Structural Analysis of the H166G Site-Directed Mutant of Galactose-1-phosphate Uridylyltransferase Complexed with either UDP-glucose or UDP-galactose: Detailed Description of the Nucleotide Sugar Binding Site[†,‡]

James B. Thoden, Frank J. Ruzicka, Perry A. Frey, Ivan Rayment,* and Hazel M. Holden*

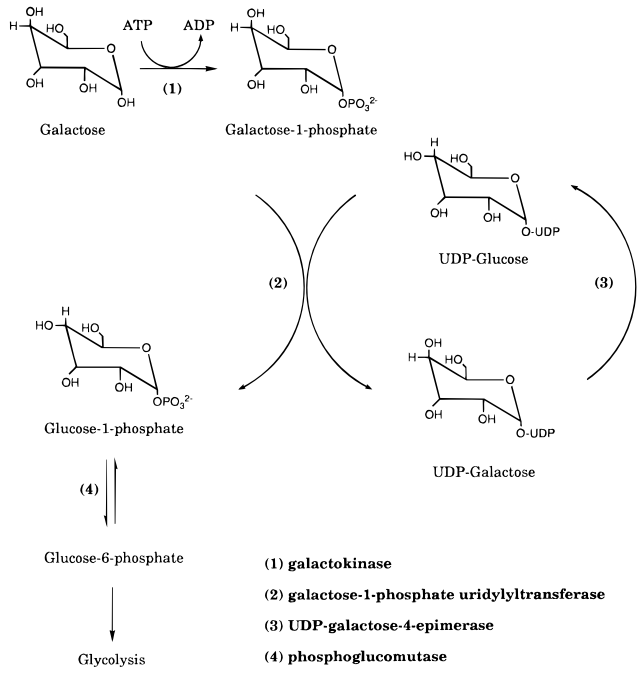
Institute for Enzyme Research, The Graduate School, and Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin−Madison, 1710 University Avenue, Madison, Wisconsin 53705

Received October 22, 1996; Revised Manuscript Received December 3, 1996[⊗]

ABSTRACT: Galactose-1-phosphate uridylyltransferase plays a key role in galactose metabolism by catalyzing the transfer of a uridine 5′-phosphoryl group from UDP-glucose to galactose 1-phosphate. The enzyme from *Escherichia coli* is composed of two identical subunits. The structures of the enzyme/UDP-glucose and UDP-galactose complexes, in which the catalytic nucleophile His 166 has been replaced with a glycine residue, have been determined and refined to 1.8 Å resolution by single crystal X-ray diffraction analysis. Crystals employed in the investigation belonged to the space group $P2_1$ with unit cell dimensions of $a =$ 68 Å, $b = 58$ Å, $c = 189$ Å, and $\beta = 100°$ and two dimers in the asymmetric unit. The models for these enzyme/substrate complexes have demonstrated that the active site of the uridylyltransferase is formed by amino acid residues contributed from both subunits in the dimer. Those amino acid residues critically involved in sugar binding include Asn 153 and Gly 159 from the first subunit and Lys 311, Phe 312, Val 314, Tyr 316, Glu 317, and Gln 323 from the second subunit. The uridylyltransferase is able to accommodate both UDP-galactose and UDP-glucose substrates by simple movements of the side chains of Glu 317 and Gln 323 and by a change in the backbone dihedral angles of Val 314. The removal of the imidazole group at position 166 results in little structural perturbation of the polypeptide chain backbone when compared to the previously determined structure for the wild-type enzyme. Instead, the cavity created by the mutation is partially compensated for by the presence of a potassium ion and its accompanying coordination sphere. As such, the mutant protein structures presented here represent valid models for understanding substrate recognition and binding in the native galactose-1-phosphate uridylyltransferase.

Galactose, a major carbohydrate constituent of dairy products, is converted into the metabolically more useful glucose 6-phosphate by a series of enzymatic reactions known as the Leloir pathway, which is outlined in Scheme 1. The focus of this investigation, galactose-1-phosphate uridylyltransferase, hereafter referred to as uridylyltransferase, catalyzes step 2 of Scheme 1, the transfer of a uridine 5′-phosphoryl group from UDP-glucose[1] to galactose 1-phosphate, thereby yielding glucose 1-phosphate and UDP-galactose. The resulting glucose 1-phosphate is converted to glucose 6-phosphate by the action of phosphoglucomutase and enters the glycolytic pathway while the UDP-galactose formed is epimerized to UDP-glucose in step 3 of Scheme 1 by UDP-galactose 4-epimerase. The most severe cases of galactosemia, a genetic disease characterized by the inability to metabolize galactose, arise from deficiencies in the uridylyltransferase activity (Kalckar, 1960).

Scheme 1: Leloir Pathway for the Metabolism of Galactose



The uridylyltransferase from *Escherichia coli* is a homodimer with each subunit containing 348 amino acids and

[†] This research was supported in part by grants from the NIH (DK47814 to H.M.H., AR35186 to I.R., and GM30480 to P.A.F.) and the NSF (BIR-9317398 shared instrumentation to I.R.).

[‡] Coordinates have been deposited in the Brookhaven Protein Data Bank (1GUP and 1GUQ).

* To whom correspondence should be addressed. I.R.: phone, 608-262-0437; fax, 608-265-2904. H.M.H.: phone, 608-262-4988; fax: 608-265-2904.

[⊗] Abstract published in *Advance ACS Abstracts,* February 1, 1997.

[1] Abbreviations: HEPPS, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; UDP-glucose, uridine 5′-diphosphate glucose; UDP-galactose, uridine 5′-diphosphate galactose; UMP-Im, uridine 5′-phosphoimidazolate.

S0006-2960(96)02651-7 CCC: $14.00

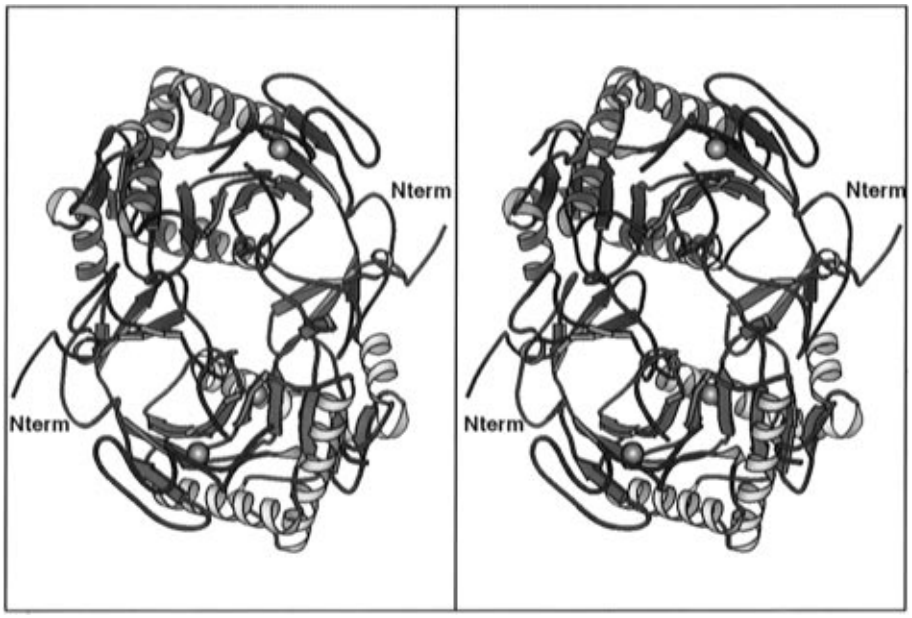


FIGURE 1: Ribbon representation of the galactose-1-phosphate uridylyltransferase from *E. coli*. X-ray coordinates employed for this figure were determined by Wedekind *et al.* (1996). The molecular 2-fold rotational axis of this dimeric enzyme is located in the center of the figure and perpendicular to the plane of the page. Subunit 1 is displayed in red and yellow while subunit 2 is shown in green and blue. The zinc and iron positions are indicated by the magenta spheres. There is a break in the polypeptide chain between Ala 36 and Gln 43 in each subunit. In this view, both the zinc ions and the polypeptide chain breaks are located toward the front. This figure and Figures 3, 4−9 were prepared with the software package MOLSCRIPT (Kraulis, 1991).

1 equiv each of zinc and iron ions (Lemaire & Müller-Hill, 1986; Cornwell *et al.*, 1987; Ruzicka *et al.*, 1995). The three-dimensional structure of the protein from *E. coli*, in complex with UDP-phenol, is known to 1.8 Å resolution and is illustrated in Figure 1 (Wedekind *et al.*, 1995). Each subunit is characterized by a nine-stranded antiparallel β-pleated sheet flanked on either side by two major α-helices. The iron ion is coordinated to His 281, His 296, His 298, and Glu 182 in a square pyramidal arrangement with Glu 182 functioning as a bidentate equatorial ligand and His 281 located in the axial position. While the iron ion is located more toward the subunit−subunit interface of the dimer where it serves as a bridge between two β-strands and an α-helix, the zinc ion is situated within approximately 8 Å of the active site and is tetrahedrally ligated to Cys 52, Cys 55, His 115, and His 164. The surface loops, defined by Ala 29 to Gln 43 in both subunits, were disordered in the original structural analysis (Wedekind *et al.*, 1995).

By various biochemical analyses, it has been demonstrated that the catalytic reaction of the uridylyltransferase proceeds by ping-pong kinetics and a double displacement mechanism, wherein a histidine residue is transiently nucleotidylated (Wong & Frey, 1974a,b; Wu *et al.*, 1974, Markus *et al.*, 1977, Yang & Frey, 1979; Hester & Raushel, 1987; Field *et al.*, 1989, Kim *et al.*, 1990). For the uridylyltransferase from *E. coli*, it has been shown that His 166 serves as the nucleophilic catalyst that is transiently modified during catalysis (Field *et al.*, 1989; Kim *et al.*, 1990). Indeed, the recent X-ray structural analysis to 1.86 Å resolution of the uridylyl-enzyme intermediate has confirmed that a covalent bond is formed between $N^{\epsilon 2}$ of His 166 and the phosphorus of UMP (Wedekind *et al.*, 1996). In addition, this high-resolution structural analysis has demonstrated that Gln 168 plays a key structural role in the stabilization of the intermediate by donating a hydrogen bond to a peripheral oxygen of the phosphoryl group as well as to the oxygen bridging the phosphorus and carbon-5′ of the uridyl moiety.

While the previous high-resolution structural investigations of the uridylyltransferase have clearly defined the polypeptide chain fold, the quaternary structure of the dimeric enzyme, and the location of the metal and nucleotide binding sites, important questions remain regarding the protein:sugar 1-phosphate interactions. It has been postulated, on the basis of the kinetic and stereochemical analyses, that galactose 1-phosphate and glucose 1-phosphate share a single binding subsite in the enzyme (Sheu *et al.*, 1979; Frey, 1982, 1992). No three-dimensional structural information on this point has been available, however.

To characterize the active site of the uridylyltransferase more fully, with special reference to the putative sugar 1-phosphate binding site, a catalytically inactive site-directed mutant of the protein, H166G, has been prepared. Although H166G does not catalyze the overall reaction of uridylyltransferase (step 2 in Scheme 1), it efficiently catalyzes the reactions of galactose 1-phosphate or glucose 1-phosphate with UMP-Im to form UDP-galactose or UDP-glucose, respectively, in reactions that are analogous to that of the uridylyl-enzyme with the sugar 1-phosphates (Kim *et al.*, 1990). The H166G mutant protein has now been crystallized in complex with either UDP-glucose or UDP-galactose. Here we describe the high-resolution X-ray analyses of these complexes.

MATERIALS AND METHODS

Molecular Biological Procedures and Protein Purification. The mutant gene of the *E. coli* galactose-1-phosphate uridylyltransferase, H166G, was prepared by site-directed mutagenesis of the replicative form of M13 containing the wild-type gene (Kim *et al.*, 1990; Kunkel, 1985). This mutant gene was transferred to the expression vector, pKK223-3 (Pharmacia) and expressed in *E. coli* CA13

Table 1: Intensity Statistics

	resolution range (Å)								
	overall	30.0−3.6	2.86	2.50	2.27	2.11	1.98	1.88	1.80
H166G/UDP-glucose									
no. of measurements	237466	46884	34629	33109	29335	26927	24502	22918	19162
no. of independent reflections	126799	16239	16612	16477	16252	16128	15653	15352	14086
% completeness	94	95	99	99	97	99	90	91	87
average $I/\sigma(I)$	35.5	81.3	38.7	22.4	16.0	12.7	9.9	7.8	6.5
R factor[a] (%)	3.4	2.5	3.5	4.5	5.5	6.5	8.1	10.1	10.8
H166G/UDP-galactose									
no. of measurements	244754	37611	48697	36422	28607	26249	24097	22817	20254
no. of independent reflections	123193	17074	16522	15918	15548	15241	14752	14449	13689
% completeness	90	98	96	94	90	91	89	86	83
average $I/\sigma(I)$	43.2	77.5	46.9	30.1	23.0	18.4	13.8	10.3	8.4
R factor[a] (%)	4.7	3.9	5.1	5.5	5.6	6.2	6.9	8.3	9.9

[a] $R_{\text{factor}} = (\sum|I - \bar{I}|/\sum I) \times 100$.

(galT118, relA1, 1-spoT1) cells (Kim *et al.*, 1990). CA13 cells containing the vector, pKF6, were grown from single colonies and subcultured in 2× YT medium (16 g of bactotryptone, 10 g of bacto yeast extract, and 5 g of NaCl per liter of distilled water) containing 100 μg/mL ampicillin. This intermediate stock was then employed to seed a batch culture (24 L of 2× YT medium + ampicillin). Cells were collected by centrifugation, frozen in liquid nitrogen, and stored at −70 °C prior to enzyme purification.

The mutant enzyme, previously referred to as UDP-hexose synthase (H166G), was purified as described with the exception that all buffers were prepared in MilliQ water without additional treatment (Ruzicka *et al.*, 1995). All buffers contained 10 mM β-mercaptoethanol.

Crystallization. X-ray quality crystals of the H166G mutant protein, complexed with either UDP-glucose or UDP-galactose, were grown by the hanging drop method of vapor diffusion at 4 °C. For such experiments, 10 μL droplets of protein at 20 mg/mL were mixed with 10 μL of 22 to 26% poly(ethylene glycol) 8000, 100 mM HEPPS, 200−300 mM KCl, and 10 mM UDP-sugar and subsequently suspended over the corresponding poly(ethylene glycol) solutions. All buffer solutions were titrated to a pH of 8.0 with potassium hydroxide.

Crystals generally appeared within 1 week and achieved maximum dimensions of 1.0 mm × 0.7 mm × 0.2 mm after 3 weeks. Both the H166G/UDP-glucose and H166G/UDP-galactose complexes crystallized in the space group $P2_1$ with unit cell dimensions of $a = 68.4$ Å, $b = 57.5$ Å, $c = 188.9$ Å, and $\beta = 100.13°$ and $a = 68.7$ Å, $b = 57.7$ Å, $c = 188.7$ Å, and $\beta = 100.08°$, respectively. The asymmetric units contained two dimers.

X-ray Data Collection and Reduction. Prior to X-ray data collection, the crystals were transferred to a synthetic mother liquor containing 26% poly(ethylene glycol) 8000, 300 mM KCl, 100 mM HEPPS, and 15 mM UDP-sugar. Following equilibration, the crystals were transferred to 28% poly-(ethylene glycol) 8000, 600 mM KCl, 100 mM HEPPS, 15 mM UDP-sugar, and 17% ethylene glycol and flash-cooled in a nitrogen stream. All X-ray data were collected at −150 °C with a Siemens HI-STAR area detector system equipped with Supper long double-focusing mirrors. The X-ray source was Cu Kα radiation from a Rigaku RU200 rotating anode generator operated at 50 kV and 90 mA and equipped with a 300 μm focal cup. Only one crystal was required per X-ray data set. The X-ray data were processed with the data reduction software package SAINT (Siemens Analytical X-ray Instruments) and scaled with the program XCALIBRE (G. Wesenberg and I. Rayment, unpublished results) to 1.8 Å resolution. Relevant X-ray data collection statistics can be found in Table 1.

Table 2: Refinement Statistics for H166G/UDP-glucose and H166G/UDP-galactose Complexes

	UDP-glucose	UDP-galactose
resolution limits (Å)	30.0−1.80	30.0−1.80
R factor (%)[a]	17.4	19.1
no. of reflections used	126300	120668
no. of protein atoms	11234	11234
no. of solvent atoms	1363	1253
(including four potassium ions)		
weighted rms deviations from ideality		
bond length (Å)	0.013	0.013
bond angle (deg)	2.37	2.38
planarity (trigonal) (Å)	0.008	0.008
planarity (other planes) (Å)	0.011	0.010
torsional angle (deg)[b]	16.4	16.5

[a] R factor $= \sum|F_o - F_c|/\sum|F_o|$, where F_o is the observed structure factor amplitude and F_c is the calculated structure factor amplitude. [b] The torsional angles were not restrained during the refinement.

Structural Determination. The structure of the H166G/UDP-glucose was solved by the method of molecular replacement with the program AMORE (Rossmann, 1972; Navaza, 1993). X-ray coordinates for the wild-type dimer complexed with UDP-phenol served as the search model (Wedekind *et al.*, 1995). Both the cross-rotational functions and translational searches were calculated with X-ray data from 8.0 to 4.0 Å. Two solutions corresponding to the two dimers in the asymmetric unit were obtained as follows: (i) $\alpha = 179.96°$, $\beta = 90.64°$, $\gamma = 270.67°$, $a = 0.3632$, $b = 0.0009$, $c = 0.3650$ and (ii) $\alpha = 179.52°$, $\beta = 90.53°$, $\gamma = 217.80°$, $a = 0.6146$, $b = 0.0722$, $c = 0.8818$. Rigid body refinement with x-ray data from 20 to 4.0 Å reduced the R factor to 35.8%. Alternate cycles of manual model building with the program FRODO (Jones, 1985) and least-squares refinement with the software package TNT (Tronrud *et al.*, 1987) reduced the R factor to 17.4% for all measured X-ray data from 30 to 1.8 Å resolution. Relevant refinement statistics are given in Table 2. Ideal stereochemistry and geometry for the UDP moiety were based on the small molecule structural investigation of Viswamitra *et al.* (1979). The final model included 1359 water molecules and four potassium ions. The identification and location of these cations will be described in the Results section.

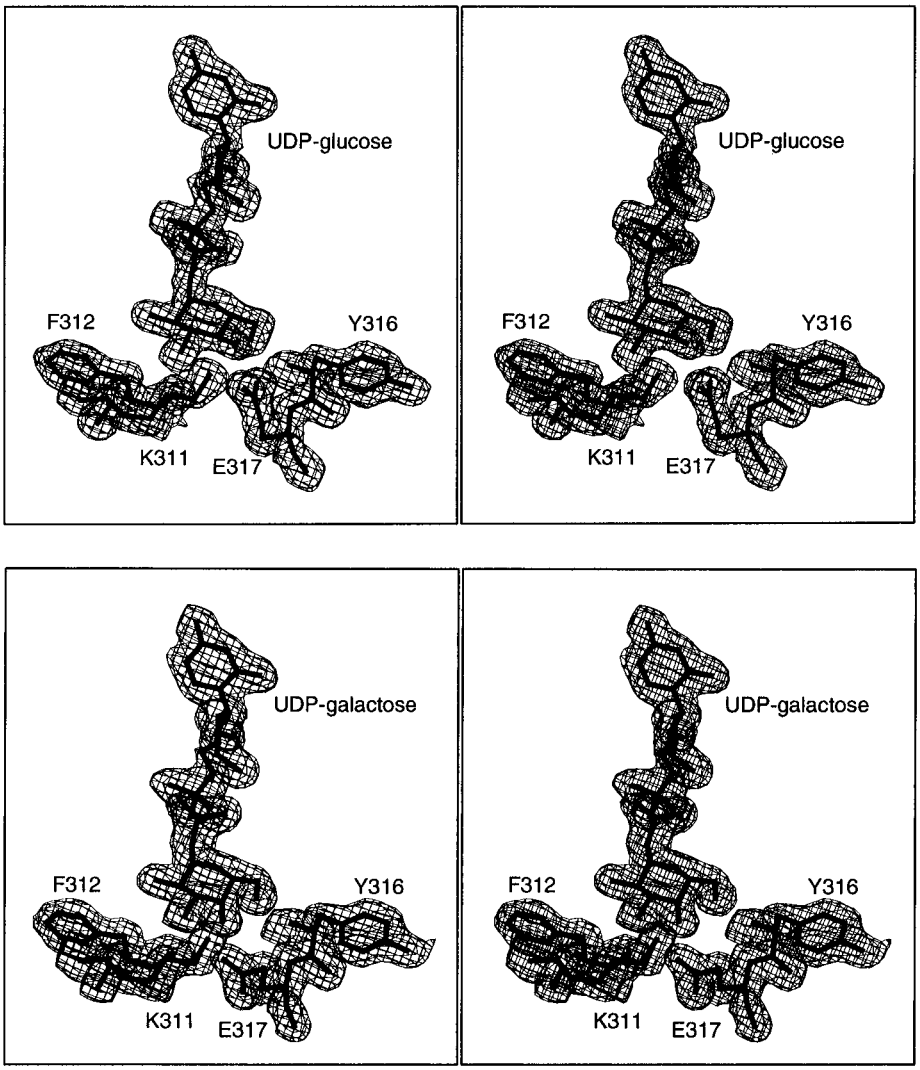


FIGURE 2: Representative electron densities for the H166G/UDP-glucose and the H166G/UDP-galactose complexes. Electron densities corresponding to Lys 311, Phe 312, Tyr 316, Glu 317, and UDP-glucose or UDP-galactose are shown in (a, top) and (b, bottom), respectively. The electron density maps displayed were calculated to 1.8 Å resolution with coefficients of the form $(2F_o - F_c)$, where F_o and F_c were the native and calculated structure factor amplitudes, respectively. The maps were contoured at 1σ. The substrates were omitted from the map calculations. This figure was prepared with the software package FROST (G. Wesenberg, University of Wisconsin—Madison).

Since the crystals of the H166G/UDP-galactose complex were isomorphous to those of the H166G/UDP-glucose species, the structure of this form of the enzyme was solved by difference Fourier methods. Alternate cycles of manual model building and least-squares refinement reduced the *R* factor to 19.1% as listed in Table 2. The final model included 1249 water molecules and four potassium ions.

RESULTS

Quality of the X-ray Models. Representative portions of the electron density maps for the H166G/UDP-glucose and H166G/UDP-galactose complexes are displayed in Figure 2. While the asymmetric units for both crystalline complexes contained two dimers, corresponding to well over 152 000 Da, the electron density maps were very well ordered. The only amino acids that were significantly disordered in the H166G/UDP-glucose complex were as follows: Thr 39 to Val 44 (subunit 1; dimer 1), Pro 40 to Leu 45 (subunit 1; dimer 2), and Ala 41 to Val 44 (subunit 2; dimer 2). The electron density corresponding to this surface loop in subunit 2; dimer 2 was reasonably well-defined. Besides these surface loops, the regions of electron densities near Pro 90 and Glu 91 were weak in all the subunits contained in the asymmetric unit. In the H166G/UDP-galactose structure the following amino acid residues were not well ordered: Thr 39 to Val 44 and Glu 91 to His 93 (subunit 1; dimer 1), Pro 90 to Glu 91 (subunit 2; dimer 1), Thr 2 to Asp 8, Arg 28 to Leu 45, and Glu 91 to His 93 (subunit 1; dimer 2), and Ala 41 to Leu 45, Glu 91 to His 93, Cys 160 to Pro 163, Phe 190 to Lys 194, and Arg 344 to Glu 345 (subunit 2; dimer 2). The first observable amino acid residue for all the subunits, in both protein/nucleotide-sugar complexes, was Thr 2. In both enzyme/substrate models, the electron densities for all subunits terminated at Glu 345, except for subunits 1; dimers 1 where electron density was observed up to the final residue, Val 348. The electron densities for the cysteines at positions 99 and 160 in all the subunits appeared modified in some manner. Since reducing agents, such as DTT, were not included in the crystallization solutions, these residues most likely became oxidized during the course of the investigation. From the original X-ray study of the uridylyltransferase, Cys 160 is known to be readily modified by β-mercaptoethanol (Wedekind *et al.*, 1995).

Ramachandran plots for both protein/substrate complexes showed that 95% of the non-glycinyl amino acids had ϕ,ψ angles lying within acceptable regions. Most of the remaining residues adopted dihedral angles falling near to the allowed regions. In the H166G/UDP-glucose model, however, Val 314, in each subunit, displayed ϕ,ψ angles of approximately $-120°$ and $-126°$, respectively. This particular residue in the H166G/UDP-galactose structure, however, adopted ϕ,ψ angles of $-140°$ and $-174°$, respectively. The difference in these observed dihedral angles resulted from the position occupied by the 4′-hydroxyl group in the UDP-glucose and UDP-galactose substrates. In the enzyme/UDP-galactose model, the carbonyl oxygen of Val 314 interacts directly with both the 4′- and 6′-hydroxyl groups of the sugar. This interaction does not occur in the enzyme/UDP-glucose complex but rather the carbonyl oxygen flips out and lies within 2.6 Å of the backbone amide nitrogen of Met 318.

For both protein/sugar complexes, the two independent dimers in the asymmetric unit are, as expected, strikingly similar. Specifically the α-carbon atoms for dimers 1 and 2 of the H166G/UDP-glucose complex are superimposable with a root-mean-square deviation of 0.24 Å. Likewise, the α-carbons for the two dimers in the asymmetric unit of the H166G/UDP-galactose complex correspond with a root-mean-square deviation of 0.27 Å. As such, the following discussions will refer only to dimers 1 of the H166G/nucleotide-sugar structures. Also, because the electron density corresponding to the surface loop delineated by Ala 36 to Gln 43 is clearest in subunit 2 of dimer 2 for both protein/substrate complexes, the following discussion will specifically refer only to this subunit unless otherwise indicated.

Description of the H166G/UDP-glucose Model. A ribbon representation of the H166G/UDP-glucose complex is displayed in Figure 3a. The disordered surface loop defined by Ala 36 to Gln 43 in the original structural investigations of the enzyme is now ordered and, as can be seen by comparing Figures 1 and 3a, forms part of the active site pocket for the neighboring subunit in the dimer. A superposition of the H166G/UDP-glucose and the uridylyl-enzyme intermediate models in this region is displayed in Figure 3b. The courses followed by the polypeptide chains in these two structures differ significantly beginning at Ser 25 where the corresponding α-carbon positions are separated by 1.1 Å. At Ala 29 and Lys 30, the polypeptide chains for these two protein structures have diverged to the extent that the α-carbons for these residues are separated by 5 and 7 Å, respectively. The two models are again superimposed at Leu 45. Other than this surface loop region and the first eight N-terminal residues, the two forms of the uridylyltransferase are very similar such that their α-carbons are superimposed with a root-mean-square deviation of 0.53 Å.

A close-up view of the active site for the H166G/UDP-glucose structure (subunit 2) is shown in Figure 4a. In the orientation depicted in Figure 4a, the top portion of the substrate binding pocket is formed by amino residues contributed by subunit 2 whereas the lower half of the active site is lined by Arg 28, Arg 31, Lys 311 to Tyr 316, and Gln 323 from subunit 1. For the most part the binding pocket is hydrophilic with the exception of Phe 312 and Tyr 316 from subunit 1. There are four ordered water molecules in the active site pocket, two of which interact with the 3′- and 4′-hydroxyl groups of the glucose, respectively, and two of which interact with the α-phosphoryl oxygens. All of these solvent/substrate interactions are within 2.8 Å.

A schematic of potential electrostatic interactions between the protein and the UDP-glucose is given in Figure 4b. The uridine moiety is anchored to the protein through interactions with the carboxylate side chain of Asp 78 and the backbone amide hydrogens of Asp 78 and Val 61 while the 2′- and 3′-hydroxyl groups of the uridyl ribose lie within hydrogen-bonding distance of the carboxamide side chain of Asn 77. Other side chain groups contributed by subunit 2 that participate in substrate binding include Asn 153 where $N^{\delta 2}$ lies within hydrogen-bonding distance to both the ring oxygen of glucose and one of the α-phosphoryl oxygens, Ser 161 where O^{γ} is located at 2.6 Å from one of the β-phosphoryl oxygens, and Gln 168 where $N^{\epsilon 2}$ is hydrogen bonded to the 5′-hydroxyl oxygen of the uridyl ribose and to both α- and β-phosphoryl oxygens as indicated in Figure 4b. The backbone carbonyl oxygen of Gly 159 in subunit 2 is hydrogen bonded to the 2′-hydroxyl group of glucose while the backbone amide hydrogen of Ser 161 lies within 2.8 Å of an α-phosphoryl oxygen. Additional interactions between the H166G mutant protein and the UDP-glucose substrate are provided by the side chains of Arg 28, Arg 31, Lys 311, Glu 317, and Gln 323 in subunit 1 and are quite extensive as indicated in Figure 4b. Further hydrogen bonds provided by subunit 1 to the substrate include those formed between the carbonyl oxygen of Phe 312 and the 3′-hydroxyl group of glucose and the amide nitrogen of Tyr 316 and the 6′-hydroxyl group of the hexose.

During the course of the least-squares refinement of the H166G/UDP-glucose model, four putative water molecules, one in each subunit of the asymmetric unit, refined with anomalously low temperature factors. The observed coordination geometry surrounding these solvent molecules suggested that, in fact, they were cations, and most likely potassium ions, since all buffers were titrated with potassium hydroxide. Furthermore, attempts to grow crystals under the same precipitant conditions using buffers titrated with sodium hydroxide and in the absence of potassium chloride were unsuccessful (Thoden, unpublished results). As such these solvents were subsequently modeled as potassium ions and refined with *B*-values ranging from 15.8 to 20.6 $Å^2$ for the ions in subunits 1 and 4, respectively. These ions reside near the H166G mutation in each subunit and most likely serve to stabilize the polypeptide chain backbone in this region. In subunits 2 and 4, the potassium ion is ligated by four solvent molecules and two carbonyl oxygens provided by Asn 153 and Gly 166 in a distorted octahedral configuration. In subunit 3, the coordination geometry of the ion lacks the fourth water molecule and thus can be described as a distorted trigonal bipyramidal arrangement. As shown in Figure 5, the potassium ion in subunit 1 is surrounded by seven oxygen-containing ligands. In all the subunits, oxygen−cation bond distances range from 2.5 to 3.0 Å. These potassium ions are likewise observed in the H166G/UDP-galactose complex.

Description of the H166G/UDP-galactose Model. The polypeptide chains for the H166G/UDP-glucose and UDP-galactose complexes are virtually identical and superimposable with a root-mean-square deviation of 0.21 Å for 691 α-carbon positions (dimers 1 in the asymmetric unit). A schematic of the potential hydrogen bonds between the

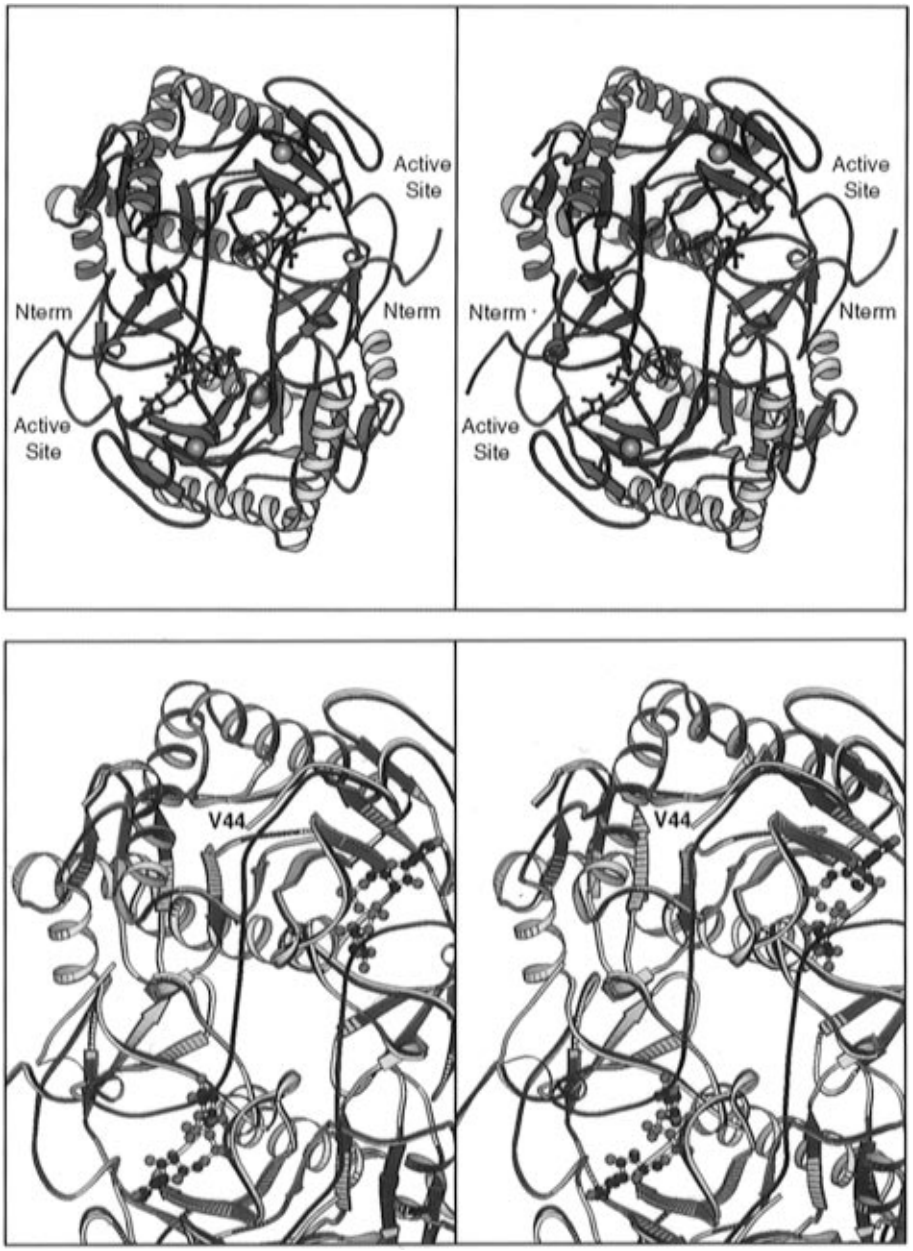


FIGURE 3: Ribbon representation of the H166G/UDP-glucose complex. (a, top) The color coding and view are the same as described for Figure 1. The UDP-glucose molecules are depicted in ball-and-stick representations. Note that the polypeptide chain near Ala 36 to Gln 43 is no longer broken. (b, bottom) A superposition of the H166G/UDP-glucose and the uridylyl-enzyme intermediate polypeptide chains near Val 44 is shown. The uridylyl-enzyme intermediate protein backbone is depicted in yellow while the H166G/UDP-glucose complex is displayed in blue. Note that this loop, provided by one subunit, forms part of the active site for the second subunit in the dimer.

galactose substrate and the protein is given in Figure 6. As can be seen, the only major structural differences arising from the binding of UDP-glucose or UDP-galactose molecules occur near the 4′-hydroxyl groups of the hexoses. In the case of the enzyme/UDP-glucose structure, a water molecule is hydrogen bonded to the 4′-hydroxyl group while in the enzyme/UDP-galactose model the carbonyl oxygen of Val 314 provides this electrostatic interaction. Likewise, the side chain carboxamide group of Gln 323, which interacts with the 4′-hydroxyl group of glucose, is displaced from the binding pocket when UDP-galactose is bound to the H166G mutant protein. These differences are highlighted in the superposition displayed in Figure 7. The most significant rearrangement of side chains occurs at Glu 317 where the position of $O^{\epsilon 2}$ differs by 2.5 Å depending upon the configuration at carbon-4 of the sugar present in the active site.

DISCUSSION

Comparison of the H166G Mutant Protein/UDP-glucose Complex and the Wild-Type Enzyme/UDP-phenol Model. The structural analyses of the H166G mutant/UDP-glucose and UDP-galactose complexes have allowed for a detailed description of the sugar binding site. The enzyme employed in this investigation, however, is that of a site-directed mutant. Consequently the question as to whether the present models represent a valid description of sugar positioning in the native enzyme must be addressed. Note that the only regions where the mutant and wild-type enzymes differ significantly are delineated by Pro 26 to Leu 45 and Ala 157 to His 164. Excluding these amino acid residues from the calculation, the α-carbon positions for these two forms of the enzyme coincide with a root-mean-square deviation of 0.52 Å. A superposition of the H166G /UDP-glucose

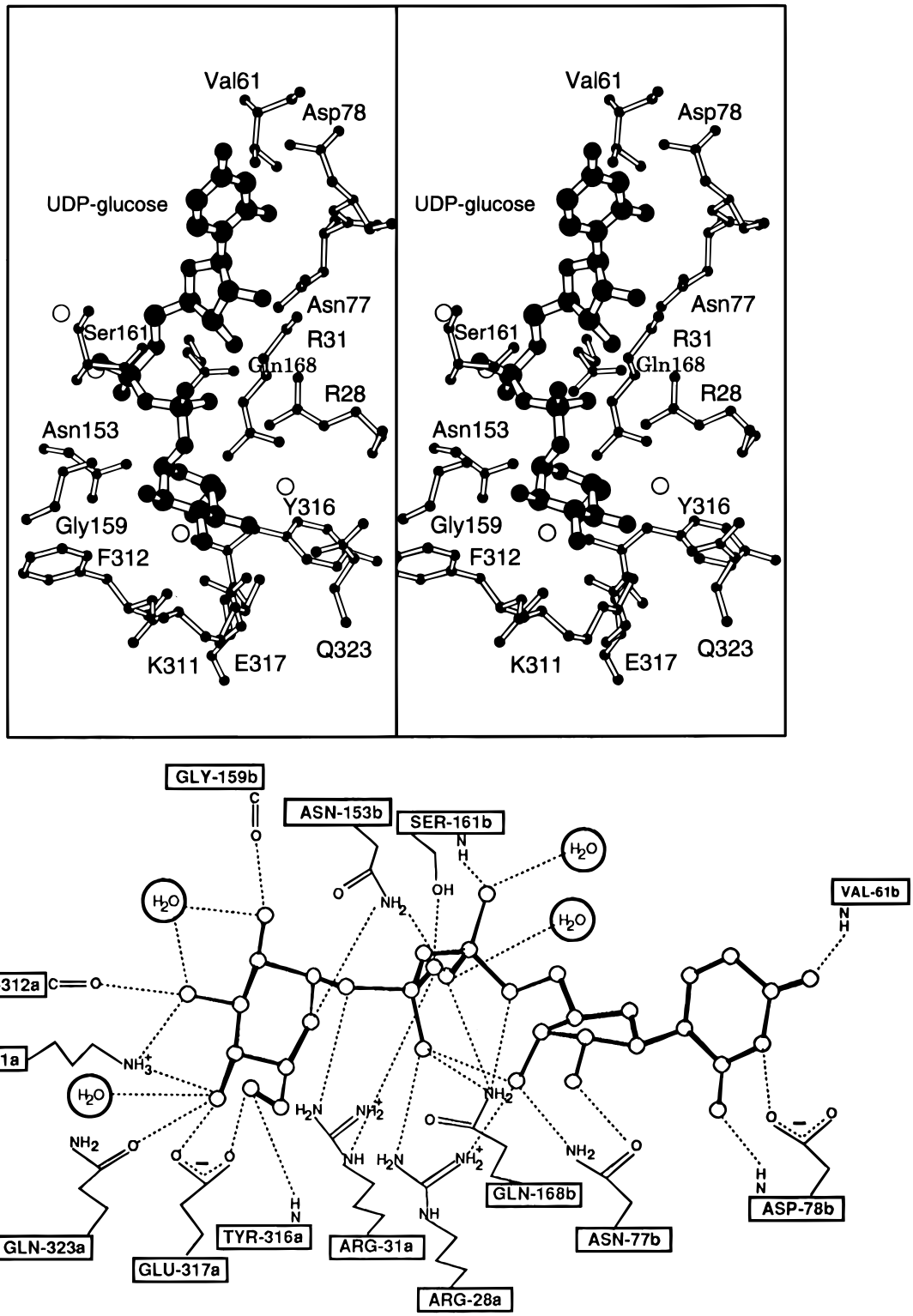


FIGURE 4: Close-up views of the active site for the enzyme/UDP-glucose complex. (a, top) Only those amino acids that are located within 3.2 Å of the substrate are displayed. The white spheres represent ordered water molecules. Those residues indicated by the three- and one-letter codes correspond to amino acids provided by subunits 2 and 1, respectively. (b, bottom) Schematic representation of the electrostatic interactions between the protein and the substrate are indicated by the dashed lines. These distances are equal to or less than 3.2 Å. Those residues with the suffix "a" belong to subunit 1 while those with the suffix "b" are provided by subunit 2 of dimer 1 in the asymmetric unit.

model upon the structure for the native enzyme with bound UDP-phenol near the mutation is displayed in Figure 8. As can be seen, the trace of the polypeptide chain backbone near Gly 166 in the mutant protein is nearly identical to that observed in the wild-type enzyme. Specifically, the α-carbons for Pro 165, H166G, and Gly 167 are superimposed with a root-mean-square deviation of 0.49 Å. Large changes between the two protein backbones occur at the N-terminal side of the mutation and between Met 158 to His 164. A difference in dihedral angles at Cys 160 of $\phi = -97°$, $\psi = 115°$ for the native enzyme and $\phi = -110°$, $\psi = 1°$ for the mutant protein results in the α-carbons for Ser 161 being

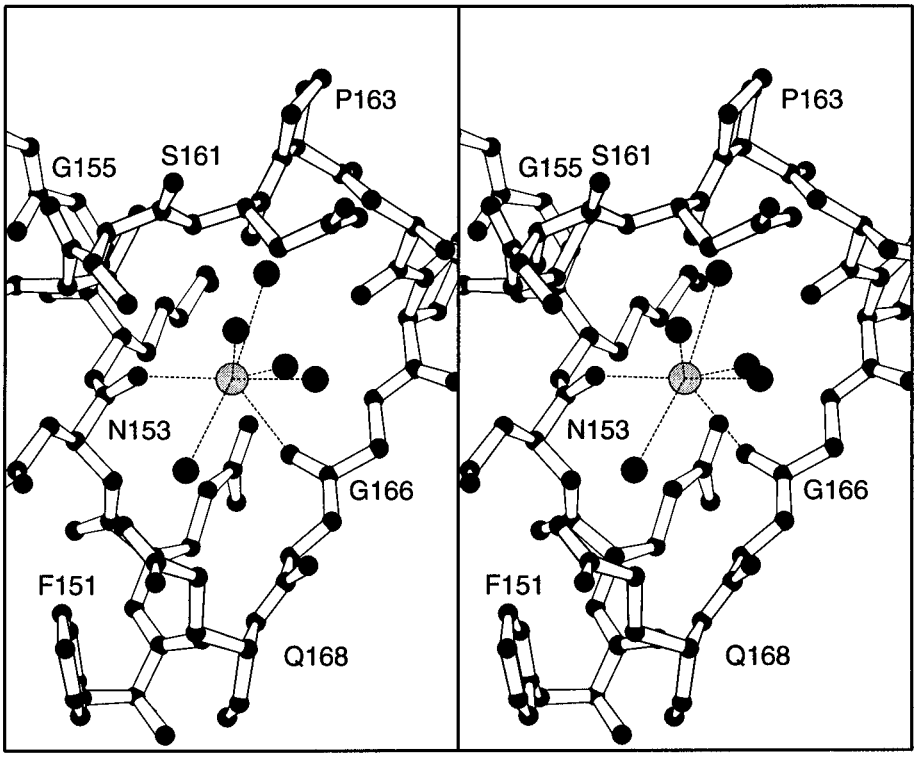


FIGURE 5: Close-up view of the potassium binding site in subunit 1, dimer 1. Water molecules are represented as solid spheres. The coordination geometry is indicated by the dashed lines.

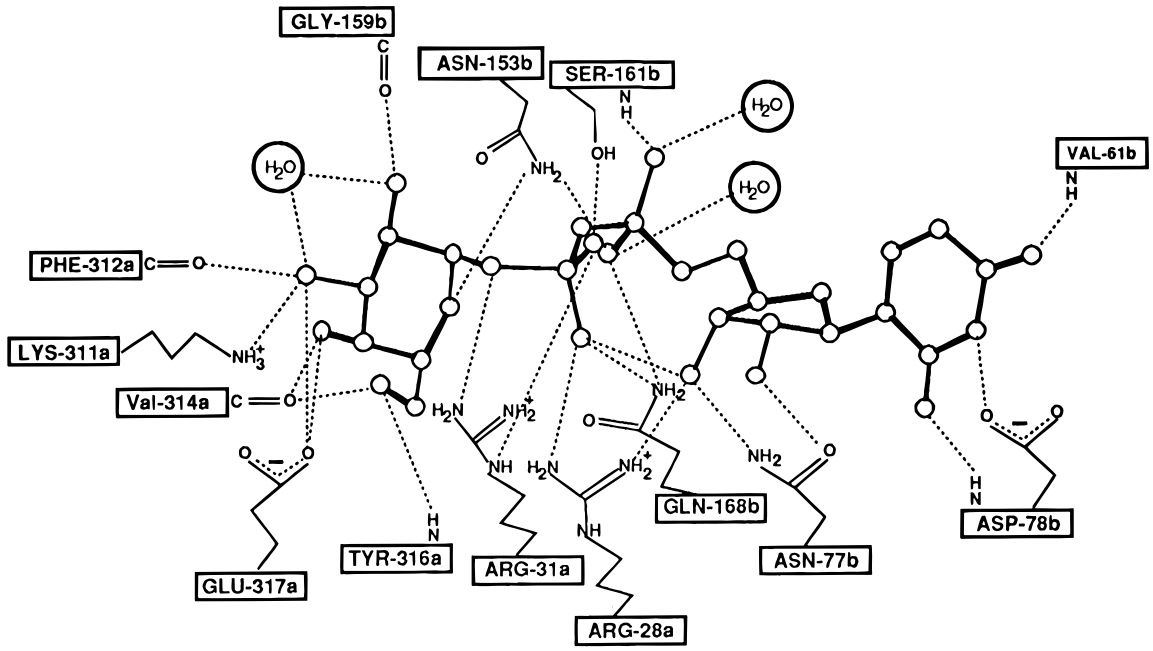


FIGURE 6: Schematic representation of the hydrogen-bonding pattern observed between the H166G mutant protein and the UDP-galactose substrate. The distances indicated by the dashed lines are equal to or less than 3.2 Å. Those residues with the suffix "a" belong to subunit 1 while those with the suffix "b" are provided by subunit 2 of dimer 1 in the asymmetric unit.

separated by 4.6 Å. The observed three-dimensional differences between the two protein forms can be attributed solely to substrate binding rather than to gross structural changes arising from the mutation, and thus the models described here provide meaningful descriptions of the active site geometry.

In addition to the structural similarity of H166G to the native uridylyltransferase, the significance and relevance of the nucleotide sugar binding to H166G are strongly supported by the fact that the mutant protein catalyzes the reaction of UMP-Im with sugar 1-phosphates to form UDP-sugars. H166G is quite active in catalyzing this unique reaction, and because the native uridylyltransferase does not catalyze it, H166G has been named UDP-hexose synthase. The synthase activity is very similar to the reaction of the native uridylyl-enzyme with sugar 1-phosphates in the normal uridylyltransferase reaction, and this similarity strongly supports the contention that the H166G/UDP-hexose struc-

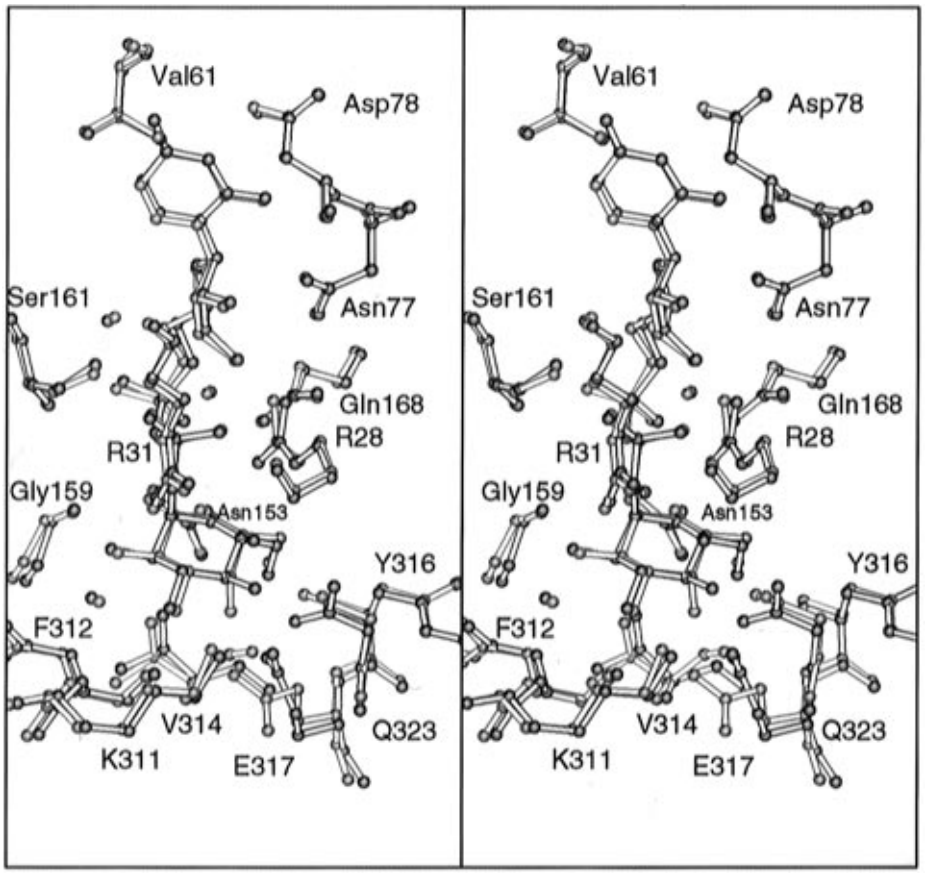


FIGURE 7: Superposition of the active sites for the protein/UDP-glucose and protein/UDP-galactose complexes. The enzyme/UDP-glucose and enzyme/UDP-galactose models are shown in black and red, respectively.

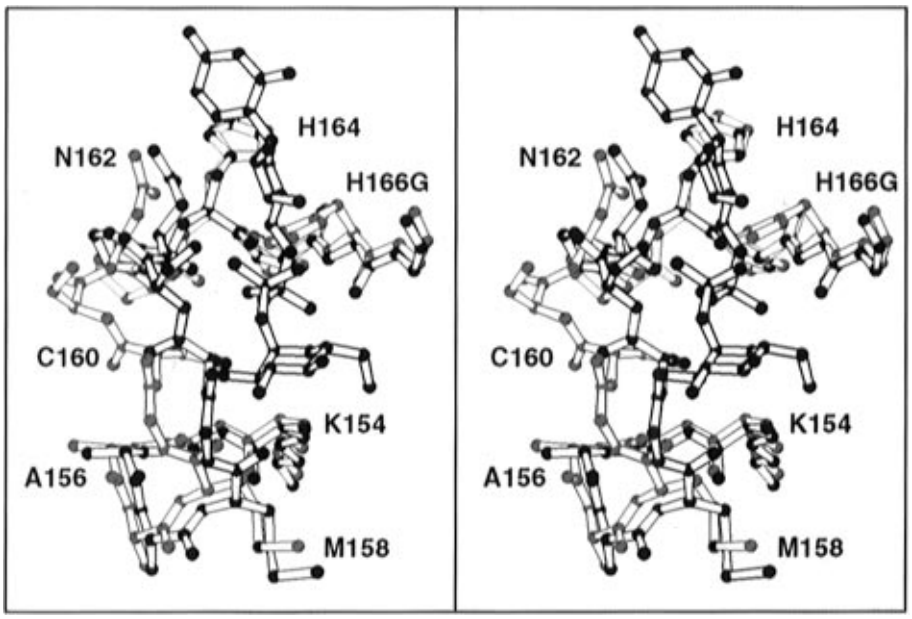


FIGURE 8: Superposition of the H166G enzyme/UDP-glucose and wild-type enzyme/UDP-phenol models near position 166. The native enzyme is shown in red while the mutant protein is depicted in black. The UDP-glucose molecule is shown in black.

tures are closely related to those of the uridylyltransferase/UDP-hexose Michaelis complexes, whose structures cannot be determined directly.

Comparison of the H166G Mutant Protein/UDP-glucose Complex and the Uridylyl-Enzyme Intermediate Model. As displayed in Figure 9, superposition of the structures of the H166G/UDP-glucose complex, the wild-type enzyme/UDP-phenol complex, and the uridylyl-enzyme intermediate indicates the direction of attack of the α-phosphate onto the nucleophilic $N^{\epsilon 2}$ of His 166. From the preceding discussion regarding the binding of UDP-glucose, it is apparent that the polypeptide chain from Gly 159 to Pro 163 forms important interactions with the hexose. The comparison of the uridylyl-enzyme intermediate structure with that of the H166G/UDP-glucose complex demonstrates that this region of the polypeptide chain, together with the loop from the adjacent subunit, serves to position the α-phosphate of the substrate opposite to $N^{\epsilon 2}$ of His 166. In the H166G mutant

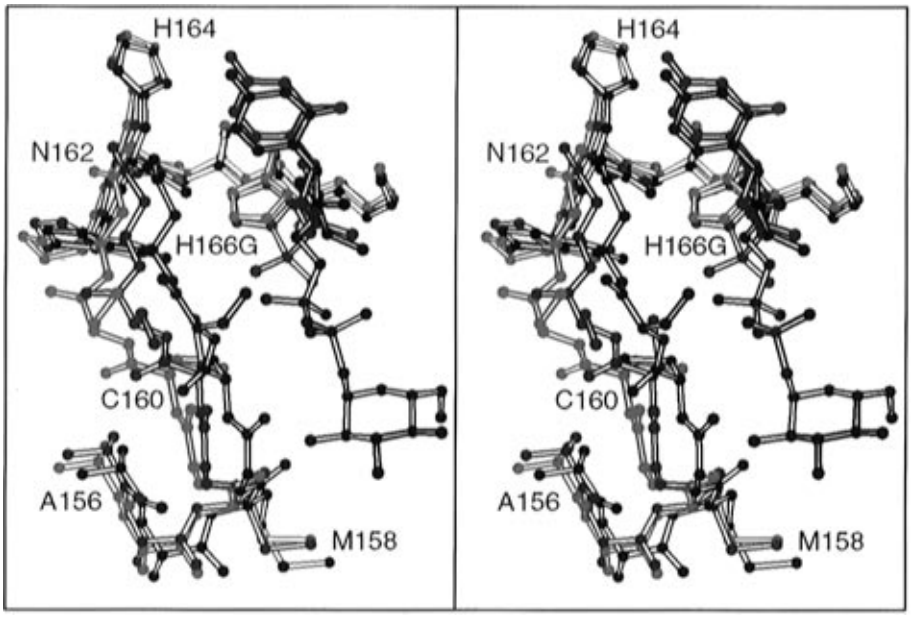


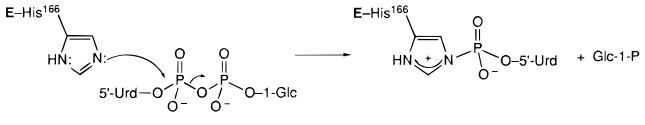
FIGURE 9: Superposition of the H166G enzyme/UDP-glucose, the wild-type enzyme/UDP-phenol, and the uridylyl-enzyme intermediate models near position 166. The wild-type enzyme/UDP-phenol model is shown in red, the H166G mutant protein with UDP-glucose is depicted in black, and the uridylyl-enzyme intermediate is displayed in blue.

protein, the loop from Gly 159 to Pro 163 moves across the sugar binding cleft, whereas in the uridylyl- and UDP-phenol complexes, the cleft is more open as shown in Figure 9. It was proposed earlier that the role of the conserved Cys 160 was to prevent the transfer of a proton from a water molecule to the α-phosphate of the uridylyl-enzyme intermediate, thereby preventing nonproductive hydrolysis (Wedekind *et al.*, 1996). The H166G/UDP-glucose structure described here reveals that, although the loop containing Cys 160 has moved into the active site to interact with the sugar, the side chain functional group of this conserved cysteine is not involved in the binding of the α-phosphate as observed in the uridylyl-enzyme intermediate. Rather, the α-phosphate is hydrogen bonded to the amide hydrogen of Ser 161. These conformational changes are consistent with an active site that favors reaction of the uridylyl-enzyme intermediate with a hexose 1-phosphate instead of water.

By mutating His 166 to a glycine residue, it has been possible to more thoroughly define the active site geometry of the galactose-1-phosphate uridylyltransferase from *E. coli*. From this investigation it is now known that the active site is formed from amino acid residues contributed by both subunits. The enzyme is able to bind the glycosyl moieties of both UDP-galactose and UDP-glucose at the same subsite by simple movements of several side chains, in particular Glu 317 and Gln 323, in accommodation of the stereochemical difference at the glycosyl C-4 position.

In addition, and somewhat surprisingly, the cavity created by substituting a glycine residue for the imidazole side chain of His 166 is partially compensated for by a potassium ion and its accompanying coordination sphere. Clearly the replacement of the histidine residue with a cation is important for the stabilization of the polypeptide chain backbone in this vicinity. The fact that crystals could not be obtained in the presence of sodium, as opposed to potassium, further suggests that the ionic radius of the cation and its ligation preference are also critical for stabilization. The fact that a cation is bound by the cavity created in H166G may have relevance for the mechanism of uridylyl group transfer. Upon

Scheme 2: Uridylylation of His 166 in Galactose-1-phosphate Uridylyltransferase



movement of the uridylyl group from a UDP-sugar to His 166, as illustrated in Scheme 2, the imidazole ring of histidine would either acquire a positive charge as shown, or a general base would have to accept a proton from the imidazolium ring. The structure of the covalent uridylyl-enzyme does not reveal the presence of a general base in the vicinity of His 166 (Wedekind *et al.*, 1996). Therefore, given that potassium ion binds to H166G/UDP-hexoses in place of the missing side chain, the imidazole ring of His 166 is likely to be positively charged in the covalent intermediate, and this would be its most reactive form.

Since those residues that participate in sugar binding have now been identified, it is possible to explain some of the molecular causes for galactosemia in three-dimensional terms. Numerous human mutations for this autosomally inherited disease have been identified including L74P, A81T, F171S, Q188R, and K334R (Elsas *et al.*, 1994; Reichardt *et al.*, 1991, 1992; Fridovich-Keil & Jinks-Robertson, 1993). On the basis of amino acid sequence alignments, these residues correspond to positions Leu 54, Val 61, Phe 151, Gln 168, and Lys 311, respectively, in the enzyme from *E. coli* (Wedekind, 1995). From previous investigations it is known that both Leu 54 and Val 61 are located within approximately 4.5 and 3.0 Å, respectively, of the uridine ring of the substrate (Wedekind *et al.*, 1995, 1996). From the present analysis it is now known that Phe 151 is positioned within 4 Å of the sugar moiety. Clearly, these three residues provide important hydrophobic surfaces for the active site. Both Gln 168 and Lys 311, on the other hand, participate in direct electrostatic interactions with the substrate. Lys 311 anchors both the 3′- and 4′-hydroxyl groups of the glucose substrate to the protein, and substitution of this residue with

an arginine side chain, as in the case of the human K334R galactosemic mutant, would most likely disrupt this interaction. Additionally, the enzyme/UDP-glucose and UDP-galactose models described here demonstrate that Gln 168 plays an essential role by binding to the phosphate backbone of the substrate. Furthermore, in the uridylyl-enzyme intermediate model, determined by Wedekind *et al.* (1996), Gln 168 also forms tight hydrogen bonds to the phosphoryl group. Consequently, it can be speculated that replacement of this glutamine by an arginine residue, as observed in human galactosemia, results in the overstabilization of the enzyme intermediate, thereby compromising its subsequent reaction with galactose 1-phosphate. The roles of these various residues in the proper functioning of galactose-1-phosphate uridylyltransferase are presently being addressed by additional site-directed mutagenesis experiments and X-ray crystallographic analyses.

REFERENCES

Cornwell, T. L., Adhya, S. L., Reznikoff, W. S., & Frey, P. A. (1987) *Nucleic Acids Res. 15*, 8116.

Elsas, L. J., Dembure, P. P., Langley, S., Paulk, E. M., Hjelm, L. N., & Fridovich-Keil, J. (1994) *Am. J. Hum. Genet. 54*, 1030−1036.

Field, T. L., Reznikoff, W. S., & Frey, P. A. (1989) *Biochemistry 28*, 2094−2099.

Frey, P. A. (1982) in *New Comprehensive Biochemistry 3. Stereochemistry* (Tamm, Ch., Ed.) pp 201−248, Elsevier, Amsterdam.

Frey, P. A. (1992) *Enzymes 20*, 141−186.

Fridovich-Keil, J. L., & Jinks-Robertson, S. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 398−402.

Hester, L. S., & Raushel, F. M. (1987) *J. Biol. Chem. 262*, 12092−12095.

Jones, T. A. (1985) *Methods Enzymol. 115*, 157−171.

Kalckar, H. M. (1960) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 19*, 984−990.

Kim, J., Ruzicka, F., & Frey, P. A. (1990) *Biochemistry 29*, 10590−10593.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488−492.

Lemaire, H. G., & Müller-Hill, B. (1986) *Nucleic Acids Res. 14*, 7705−7711.

Markus, H. B., Wu, J. W., Boches, F. S., Tedesco, T. A., Mellman, W. J., & Kallen, R. G. (1977) *J. Biol. Chem. 252*, 5363−5369.

Navaza, J. (1993) *Acta Crystallogr. D49*, 588−591.

Reichardt, J. K. V., Packman, S., & Woo, S. L. C. (1991) *Am. J. Hum. Genet. 49*, 860−867.

Reichardt, J. K. V., Levy, H. L., & Woo, S. L. C. (1992) *Biochemistry 31*, 5430−5433.

Rossmann, M. G. (1972) *The Molecular Replacement Method*, Gordon and Breach, New York.

Ruzicka, F. J., Wedekind, J. E., Kim, J., Rayment, I., & Frey, P. A. (1995) *Biochemistry 34*, 5610−5617.

Sheu, K.-F., Richard, J. P., & Frey, P. A. (1979) *Biochemistry 18*, 5548−5556.

Tronrud, D. E., Ten Eyck, L. F., & Matthews, B. W. (1987) *Acta Crystallogr., Sect. A 43*, 489−501.

Viswamitra, M. A., Post, M. L., & Kennard, O. (1979) *Acta Crystallogr. B35*, 1089−1094.

Wedekind, J. E. (1995) Ph.D. Thesis, University of Wisconsin, Madison, WI.

Wedekind, J. E., Frey, P. A., & Rayment, I. (1995) *Biochemistry 34*, 11049−11061.

Wedekind, J. E., Frey, P. A., & Rayment, I. (1996) *Biochemistry 35*, 11560−11569.

Wong L.-J., & Frey, P. A. (1974a) *Biochemistry 13*, 3889−3894.

Wong, L.-J., & Frey, P. A. (1974b) *J. Biol. Chem. 249*, 2322−2324.

Wu, J. W., Tedesco, T. A., Kallen, R. G., & Mellman, W. J. (1974) *J. Biol. Chem. 249*, 7038−7039.

Yang, S.-L. L., & Frey, P. A. (1979) *Biochemistry 18*, 2980−2984.

BI9626517