting reports for the chromosomal location of human galactokinase have been reconciled through the recent studies of Pastuszak *et al.* (13, 14), who demonstrated that the gene located on chromosome 15 does not, in fact, encode a *bona fide* galactokinase but rather a GalNAc kinase (Scheme 1). Given the ubiquitous role of GalNAc in complex carbohydrates, it is not surprising that this enzyme has been postulated to function in a salvage pathway for recycling free GalNAc. A detailed analysis of the porcine protein has demonstrated that the enzyme is very specific for GalNAc, in striking contrast to galactokinase, which has no activity against this sugar. Additionally, ATP has been shown to be the best phosphate donor (14).

In an effort to understand the observed substrate specificity of the enzyme, we initiated a structural analysis of human GalNAc kinase. Here we report the high resolution three-dimensional structure of the enzyme in the presence of either GalNAc and MnAMPPNP or its reaction products, GalNAc 1-phosphate and MgADP. From this x-ray analysis, the overall molecular architecture of the enzyme has been revealed, and the observed substrate specificity of the enzyme has been defined on a molecular level.

EXPERIMENTAL PROCEDURES

Cloning of the Human GK2 Gene—The human *GK2* gene was PCR-amplified from IMAGE clone 3504930 (The I.M.A.G.E. Consortium)

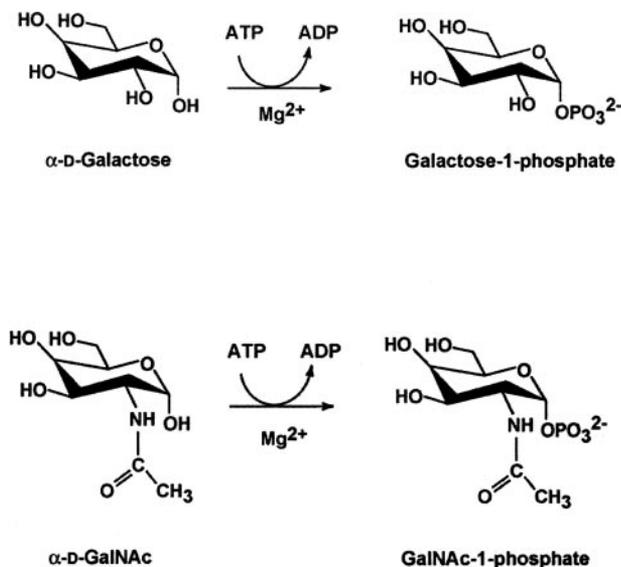
* This work was supported by National Institutes of Health Grant DK47814 (to H. M. H.). Use of the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source was supported by the United States Department of Energy, Office of Energy Research, under Contract W-31-109-ENG-38. 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 (codes 2A2C and 2A2D) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. 1 and 2.

¹ To whom correspondence should be addressed. Tel.: 608-262-4988; Fax: 608-262-1319; E-mail: Hazel_Holden@biochem.wisc.edu.

² The abbreviations used are: AMPPNP, 5'-adenylyl- β , γ -imidodiphosphate; MES, 4-morpholineethanesulfonic acid; ATP- γ S, adenosine 5'-O-(thiotriphosphate).



purchased from ATCC such that the forward and reverse primers added NdeI and XhoI 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). This vector was used to transform *Escherichia coli* DH5 α cells. The *GK2* gene was subsequently sequenced in the pGEM-T vector construct with the ABI PrismTM Big Dye Primer Cycle Sequencing Kit (Applied Biosciences, Inc.) to confirm the integrity of the PCR product. The pGEM-T vector construct was then digested with NdeI and XhoI, and the gene was separated from digestion by-products on a 1.0% agarose gel. The gene was excised from the gel, purified with the QIAquick gel extraction kit (Qiagen), and ligated into the expression vector pET-28b(+) (Novagen) that was previously digested with NdeI and XhoI. The pET28 plasmid had also been previously altered such that its thrombin cleavage site had been replaced with the recognition sequence for TEV protease, resulting in a construct that contains an additional 20 amino acid residues at the N terminus with the sequence MGSSHHH-HHHSSENLYFQGH. *E. coli* DH5 α cells were transformed with the ligation mixture and grown on LB/agar plates for selection with kanamycin. Individual colonies were selected and cultured overnight. The plasmid DNA was subsequently extracted with QIAprep Spin Miniprep Kit (Qiagen). Plasmids were tested for insertion of the *GK2* gene by digestion with NdeI and XhoI.

Protein Expression and Purification—For protein expression, the pET28-GK2 plasmid was used to transform *E. coli* HMS174(DE3) cells (Novagen). A starter culture from a single colony was grown overnight at 37 °C in LB medium supplemented with kanamycin. Subsequently, 10 ml were transferred to 1000 ml of supplemented TB media (50 mg/liter kanamycin) in a 2-liter shaker flask and grown at 37 °C until an optical density of ~ 0.8 was achieved at 600 nm. The cultures were then transferred to a shaker at room temperature (~ 20 °C) and allowed to grow until an optical density of greater than 1.8 was obtained, at which point isopropyl 1-thio- β -D-galactopyranoside was added to a final concentration of 0.05 mM. Cell growth was allowed to continue at room temperature for an additional 18 h.

The cells were harvested by centrifugation at 4000 $\times g$ for 15 min and frozen in liquid nitrogen. Frozen cells (125 g) were thawed in 500 ml of lysis buffer consisting of 50 mM NaH₂PO₄, 10 mM imidazole, and 300

mM NaCl (pH 8.0). The thawed cells were placed in an ice bath and disrupted by five rounds of sonication (1-min duration each) separated by 5 min of cooling. Cellular debris was removed by centrifugation at 20,000 $\times g$ for 25 min. The clarified supernatant was loaded onto a 12-ml column of Ni²⁺-nitrilotriacetic acid-agarose (Qiagen) that had been previously equilibrated with lysis buffer. The column was then washed with lysis buffer until the absorbance reading at 280 nm reached background level. The protein was eluted with a linear gradient of 10–250 mM imidazole in lysis buffer. Protein-containing fractions were pooled based on purity as judged by SDS-PAGE and dialyzed against 10 mM Tris and 25 mM NaCl (pH 8.0). The dialyzed protein was further purified by anion exchange high pressure liquid chromatography using a 6-ml Resource-Q column. The protein was eluted at pH 8.0 (25 mM Tris) with a linear gradient from 25 to 250 mM NaCl. Protein-containing fractions were again pooled based on purity as judged by SDS-PAGE and dialyzed against 10 mM Tris and 200 mM NaCl (pH 8.0) and then concentrated to 14.5 mg/ml based on the extinction coefficient of 1.35 cm/(mg·ml) as calculated with the program Protean (DNASTAR, Inc., Madison, WI). A typical yield was ~ 200 mg of protein/125 g of cells. Note that the His tag was not removed prior to crystallization trials.

Crystallization of GalNAc Kinase—A search for crystallization conditions was conducted at both room temperature and at 4 °C via the hanging drop method of vapor diffusion utilizing an “in-house” designed sparse matrix screen composed of 144 conditions. The best crystals were observed growing at pH 6 from polyethylene glycol 3400 solutions in the presence of either MgAMPPNP or MnAMPPNP and GalNAc at room temperature. These conditions were optimized with crystals being routinely grown in 7–14 days from 10 mM MnAMPPNP, 100 mM GalNAc, 16–20% polyethylene glycol 3400, 200 mM LiCl, and 100 mM MES (pH 6.0). They achieved maximum dimensions of $\sim 0.5 \times 0.5 \times 0.2$ mm and belong to the space group P6₅ with unit cell dimensions of approximately $a = b = 118$ Å and $c = 65$ Å. The asymmetric unit contained one monomer. Crystals were also obtained in the presence of GalNAc and MgATP (which turned over to product in the active site). Whereas these belonged to the same space group, P6₅, the unit cell dimensions were slightly different at $a = b = 124$ Å, and $c = 60$ Å.

Preparation of Selenomethionine-labeled Protein—Cultures of *E. coli* HMS174(DE3) cells harboring the plasmid encoding the human *GK2* gene were grown overnight in M9 minimal media at 37 °C. Subsequently, 15 ml of the overnight culture were used to inoculate each of 12 \times 2-liter baffled flasks containing 500 ml of M9 minimal media supplemented with 5 mg/liter thiamine. Cultures were grown at 37 °C to an optical density of ~ 0.9 at 600 nm 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 $\times g$ for 10 min, and the cell paste was frozen in liquid nitrogen for storage at -80 °C. Selenomethionine-labeled protein was purified to homogeneity according to the procedure described above.

Crystallization of Selenomethionine-labeled GalNAc Kinase—Crystals of the selenomethionine-labeled protein were grown under the same conditions as the wild-type protein, except that MgATP γ S was employed as the nucleotide analog, since it yielded better crystals than using MgAMPPNP or MnAMPPNP. The observed unit cell dimensions were similar to those obtained for the wild-type enzyme complexed with MnAMPPNP.

TABLE ONE

X-ray data collection statistics

	Wavelength	Resolution	Independent reflections	Completeness	Redundancy	Avg $I/\text{Avg } \sigma(I)$	R_{sym}^a
	Å	Å		%			%
Peak	0.97926	50.0 to 2.30	23,644	100.0	11.0	51.2	7.7
		(2.38 to 2.30) ^b	(2340)	(100.0)	(11.0)	(14.1)	(15.6)
Inflection	0.97940	50.0 to 2.30	23,662	100.0	11.0	45.5	7.0
		(2.38 to 2.30)	(2352)	(100.0)	(11.0)	(14.0)	(15.3)
Remote	0.96411	50.0 to 2.30	23,665	100.0	11.0	47.5	6.3
		(2.38 to 2.30)	(2355)	(100.0)	(11.0)	(15.3)	(13.6)
AMPPNP GalNAc complex	1.5418	30.0 to 2.20	26,919	98.9	8.7	15.9	8.1
		(2.30 to 2.20)	(3322)	(98.4)	(4.8)	(3.1)	(35.4)
ATP GalNAc complex	0.96411	50.0 to 1.65	59,826	94.3	8.6	46.3	5.4
		(1.71 to 1.65)	(4607)	(72.7)	(3.0)	(3.5)	(17.4)

^a $R_{\text{sym}} = (\sum |I - \bar{I}| / \sum I) \times 100$.

^b Statistics for the highest resolution bin are shown in parentheses.

X-ray Data Collection—Crystals were stabilized for x-ray data collection by first harvesting them into a synthetic mother liquor containing 25% polyethylene glycol 3400, 200 mM NaCl, 100 mM LiCl, 10 mM MnAMPPNP, 100 mM GalNAc, and 100 mM MES (pH 6.0). The crystals were frozen by rapid transfer into a cryoprotectant solution composed of 35% polyethylene glycol 3400, 300 mM NaCl, 300 mM LiCl, 10 mM MnAMPPNP, 100 mM GalNAc, 12% ethylene glycol, and 100 mM MES (pH 6.0). X-ray data were collected from crystals of the selenomethionine-substituted protein and from crystals of the enzyme complexed with MgATP and GalNAc on a CCD detector at SBC Beamline 19-BM (Advanced Photon Source, Argonne National Laboratory, Argonne, IL). The x-ray data were processed and scaled with HKL2000 (15). For the complex with MnAMPPNP and GalNAc, the x-ray data were collected at 100 K with a Bruker AXS Platinum 135 CCD detector controlled with the Proteum software suite (Bruker AXS Inc., Madison, WI). The x-ray source was CuK α radiation from a Rigaku RU200 x-ray generator equipped with montel optics and operated at 50 kV and 90 mA. The x-ray data were processed with SAINT (version V7.06A; Bruker AXS) and internally scaled with SADABS (version 2005/1; Bruker AXS). Relevant x-ray data collection statistics are presented in TABLE ONE.

X-ray Structural Analyses—The structure of human GalNAc kinase was solved via MAD phasing with crystals of the selenomethionine-substituted protein. The software package SOLVE was utilized to determine the positions of 11 out of the 13 selenium atoms in the asymmetric unit and to generate initial protein phases (figure of merit = 0.70) (16). Solvent flattening with RESOLVE (figure of merit = 0.85) resulted in an interpretable electron density map calculated to 2.3 Å resolution (17). The map allowed for a complete tracing of the polypeptide chain (458 amino acids) with the exception of the residues between Ala⁴³⁰ and Lys⁴⁴⁰.

The structure obtained from the MAD phasing was then employed as a search model to solve the structure of the enzyme complexed with MgATP and GalNAc via molecular replacement with the program AMORE (18). Alternate cycles of manual model building and least squares refinement with the software package TNT reduced the *R*-factor to 16.8% for all measured x-ray data from 30.0 to 1.65 Å resolution (19). Relevant refinement statistics are presented in TABLE TWO. In this model, there are two breaks in the polypeptide chain between Asn⁹⁷ and Ile¹⁰⁰ and between Ala⁴³⁰ and Lys⁴⁴⁰.

The complex with MnAMPPNP and GalNAc was solved by difference Fourier techniques using the original model derived from the MAD phasing experiment. Alternate cycles of manual model building and least squares refinement with the software package TNT reduced

the *R*-factor to 20.7% for all measured x-ray data from 30.0 to 2.20 Å resolution (19). Again, relevant refinement statistics are given in TABLE TWO. Ramachandran plots for both complexes are given in the supplemental materials. There are no significant outliers in either model.

RESULTS AND DISCUSSION

Overall Structure of the Enzyme with Bound Products—Human GalNAc kinase contains 458 amino acid residues and packs in the P6₅ unit cell with one polypeptide chain per asymmetric unit. On the basis of the symmetry elements contained within this space group, which include only 2-fold, 3-fold, and 6-fold screw axes, it can be concluded that the enzyme exists as a monomer. GalNAc kinase from porcine kidney has also been reported to function as a monomer (14).

A ribbon representation of the molecule, with overall dimensions of ~69 × 54 × 64 Å, is presented in Fig. 1a. The N-terminal domain is dominated by a seven-stranded mixed β -sheet with the two parallel strands formed by Ala⁷ to Val¹¹ and Leu⁴⁵³ to Ala⁴⁵⁸. The C-terminal domain contains two distinct regions of anti-parallel β -sheet, with each layer containing four strands. A total of 14 major α -helices surround the β -sheets of the N- and C-terminal domains. Pro²⁰⁹, which is located in a random coil region connecting β -strands 9 and 10 in the C-terminal domain, adopts the cis-conformation. Ser¹⁴⁶–Leu¹⁶² form an α -helix that is situated with its positive helix dipole moment projecting toward the β -phosphoryl group of the ADP (Fig. 1a). In this complex, the polypeptide chain extends from Ala² to Ala⁴⁵⁸ with two breaks between Asn⁹⁷ and Ile¹⁰⁰ and Ala⁴³⁰ to Lys⁴⁴⁰ (Fig. 1a).

The electron density map for the protein crystallized in the presence of its substrates, GalNAc and MgATP, clearly reveals the presence of products in the active site as can be seen in Fig. 1b. The ligands are well ordered with average *B*-factors of 16.7 and 16.8 Å² for the sugar and nucleotide, respectively. The Mg²⁺ ion has a *B* value of 9.4 Å². The ribose of the ADP adopts the C_{3'}-endo pucker, and the adenine ring is in the anti-conformation. Both Tyr⁸⁸ and Trp¹⁰⁷ form stacking interactions with the adenine ring of the nucleotide (Fig. 1b). Trp¹⁰⁷ is strictly conserved among galactokinase sequences deposited in the Swiss-Prot data bank thus far.

A stereo view of the active site, within 3.2 Å of the ligands, is presented in Fig. 1c. The Mg²⁺ ion is surrounded in an octahedral coordination sphere with metal-ligand bond distances ranging between 2.2 and 2.4 Å. O^γ of Ser¹⁴⁷ and a phosphoryl oxygen of GalNAc 1-phosphate provide the axial ligands, whereas two well ordered water molecules (*B* values of 14.1 and 14.6 Å²) and an α - and a β -phosphoryl oxygen of the ADP serve as the equatorial ligands. Most of the interactions between

TABLE TWO

Least-squares refinement statistics

	AMPPNP·GalNAc	ATP·GalNAc
Resolution limits (Å)	30.0 to 2.20	30.0 to 1.65
<i>R</i> -factor ^a (overall) %/no. of reflections	20.7/26,919	16.8/59826
<i>R</i> -factor (working) %/no. of reflections	20.3/24,211	16.7/53716
<i>R</i> -factor (free) %/no. of reflections	24.7/2693	20.1/5980
No. of protein atoms	3462 ^b	3470 ^c
No. of heteroatoms	303 ^d	457 ^e
Average <i>B</i> values (Å ²)		
Protein atoms	33.3	30.2
AMPPNP	28.2	NA ^f
ADP	NA	16.8
GalNAc	14.2	NA
GalNAc 1-phosphate	NA	16.7
Solvents	37.9	40.9
Weighted root mean square deviations from ideality		
Bond lengths (Å)	0.013	0.013
Bond angles (degrees)	2.49	2.37
Trigonal planes (Å)	0.006	0.006
General planes (Å)	0.013	0.012
Torsional angles ^g (degrees)	18.2	16.6

^a R -factor = $(\sum |F_o - F_c| / \sum |F_o|) \times 100$, where F_o is the observed structure factor amplitude, and F_c is the calculated structure factor amplitude.

^b These include multiple conformations for residues Gln²⁶¹ and Thr³²⁰.

^c These include multiple conformations for residues Met²⁵, Ile⁴⁶, Leu⁵⁹, Ser¹²⁵, Ile²⁰⁵, Ile²²⁵, Met²³², Arg²⁴⁸, and Arg³⁰².

^d These include one AMPPNP, one manganese ion, one GalNAc, and 256 water molecules.

^e These include one ADP, one magnesium ion, one GalNAc 1-phosphate, one chloride ion, one sodium ion, and 357 water molecules.

^f NA, not applicable.

^g The torsional angles were not restrained during the refinement.

the protein and the adenine and ribose groups of ADP are mediated via water molecules. There is one direct hydrogen bond between the adenine ring and the side chain of Asn⁸³. The pyrophosphate group of ADP is anchored to the protein via the backbone amide groups of Ser¹⁴¹, Gly¹⁴³, Ser¹⁴⁶, Ser¹⁴⁷, and Lys²³⁴. Additionally, O^γ of Ser¹⁴⁶, N^{δ2} of Asn²³³, N^ε of Lys²³⁴, and two water molecules lie within 3.2 Å of the pyrophosphate moiety. Lys²³⁴ is very well ordered, with an average *B*-factor of 18.1 Å². Key amino acid residues involved in positioning GalNAc 1-phosphate within the active site include Arg⁴³, Glu⁴⁹, His⁵⁰, Asp⁵², Ser¹⁴⁵, Gly¹⁸⁷, Asp¹⁹⁰, and Ala⁴⁰². There are four water molecules that also interact with the sugar moiety. The 1-phosphorus of the sugar product is situated at 4.8 Å from the β-phosphorus of the ADP.

Structure of the Enzyme in the Presence of MnAMPPNP and GalNAc—GalNAc kinase belongs to the GHMP superfamily, which also includes homoserine kinase (20), mevalonate kinase (21, 22), and phosphomevalonate kinase (23), among others. The catalytic mechanisms for these proteins have been the subject of recent investigations, and two quite different scenarios have been proposed (21, 24, 25). In the case of mevalonate kinase, it has been postulated that a conserved aspartate serves as an active site base to abstract the proton from the C-5 hydroxyl group of mevalonate that is ultimately phosphorylated (21, 22). There is an apparent absence of a catalytic base in homoserine kinase, however (24). Rather, it has been suggested that the catalytic mechanism of homoserine kinase proceeds through a mechanism whereby the close positioning of the substrate and the γ-phosphate of the ATP results in direct transfer of the proton from the δ-OH group of homoserine to the γ-phosphate of ATP and attack of the δ-oxygen on the γ-phosphorus. In both of these proposals, there is the tacit assumption that phosphoryl transfer occurs through an associative transition state. This is in striking contrast to studies demonstrating that reactions of ATP in solution proceed via a dissociative, metaphosphate-like transition state (26).

Both human galactokinase and GalNAc kinase contain a similarly positioned aspartate residue as that observed in mevalonate kinase. Yet

kinetic analyses of both yeast and human galactokinases suggest that proton transfer does not play a key role in the rate-determining step of catalysis (27, 28). In an attempt to mimic the Michaelis complex for GalNAc kinase, the protein was crystallized in the presence of the non-hydrolyzable analog AMPPNP (29) and GalNAc. Overall, the polypeptide chains for this complex and that with bound products are virtually identical and superimpose with a root mean square deviation of 0.34 Å. A close-up view of the environment around the GalNAc ligand is displayed in Fig. 2a. Again, the metal ion, in this case manganese, is octahedrally coordinated by three phosphoryl oxygens donated by the nucleotide: two water molecules and O^γ of Ser¹⁴⁷. Both Arg⁴³ and Lys²³⁴ are located near the γ-phosphoryl group, possibly enhancing its electrophilic character. In this complex, however, Lys²³⁴ is not well ordered (*B*-factor of 43.8 Å²), and difference electron density maps indicate a second conformation for this residue that swings out away from the active site. Asp¹⁹⁰, which is absolutely conserved among the galactokinases, lies within 3.1 Å of the 1-hydroxyl group of the sugar. Importantly, however, O^{δ1} and O^{δ2} of Asp¹⁹⁰ also lie within 2.8 and 3.1 Å, respectively, of N^{η2} of Arg⁴³. Given the surrounding environment, it is difficult to envision how Asp¹⁹⁰ could function as an active site base to remove the proton from the 1-hydroxyl group of GalNAc.

Note that the 1-hydroxyl group of GalNAc is positioned at 3.1 and 2.5 Å from two of the γ-phosphoryl oxygens and 3.1 Å of the γ-phosphorus of the AMPPNP. Shown in Fig. 2b is a superposition of the active sites before and after catalysis. As can be seen, the active site is ideally suited for positioning the 1-hydroxyl group of the sugar substrate in the correct orientation for direct attack at the γ-phosphorus. The reaction may not, in fact, require a catalytic base but rather proceed via approximation and perhaps through a metaphosphate-like transition state. In this situation, the bond between the γ-phosphorus and the bridging oxygen is largely broken, which results in a build-up of negative charge on the bridging oxygen and positive charge on the phosphorus as discussed in Ref. 26. Movement of the γ-phosphorus toward the 1-hydroxyl group of

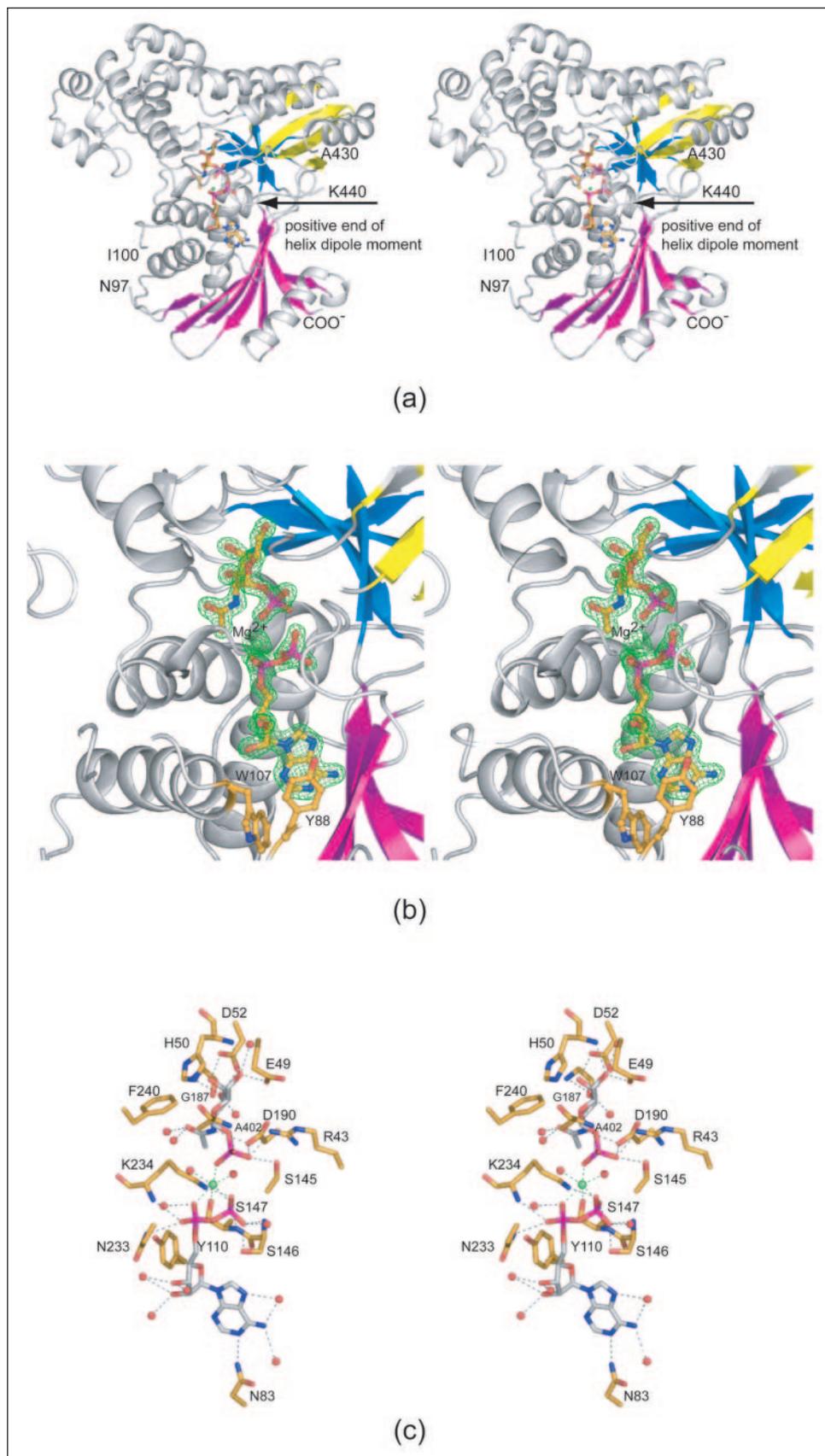


FIGURE 1. Ribbon representation of human GalNAc kinase. Human GalNAc kinase folds into two motifs referred to as the N- and C-terminal domains. As indicated in *a*, the N-terminal domain is dominated by a seven-stranded mixed β -sheet highlighted in *magenta*. Two layers of antiparallel β -sheet, colored in *yellow* and *blue*, characterize the C-terminal domain. The location of the active site is indicated by the ball-and-stick representations for GalNAc-1-phosphate and MgADP, with the Mg^{2+} ion highlighted in *green*. Electron density corresponding to GalNAc-1-phosphate and MgADP is displayed in *b*. While the protein was crystallized in the presence of GalNAc and MgATP, the electron density clearly reveals that the enzyme is active in the crystalline lattice. The map, contoured at 2σ , 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 coordinates for the ligands. A close-up view of the active site with bound products is depicted in *c*. Only those residues located within ~ 3.2 Å of the ligands are shown. The *green dashed lines* indicate coordinate covalent bonds between the magnesium ion and its ligands. The *black dashed lines* indicate potential hydrogen bonding interactions. For clarity, Ser¹⁴¹ and Gly¹⁴³ were omitted from the figure.

GalNAc could effectively lower its pK_a , resulting in proton transfer to Asp¹⁹⁰. Similar mechanisms have been proposed previously for GTP hydrolysis by transducin α (30) and p21^{ras} (31, 32) and ATP hydrolysis by myosin subfragment-1 (33, 34).

That different members of the GHMP superfamily have been suggested to proceed via contrasting reaction mechanisms is, indeed, intriguing. It is possible that, in fact, they actually function by similar mechanisms that are not completely understood at this time.

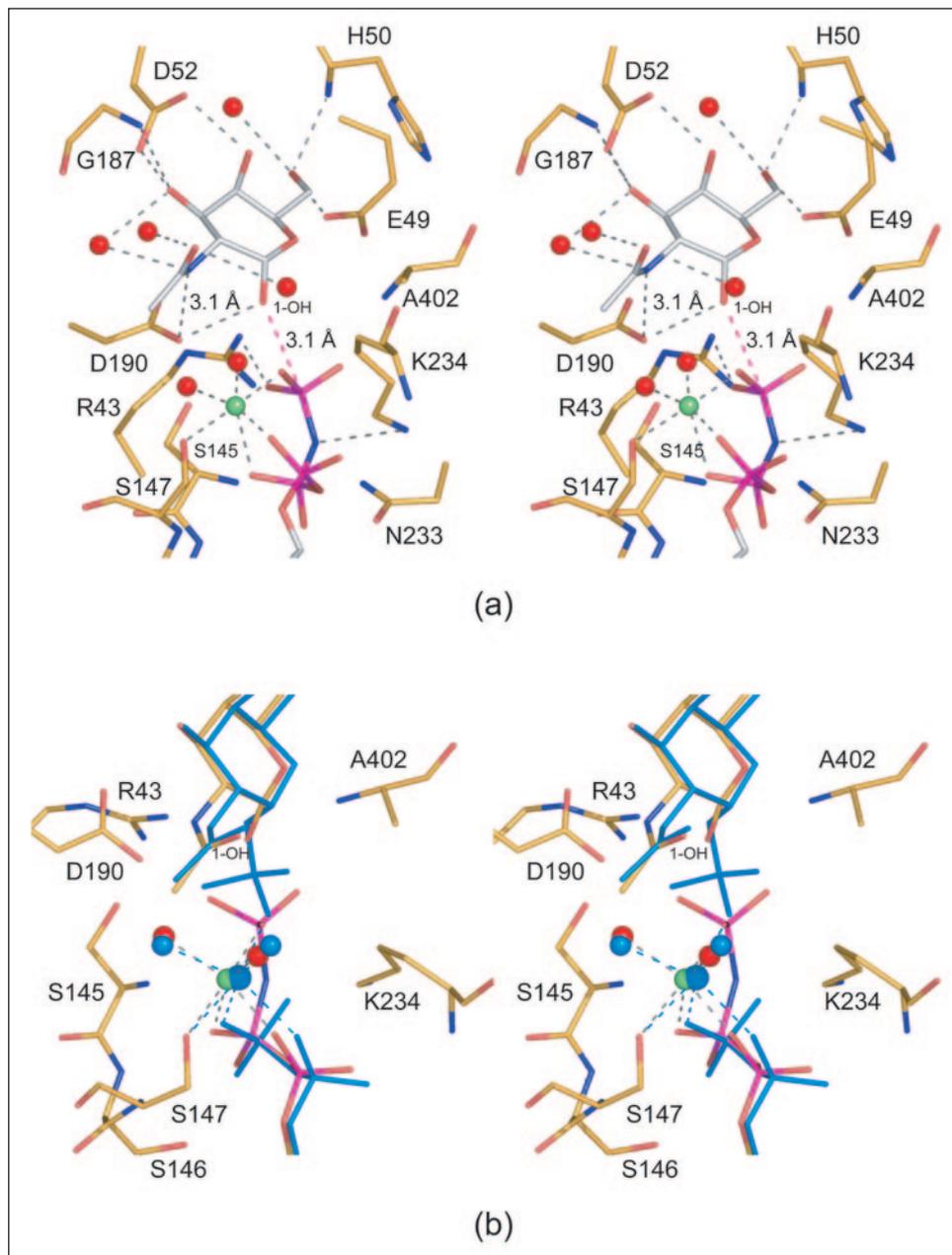


FIGURE 2. Close-up view of the GalNAc binding region. Those residues located within ~ 3.2 Å of GalNAc are shown in *a*. Note the close approach of the 1-hydroxyl group of the sugar to the γ -phosphorus of the nucleotide (indicated by the red dashed line). A superposition of the AMPPNP-GalNAc and ADP-GalNAc 1-phosphate complexes near the sugar ligand is displayed in *b*. In that there are minimal changes in side chain conformations between the two models, only the side chains for the AMPPNP-GalNAc complex are shown. The positions of the Mn^{2+} ion and its associated water molecules are indicated by the green and red spheres, respectively. The MgADP and GalNAc 1-phosphate ligands are displayed in aquamarine.

Comparison of Human GalNAc Kinase and Human Galactokinase—GalNAc kinase is substantially larger, with 458 versus 392 amino acid residues for human galactokinase. Highlighted in Fig. 3*a* are the three regions where these two enzymes differ significantly. Excluding these areas, however, the two proteins superimpose with a root mean square deviation of 1.3 Å for 332 structurally equivalent α -carbon atoms. The first region where these two proteins differ is at the N terminus, labeled *A* in Fig. 3*a*. While both polypeptides initiate with an extended chain, in human galactokinase, this region abuts the C-terminal domain. In GalNAc kinase, however, residues Ala⁷ to Val¹¹ actually form the seventh strand of the mixed β -sheet in the N-terminal domain. The second significant difference between these two proteins occurs at Gly²⁵¹ in galactokinase and Ala²⁵⁵ in GalNAc kinase. There is a 19-residue insertion in GalNAc kinase that folds into two additional helices. The largest insertion in GalNAc kinase, relative to galactokinase, occurs at Leu²⁹⁰. Here a 31-residue insertion folds into two α -helices defined by Pro²⁹⁷–Leu³⁰⁴ and Leu³⁰⁸–Gln³¹⁴.

The substrate specificity for human GalNAc was addressed in a study published several years ago (13). While the most favored substrate for the enzyme is GalNAc, it can also turn over galactose, albeit when the

sugar is present in millimolar concentrations. Substrate specificity studies of human galactokinase have also been recently conducted and have revealed that the enzyme cannot turn over GalNAc (35). The two enzymes share an approximate 35% amino acid sequence identity. A superposition near the sugar binding pockets of human galactokinase and GalNAc kinase is presented in Fig. 3*b*. The first noticeable difference in the binding pockets is the interaction in galactokinase between the 4-hydroxyl group of galactose and O ^{η} of Tyr²³⁶. In GalNAc kinase, this tyrosine is replaced with a phenylalanine. The more important difference, however, is near the *N*-acetyl group of GalNAc. In human galactokinase, the positions of both Met¹⁸⁰ and Cys¹⁸² preclude the possibility of the enzyme accepting GalNAc as a substrate. The more bulky GalNAc ligand is accommodated in GalNAc kinase by the substitution of Met¹⁸⁰ with a threonine and Cys¹⁸² with a glycine.

Conclusions—Sugars often play major roles in the biological effectiveness of such clinically relevant compounds as erythromycin, vancomycin, novobiocin, and digitoxin (36). As such, there is an increasing effort among various laboratories to alter the glycosylation patterns of

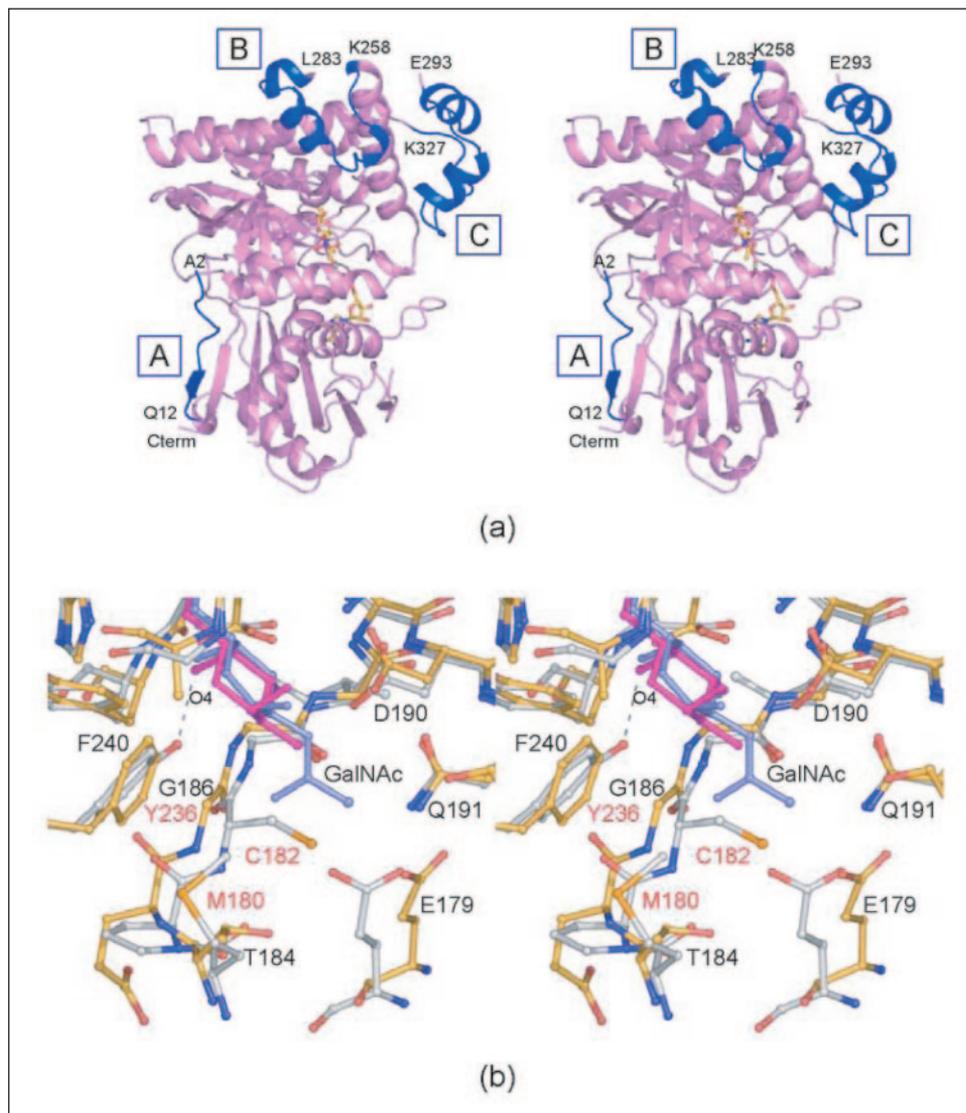


FIGURE 3. A comparison of human galactokinase versus GalNAc kinase. The human forms of galactokinase and GalNAc kinase differ in three specific regions labeled A, B, and C and indicated by the blue ribbons in a. A superposition of the sugar binding regions for these enzymes is given in b. The yellow bonds correspond to GalNAc kinase, with the sugar ligand depicted in aquamarine. The white bonds correspond to galactokinase, with the sugar ligand highlighted in magenta. The red and black labels correspond to residues in galactokinase and GalNAc kinase, respectively. Coordinates for the human galactokinase were from this laboratory (Protein Data Bank accession number 1WUU).

these drugs to produce novel therapeutics. One approach that has been championed of late is the so-called “*in vitro* glycorandomization” method whereby chemically synthesized sugars are activated and subsequently attached to natural products via promiscuous nucleotidyltransferases and glycosyltransferases (37).

Galactokinase, as a biological catalyst, has sparked renewed interest for its use in the preparation of modified sugar phosphates. Thus far, a variant of the *E. coli* galactokinase (Y371H) has demonstrated a marked increase in substrate flexibility with respect to modifications at C-2, C-3, and C-5 but not at C-4 (38). Quite strikingly, the same mutation in the *L. lactis* enzyme (Y385H) resulted in a protein displaying kinase activity toward several C-4-substituted sugars (39). These contrasting results from the *E. coli* and *L. lactis* enzymes emphasize the advantages of optimizing additional enzymes for synthetic purposes. The structure of human GalNAc kinase presented here represents yet another protein platform for the eventual production of novel sugar phosphates.

Acknowledgments—We thank Dr. Jurek Osipiuk for assistance during x-ray data collection at the Advanced Photon Source. The helpful discussions of Drs. Victor Hong, David J. Timson, Richard J. Reece, and Ivan Rayment are gratefully acknowledged.

REFERENCES

- Holden, H. M., Rayment, I., and Thoden, J. B. (2003) *J. Biol. Chem.* **278**, 43885–43888
- Holden, H. M., Thoden, J. B., Timson, D. J., and Reece, R. J. (2004) *Cell Mol. Life Sci.* **61**, 2471–2484
- Thoden, J. B., and Holden, H. M. (2003) *J. Biol. Chem.* **278**, 33305–33311
- 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
- Thoden, J. B., Timson, D. J., Reece, R. J., and Holden, H. M. (2005) *J. Biol. Chem.* **280**, 9662–9670
- Reece, R. J. (2000) *Cell Mol. Life Sci.* **57**, 1161–1171
- Platt, A., and Reece, R. J. (1998) *EMBO J.* **17**, 4086–4091
- Zenke, F. T., Engles, R., Vollenbroich, V., Meyer, J., Hollenberg, C. P., and Breunig, K. D. (1996) *Science* **272**, 1662–1665
- Platt, A., Ross, H. C., Hankin, S., and Reece, R. J. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 3154–3159
- Elsevier, S. M., Kucherlapati, R. S., Nichols, E. A., Creagan, R. P., Giles, R. E., Ruddle, F. H., Willecke, K., and McDougall, J. K. (1974) *Nature* **251**, 633–636
- Stambolian, D., Ai, Y., Sidjanin, D., Nesburn, K., Sathe, G., Rosenberg, M., and Bergsma, D. J. (1995) *Nat. Genet.* **10**, 307–312
- Lee, R. T., Peterson, C. L., Calman, A. F., Herskowitz, I., and O'Donnell, J. J. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10887–10891
- Pastuszak, I., O'Donnell, J., and Elbein, A. D. (1996) *J. Biol. Chem.* **271**, 23653–23656
- Pastuszak, I., Drake, R., and Elbein, A. D. (1996) *J. Biol. Chem.* **271**, 20776–20782
- Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
- Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D. Biol. Crystallogr.* **55** (Pt 4), 849–861

17. Terwilliger, T. C. (2000) *Acta Crystallogr. D. Biol. Crystallogr.* **56** (Pt 8), 965–972
18. Navaza, J. (1987) *Acta Crystallogr. Sect. A* **43**, 645–653
19. Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987) *Acta Crystallogr. Sect. A* **43**, 489–501
20. Zhou, T., Daugherty, M., Grishin, N. V., Osterman, A. L., and Zhang, H. (2000) *Structure Fold Des.* **8**, 1247–1257
21. Fu, Z., Wang, M., Potter, D., Miziorko, H. M., and Kim, J. J. (2002) *J. Biol. Chem.* **277**, 18134–18142
22. Yang, D., Shipman, L. W., Roessner, C. A., Scott, A. I., and Sacchettini, J. C. (2002) *J. Biol. Chem.* **277**, 9462–9467
23. Romanowski, M. J., Bonanno, J. B., and Burley, S. K. (2002) *Proteins* **47**, 568–571
24. Krishna, S. S., Zhou, T., Daugherty, M., Osterman, A., and Zhang, H. (2001) *Biochemistry* **40**, 10810–10818
25. Pilloff, D., Dabovic, K., Romanowski, M. J., Bonanno, J. B., Doherty, M., Burley, S. K., and Leyh, T. S. (2003) *J. Biol. Chem.* **278**, 4510–4515
26. Admiraal, S. J., and Herschlag, D. (1995) *Chem. Biol.* **2**, 729–739
27. Timson, D. J., and Reece, R. J. (2002) *Biochimie (Paris)* **84**, 265–272
28. Timson, D. J., and Reece, R. J. (2003) *Eur. J. Biochem.* **270**, 1767–1774
29. Nakamaye, K. L., Wells, J. A., Bridenbaugh, R. L., Okamoto, Y., and Yount, R. G. (1985) *Biochemistry* **24**, 5226–5235
30. Sondek, J., Lambright, D. G., Noel, J. P., Hamm, H. E., and Sigler, P. B. (1994) *Nature* **372**, 276–279
31. Schweins, T., Langen, R., and Warshel, A. (1994) *Nat. Struct. Biol.* **1**, 476–484
32. Schweins, T., Geyer, M., Scheffzek, K., Warshel, A., Kalbitzer, H. R., and Wittinghofer, A. (1995) *Nat. Struct. Biol.* **2**, 36–44
33. Fisher, A. J., Smith, C. A., Thoden, J. B., Smith, R., Sutoh, K., Holden, H. M., and Rayment, I. (1995) *Biochemistry* **34**, 8960–8972
34. Smith, C. A., and Rayment, I. (1996) *Biochemistry* **35**, 5404–5417
35. Timson, D. J., and Reece, R. J. (2003) *BMC Biochem.* **4**, 16
36. Thorson, J. S., and Vogt, T. (2002) *Glycosylated Natural Products. Carbohydrate-Based Drug Discovery* (Wong, C.-W., ed), Wiley-VCH, Weinheim
37. Thorson, J. S., Barton, W. A., Hoffmeister, D., Albermann, C., and Nikolov, D. B. (2004) *Chembiochem.* **5**, 16–25
38. Hoffmeister, D., Yang, J., Liu, L., and Thorson, J. S. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13184–13189
39. Yang, J., Liu, L., and Thorson, J. S. (2004) *Chembiochem* **5**, 992–996