Crystallographic Evidence for Tyr 157 Functioning as the Active Site Base in Human UDP−Galactose 4-Epimerase[†,‡]

James B. Thoden,[§] Travis M. Wohlers,[||] Judith L. Fridovich-Keil,[⊥] and Hazel M. Holden*[,§]

Department of Biochemistry, University of Wisconsin, Madison, Wisconsin, 53705, and Graduate Program in Genetics and Molecular Biology and Department of Genetics, Emory University School of Medicine, Atlanta, Georgia 30322

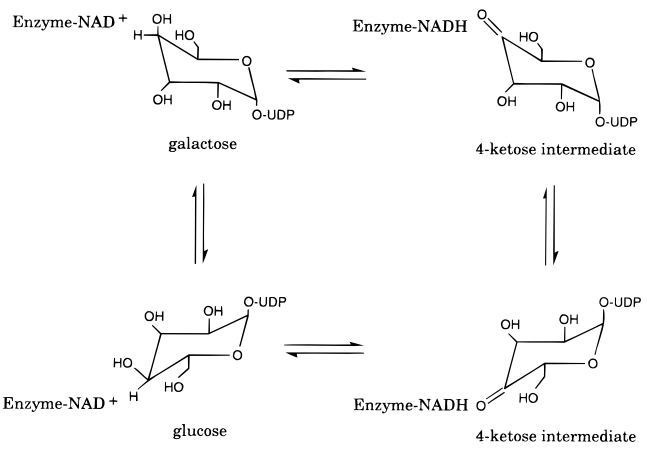
Received February 1, 2000; Revised Manuscript Received March 8, 2000

ABSTRACT: UDP-galactose 4-epimerase catalyzes the interconversion of UDP-glucose and UDP-galactose during normal galactose metabolism. In humans, deficiencies in this enzyme lead to the complex disorder referred to as epimerase-deficiency galactosemia. Here, we describe the high-resolution X-ray crystallographic structures of human epimerase in the resting state (i.e., with bound NAD^+) and in a ternary complex with bound NADH and UDP-glucose. Those amino acid side chains responsible for anchoring the NAD^+ to the protein include Asp 33, Asn 37, Asp 66, Tyr 157, and Lys 161. The glucosyl group of the substrate is bound to the protein via the side-chain carboxamide groups of Asn 187 and Asn 207. Additionally, O^{γ} of Ser 132 and O^{η} of Tyr 157 lie within 2.4 and 3.1 Å, respectively, of the 4′-hydroxyl group of the sugar. Comparison of the polypeptide chains for the resting enzyme and for the protein with bound NADH and UDP-glucose demonstrates that the major conformational changes which occur upon substrate binding are limited primarily to the regions defined by Glu 199 to Asp 240 and Gly 274 to Tyr 308. Additionally, this investigation reveals for the first time that a conserved tyrosine, namely Tyr 157, is in the proper position to interact directly with the 4′-hydroxyl group of the sugar substrate and to thus serve as the active-site base. A low barrier hydrogen bond between the 4′-hydroxyl group of the sugar and O^{γ} of Ser 132 facilitates proton transfer from the sugar 4′-hydroxyl group to O^{η} of Tyr 157.

The human enzyme UDP-galactose 4-epimerase (hGALE), hereafter referred to as epimerase, normally catalyzes the second step in the Leloir pathway of galactose metabolism, namely the interconversion of UDP-galactose and UDP-glucose as shown in Scheme 1. Impairment of this enzyme results in the complex disorder epimerase-deficiency galactosemia, which affects as many as 1/6700 individuals, at least in some ethnic groups (*1*, *2*). Clinically, epimerase-deficiency galactosemia can range from benign to severe (*3*−*8*). The molecular basis of this variance remains unclear, although allelic heterogeneity observed in the patient population (*3*, *4*, *9*, *10*) raises the intriguing possibility of a genotype/phenotype correlation.

UDP-galactose 4-epimerase from *Escherichia coli* has been the focus of numerous structural investigations within recent years (*11*−*15*). The bacterial enzyme is a homodimer with each subunit containing 338 amino acid residues and one tightly bound NAD^+ or NADH moiety (*12*, *16*, *17*). From these studies, it is now known that each subunit folds into two distinct domains: an N-terminal or nucleotide-binding motif consisting of a seven-stranded parallel β-pleated sheet flanked on either side by α-helices and a smaller C-terminal domain responsible for binding the UDP-sugar substrates. The active site is positioned within the crevice between these two domains.

Scheme 1



As indicated in Scheme 1, there are three key features in the reaction mechanism of epimerase: (1) abstraction of the 4′-hydroxyl hydrogen of the sugar by an enzymatic base, (2) transfer of a hydride from C4 of the sugar to the C4 of NAD^+ leading to a 4′-ketopyranose intermediate and NADH, and (3) rotation of the resulting 4′-ketopyranose intermediate in the active site to allow return of the hydride from NADH to the opposite face of the sugar. Recent X-ray crystal-

[†]This research was supported in part by grants from the NIH (DK47814 to H.M.H. and DK46403 to J.L.F.-K.). T.M.W. was supported in part by funds provided by NIH Predoctoral Training Grant GM08490.

[‡] X-ray coordinates have been deposited in the Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (1EK5 and 1EK6) and will be released upon publication.

* To whom correspondence should be addressed. E-mail: Hazel_Holden@biochem.wisc.edu. Phone: (608) 262-4988. Fax: (608) 262-1319.

[§] Department of Biochemistry.

[||] Graduate Program in Genetics and Molecular Biology.

[⊥] Department of Genetics.

10.1021/bi000215l CCC: $19.00
Published on Web 04/19/2000

lographic and site-directed mutagenesis investigations have suggested that the active-site base is not a simple amino acid residue but rather a combination of side chains contributed by Ser 124, Tyr 149, and Lys 153 which act in concert to abstract the 4′-hydroxyl hydrogen from the UDP-sugar (*14*, *18*). Additionally, these studies have shown that the epimerase is a member of the "short-chain" dehydrogenase family which contains a characteristic Tyr-Lys couple (*19*, *20*). Members of this family are involved in a variety of metabolic processes including fertility, growth, and hypertension, among others (*21*). With respect to the rotation of the presumed 4′-ketopyranose intermediate, a recent structural investigation has revealed the manner in which both UDP-glucose and UDP-galactose are accommodated within the active site (*15*). This study has demonstrated that Asn 179 and Tyr 299 form a critical template for the proper positioning of the two different sugars and, further, has shown that the active site is large enough to allow for the presumed rotation of the 4′-ketopyranose intermediate.

Homology-modeling studies of the human epimerase sequence onto the *E. coli* protein model have offered useful predictions regarding the functional impact of specific patient mutations on enzyme function. For example, one of the first patient hGALE alleles to be identified encoded a substitution of serine for asparagine at residue 34 (N34S). Under standard assay conditions of excess substrate and excess NAD^+, the N34S protein demonstrated approximately 70% wild-type levels of activity (*4*). Inspection of the *E. coli* epimerase crystal structure revealed that residue N32, which corresponds to N34 in the human sequence, forms a hydrogen bond with the adenine ring of the enzyme-bound NAD^+ (*12*). A substitution of serine at that position would be expected to interfere with this interaction. As predicted, subsequent studies of the N34S protein activity versus wild-type as a function of NAD^+ concentration revealed that under conditions where NAD^+ became limiting, the discrepancy between the mutant and wild-type epimerase activities became much more pronounced (*4*).

Although the predictions regarding the N34S mutant protein proved useful, such cross-species predictions are risky at best considering that the human and *E. coli* enzymes share only 55% amino acid sequence identity. Strikingly, human epimerase catalyzes an additional reaction not reported for the bacterial enzyme, namely the interconversion of UDP−GalNAc[1] and UDP-GlcNAc (*8*, *22*−*24*). Why can human epimerase interconvert both sets of substrates, while the bacterial enzyme only interconverts one? Furthermore, why do some patient mutations in epimerase, for example, V94M (*8*), result in differential impairment of the two enzymatic activities? Why do some pairs of patient hGALE alleles interact, at least in a yeast model system, in a partial dominant negative manner (*4*), while others do not (*8*)? What is the active-site base in the human enzyme?

Here we describe the high-resolution X-ray crystal structures of human epimerase with bound NAD^+ and in a ternary complex with NADH and UDP-glucose. These studies serve as the initial template for addressing the various structural/functional questions outlined above.

MATERIALS AND METHODS

Expression of Human Epimerase in Pichia pastoris. Human UDP-galactose 4-epimerase was overexpressed in the methylotrophic yeast *Pichia pastoris*. In brief, a 1.1 kb *Eco*RI−*Sal*I fragment containing the wild-type epimerase coding sequence was incubated with Klenow enzyme and then ligated into the high copy number *Pichia* expression plasmid pPIC3.5K (Invitrogen), which had been linearized with *Sna*BI. Correct orientation of the insert was verified by restriction digestion. The resultant plasmid, pPIC3.5KhGALE, was then linearized using *Sal*I and integrated in multiple copy into the genome of *Pichia* strain GS115, as recommended by Invitrogen. Transformants were selected initially on histidine-deficient medium and subsequently on increasing concentrations of G418 (U.S. Biological) to reveal the highest level expressing clones. Expression of human epimerase in these clones was verified by western blot analysis using a rabbit polyclonal antiserum (EU69) raised against purified hexahistidine-tagged human epimerase (*8*). The clone demonstrating the highest level of epimerase expression was then expanded and cultured in a New Brunswick Scientific Bioflo 3000 fermenter, harvested, and lysed by agitation with glass beads in breaking buffer (50 mM sodium phosphate pH 7.4, 1 mM PMSF, 1 mM EDTA and 5% glycerol) using a Beadbeater (Biospec). Cell lysates were clarified by centrifugation two times at 12 000 rpm for 10 min in a Sorvall GS-3 rotor at 4 °C.

Protein Purification. Ammonium sulfate to 5% saturation was added to 500 mL of the cell lysate followed by centrifugation with a Beckman JA14 rotor at 10 000 rpm for 30 min at 4 °C. The precipitate was discarded. Ammonium sulfate was added to the supernatant to 55% saturation and the solution was allowed to stir for approximately 30 min. This sample was centrifuged at 12 000 rpm for 30 min in a Beckman JA14 rotor at 4 °C and the pellet was recovered and resuspended in 20 mM potassium phosphate (pH 7.4), 1 mM EDTA, and 1 mM PMSF, followed by overnight dialysis against four liters of the same buffer. Next, the sample was centrifuged at 15 000 rpm for 30 min at 4 °C using a Beckman JA20 rotor to remove any particulate matter and then loaded onto a hydroxyapatite column (5 cm × 12.5 cm) that had been equilibrated with 20 mM potassium phosphate (pH 7.4), 1 mM EDTA, and 1 mM PMSF. Fractions of 10.0 mL in size were collected and those containing significant amounts of epimerase, as judged by SDS gels, were pooled. The protein sample was diluted 1:1 with 20 mM potassium phosphate (pH 7.5) and loaded onto a Q-sepharose column (2.5 cm × 25.0 cm) equilibrated with 20 mM potassium phosphate (pH 7.5). The column was washed with Na^+/K^+ phosphate until the flow-through reached a background absorbance level as measured at 280 nm and no protein could be observed in the SDS gels. A linear gradient from 20 to 300 mM Na^+/K^+ phosphate was applied to the column. Fractions of 4.0 mL in size were collected, and these were pooled on the basis of purity as judged by SDS gels. The pooled fractions were concentrated with an Amicon pressure filtration stirred cell and then dialyzed overnight against 10 mM potassium phosphate (pH 7.5) and 150 mM NaCl. Subsequently, the protein sample

[1] Abbreviations: EDTA, ethylenediaminetetraacetic acid; Hepps, *N*-2-hydroxyethylpiperazine-*N*′-3-propanesulfonic acid; Mes, 2-(*N*-morpholino)ethanesulfonic acid; PMSF, phenylmethanesulfonyl fluoride; SDS, sodium dodecyl sulfate; UDP-GalNAc, UDP-*N*-acetylgalactosamine; UDP-GlcNAc, UDP-*N*-acetylglucosamine.

Table 1: X-ray Data Collection Statistics

data set	resolution (Å)	independent reflections	completeness (%)	redundancy	avg I/avg $\sigma(I)$	R_{sym}[a] (%)
			Epimerase/NAD^+ Complex			
native	30.0−1.80	34 184	95.2	2.9	16.5	5.9
	1.88−1.80[b]	3761	87.0	1.8	3.0	25.8
methylHgOAc	30.0−2.00	24 826	94.3	2.5	14.6	7.5
	2.09−2.00	2516	78.3	1.5	2.9	25.1
$KAu(CN)_2$	30.0−3.00	22 555	85.7	2.2	11.8	7.3
	2.09−2.00	1926	60.0	1.4	2.9	24.0
			Epimerase/NADH/UDP-Glucose Complex			
native	30.0−1.50	96 156	96.1	5.5	29.9	5.0
	1.55−1.50	8401	84.7	3.6	2.9	28.6

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

was concentrated to ∼17 mg/mL as determined with a Bradford assay employing bovine serum albumin and the epimerase from *E. coli* as standards. An approximate extinction coefficient at 280 nm of 0.85 mL/(mg·cm) was determined. The total yield of purified epimerase was ∼400 mg.

Crystallization of the Epimerase/NAD^+ Complex. A search for suitable crystallization conditions was conducted both at room temperature and at 4 °C using a sparse matrix screen and the hanging drop method of vapor diffusion. These crystallization trials were conducted both in the presence (2 mM) and absence of NAD^+. The best crystals were observed growing from poly(ethylene glycol) 8000 and potassium chloride and in the presence of 2 mM NAD^+. Typically large crystals were obtained from 18 to 20% poly(ethylene glycol) 8000, 100 mM Hepps (pH 7.5), and 250 mM KCl at room temperature. The protein concentration was 17 mg/mL, and the sample contained 2 mM NAD^+. Crystals achieved maximum dimensions of 0.2 mm × 0.2 mm × 1.0 mm in about 1 week. These crystals belonged to the space group $C222_1$ with unit cell dimensions of $a = 50.3$ Å, $b = 116.4$ Å, and $c = 130.4$ Å. The asymmetric unit contained one monomer.

Structural Analysis of the Epimerase/NAD^+ Complex. An initial native X-ray data set was collected to 1.8 Å resolution at 0 °C with a Bruker HiStar area detector system. The X-ray source was nickel-filtered CuKα radiation from a Rigaku RU200 X-ray generator operated at 50 kV and 90 mA and equipped with a 300 μm focal cup. The X-ray data were processed with the software package SAINT (Bruker AXS, Inc.) and internally scaled with XCALIBRE (Drs. G. Wesenberg and I. Rayment, unpublished results). Relevant X-ray data collection statistics are presented in Table 1.

Two isomorphous heavy-atom derivatives were prepared by soaking native epimerase crystals in either 1 mM methyl mercury acetate for 1.5 days or 5 mM $KAu(CN)_2$ for 3 days. X-ray data sets for these derivatives were collected to 2.0 Å resolution in a manner similar to that described for the native X-ray data set. Each heavy-atom derivative data set was placed on the same scale as the native data set by a "local" scaling procedure developed by Drs. G. Wesenberg, W. Rypniewski, and I. Rayment. The *R*-factors between the native and the mercury and gold heavy-atom derivative data sets were 29.7 and 21.0%, respectively, to 2.0 Å resolution (where $R = (\sum|F_N - F_H|/\sum|F_N|) \times 100$ and $|F_N|$ is the native structure factor amplitude and $|F_H|$ is the derivative structure factor amplitude).

The positions of the heavy-atom binding sites were determined by inspection of difference Patterson maps calculated with X-ray data from 30 to 5.0 Å. Five heavy-atom binding sites for the mercury derivative and two heavy-atom binding sites for the gold derivative were identified. These heavy-atom derivatives were placed on a common origin by difference Fourier maps and the positions and occupancies for each heavy-atom binding site were refined by the origin-removed Patterson-function correlation method to 2.0 Å resolution (*25*). Protein phases were calculated with the program HEAVY (*25*). The overall figure-of-merit was 0.55 from 30 to 2.0 Å.

A model of human epimerase was built into an electron density map calculated to 2.0 Å resolution. The model was subsequently refined to 1.8 Å resolution by least-squares analysis as implemented in TNT (*26*). Relevant refinement statistics are given in Table 2. The final *R*-factor was 19.0% for all measured X-ray data from 30 to 1.8 Å resolution. The only significant outlier on the Ramachandran plot of all nonglycinyl residues was Phe 186 which adopts dihedral angles of $\phi = -97.8°$ and $\psi = -104.8°$. The electron density was unambiguous in this region. For the refinement of the human epimerase model, 10% of the X-ray data were excluded for the required calculation of R_{free} as listed in Table 2. In that all X-ray data are important for the Fourier synthesis; however, these data were ultimately included in the final stages of the refinement and model building. A portion of the electron density map corresponding to the bound cofactor is shown in Figure 1. The average temperature factors for the waters, the polypeptide chain, and the NAD^+ were 51.2, 34.7, and 19.4 $Å^2$, respectively.

Crystallization of the Epimerase/NADH/UDP-Glucose Ternary Complex. The abortive complex was prepared by treating the epimerase (17 mg/mL in the final dialysis buffer) with 10 mM NADH and 10 mM UDP-glucose and allowing the solution to stand for 24 h at 4 °C. The sample was subsequently frozen in 100 μL aliquots. A sparse matrix screen, conducted at both room temperature and at 4 °C, yielded several crystal forms. The form initially demonstrating the best X-ray diffraction properties grew in the presence of poly(ethylene glycol) 3400 and $MgCl_2$ at 4 °C. Large crystals employed for the X-ray analysis described here were prepared by macroseeding in batch setups. Specifically, the abortive complex, after thawing, was diluted 1:1 with 100 mM Mes (pH 6.0). This solution was then mixed 1:1 with a precipitant solution such that the final mother liquor contained 8.5% poly(ethylene glycol) 3400 and 75 mM $MgCl_2$.

Table 2: Relevant Least-Squares Refinement Statistics

model	epimerase/NAD$^+$ complex	epimerase/NADH/ UDP-glucose complex
resolution limits (Å)	30.0−1.80	30.0−1.50
R-factor[a] (overall) (%)/no. of rflns	19.1/34 153	16.9/96 012
R-factor (working) (%)/no. of rflns	19.0/30 738	16.8/86 401
R-factor (free) (%)/no. of rflns	22.3/3415	19.8/9611
no. of protein atoms	2676	5548[b]
no. of heteroatoms	240[c]	1033[d]
weighted root-mean-square deviations from ideality		
bond lengths (Å)	0.011	0.010
bond angles (deg)	2.36	2.04
trigonal planes (Å)	0.006	0.007
general planes (Å)	0.010	0.011
torsional angles (deg)	16.6	15.7

[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] These include multiple conformations for E61, M65, M83, E96, T110, M167, K176, and D231 in subunit I and E3, E63, K78, T110, K120, T134, M167, E221, and C307 in subunit II. [c] These include 196 water molecules and one NAD$^+$ molecule. [d] These include two NADH molecules, two UDP-glucose molecules, 856 waters, two magnesium ions, and three tetramethylammonium ions.

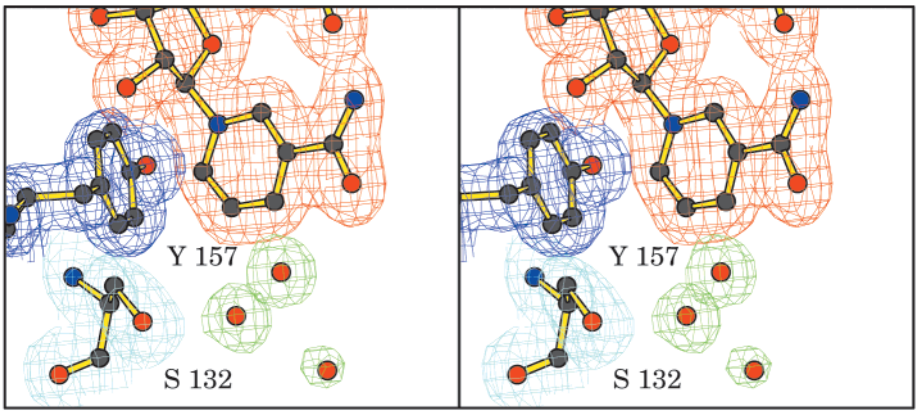


FIGURE 1: Representative electron density for the human epimerase/NAD$^+$ complex. The map was contoured at 1σ and calculated with coefficients of the form $(2F_o - F_c)$, where F_o was the native structure factor amplitude and F_c was the calculated structure factor amplitude. The nicotinamide ring of the cofactor is displayed in red contours. Ordered solvent molecules are indicated by the green contours.

Typically, the crystals achieved maximum dimensions of 0.10 mm × 0.15 mm × 0.80 mm in approximately 1−2 weeks.

Structural Analysis of the Epimerase/NADH/UDP-Glucose Complex. For X-ray data collection, the crystals were first transferred into a synthetic mother liquor containing 15% poly(ethylene glycol) 3400, 75 mM $MgCl_2$, 2.5 mM NADH, 2.5 mM UDP-glucose, and 100 mM Mes (pH 6.0). After 1 h of equilibration, the crystals were transferred (for 10−15 s) to a cryoprotectant solution containing 20% poly(ethylene glycol) 3400, 75 mM $MgCl_2$, 500 mM NaCl, 1 M tetramethylammonium chloride, 2.5 mM NADH, 2.5 mM UDP-glucose, and 100 mM Mes (pH 6.0). These crystals were suspended in a loop of 20 μm surgical thread and immediately flash-frozen in a stream of nitrogen gas.

The crystals belonged to the space group $P2_12_12_1$ with unit cell dimensions of $a = 74.7$ Å, $b = 86.9$ Å, and $c = 95.3$ Å. The asymmetric unit contained one dimer. A native X-ray data set to 1.5 Å resolution was collected at the Advanced Photon Source, Structural Biology Center beamline 19-ID. These data were processed with HKL2000 and scaled with SCALEPACK (*27*). The structure was solved via molecular replacement with the program AMORE (*28*) employing as the search model the structure of the human epimerase dimer with bound NAD$^+$. Initial refinement with the package TNT reduced the *R*-factor to ∼36% at 1.5 Å resolution. The electron density map was subsequently improved by molecular averaging with the software AVE found in the RAVE package of programs (*29*, *30*). On the basis of this averaged electron density, the model was adjusted and placed back into the unit cell, and another cycle of least-squares refinement applied. The *R*-factor dropped to ∼24%. Next the model was adjusted within the unit cell and solvent molecules were added to the X-ray coordinate file. The final *R*-factor was 16.9% for all measured X-ray data from 30 to 1.5 Å. Relevant refinement statistics can be found in Table 2. The electron density map near the bound NADH and UDP-glucose molecules is shown in Figure 2. As in the human epimerase/NAD$^+$ complex, the only significant outlier in the Ramachandran plot was Phe 186 in both subunits. The average temperature factors for the waters, the polypeptide chain, the NADH, and the UDP-glucose were 34.5, 19.2, 11.4, and 18.3 Å^2, respectively

RESULTS AND DISCUSSION

Molecular Structure of the Epimerase/NAD$^+$ Complex. Mammalian UDP-galactose 4-epimerases are known to be dimers in solution (*31*). A ribbon representation of one subunit of the human epimerase dimer is shown in Figure 3. As color-coded in blue and green, respectively, the

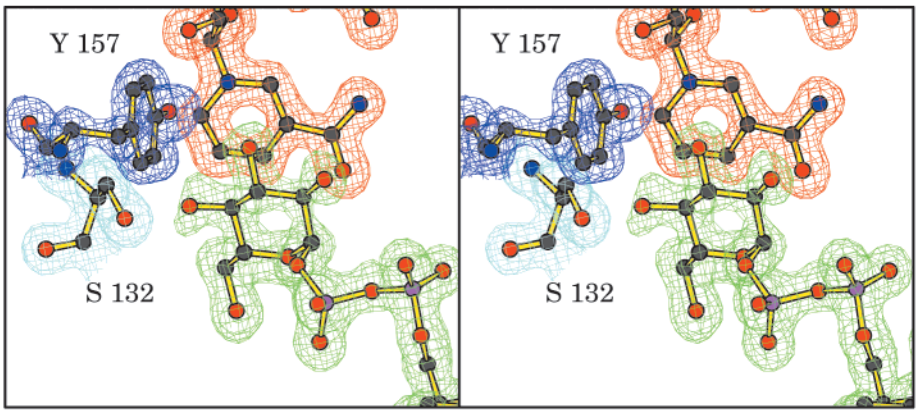


FIGURE 2: Representative electron density for the human epimerase/NADH/UDP-glucose ternary complex. The map was calculated and contoured in the same manner as described in Figure 1. The nicotinamide ring of the dinucleotide and the glucosyl moiety of the substrate are displayed in red and green, respectively.

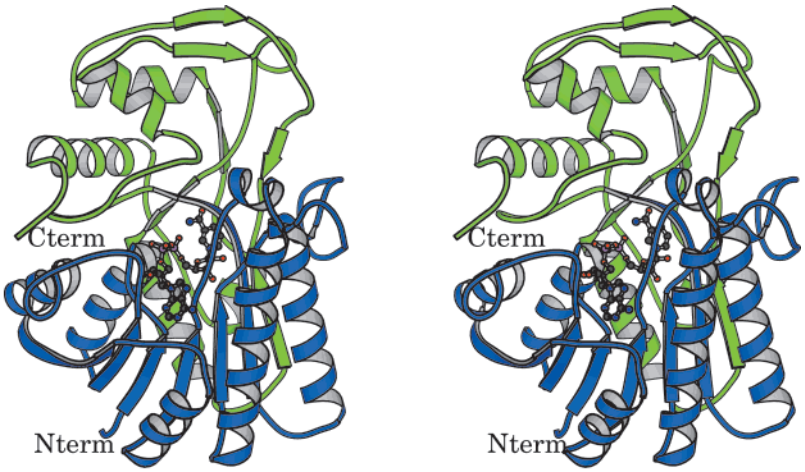


FIGURE 3: Ribbon representation of one subunit of human epimerase. The N-terminal and C-terminal domains are color-coded in blue and green, respectively. The NAD^+ molecule is drawn in a ball-and-stick representation.

molecule folds into two structural domains delineated by Met 1 to Thr 189 and Gly 190 to Ala 346. The N-terminal motif is dominated by a seven-stranded parallel β-sheet with the β-strands ranging in length from 4 to 11 amino acid residues. Five α-helices, two on one side and three on the other, abut this β-sheet. These helices contain anywhere from 6 to 19 amino acid residues. This domain is responsible for positioning the NAD^+ at the C-terminal end of the β-sheet as indicated in Figure 3. In addition to these α-helices and β-strands, there is another short α-helix from Val 94 to Gln 99 and eight classical reverse turns (six type I, one type II, and one type III).

The C-terminal motif, or substrate binding domain, is composed of six strands of β-sheet ranging in length from three to six amino acid residues. The third β-strand of the C-terminal domain, delineated by Cys 264 to Asn 268, forms the seventh strand of the parallel β-sheet located in the N-terminal domain. Additionally, the C-terminal domain contains five α-helices, ranging in length from 6 to 14 amino acid residues, one type III turn and four type I turns.

A close-up view of the NAD^+-binding pocket is shown in Figure 4a. There are 14 water molecules located within approximately 4.0 Å of the dinucleotide cofactor, eight of which participate directly in hydrogen bonds to the NAD^+. The nicotinamide ring of the cofactor adopts the *syn*-conformation with an intramolecular hydrogen bond occurring between the nitrogen of its carboxamide group and a phosphoryl oxygen. This type of conformation was also observed for the *E. coli* epimerase which is known to be a *B*-side specific enzyme (*12*). Both riboses of the NAD^+ adopt the $C_{2'}$-endo conformation. Note that Phe 186, whose dihedral angles lie outside of the allowed regions in the Ramachandran plot, is located at the top of the binding pocket in Figure 4a.

A cartoon of the hydrogen-bonding pattern observed between the oxidized cofactor and the protein is displayed in Figure 4b. Five side chains, namely those provided by Asp 33, Asn 37, Asp 66, Tyr 157, and Lys 161, are directly involved in nucleotide binding. Additionally, two peptidic NH groups contributed by Tyr 13 and Ile 14 form hydrogen bonds with the phosphoryl oxygens of the dinucleotide. There are 19 hydrogen bonds formed between NAD^+ and the protein or solvent. Strikingly, five of the eight water molecules shown in Figure 4b are located in nearly identical positions to those observed in the NAD^+-binding pocket for the *E. coli* epimerase (*12*). In the *E. coli* enzyme, removal of the dinucleotide cofactor results in irreversible denaturation of the enzyme, while in the human protein, the NAD^+ can be readily removed. Not surprisingly, a comparison of the two NAD^+-binding pockets for the human and bacterial enzymes reveals significantly more interactions between the

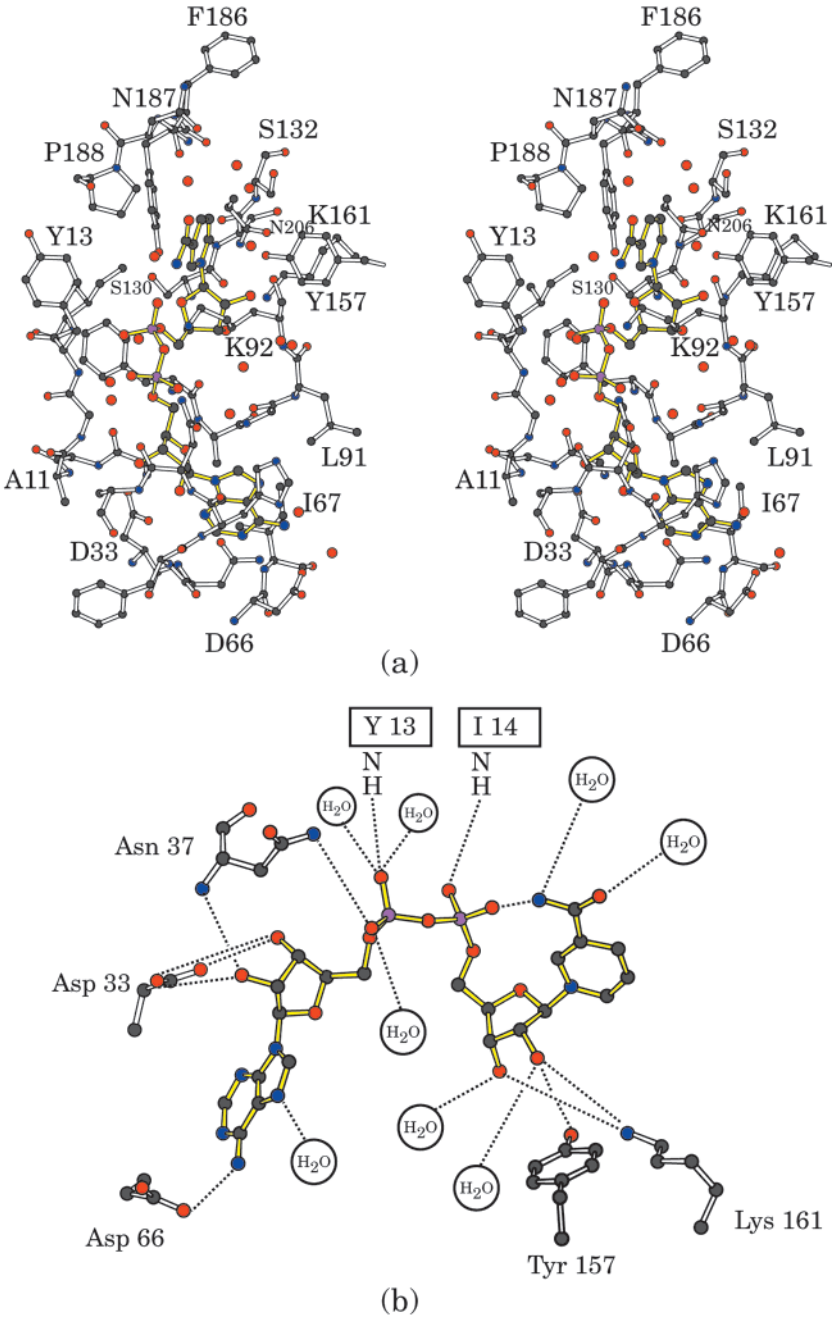


FIGURE 4: Close-up view of the human epimerase dinucleotide cofactor binding pocket. Those residues that lie within approximately 4.0 Å of atoms of the NAD^+ are displayed in panel a. A cartoon of potential hydrogen-bonding interactions between the dinucleotide cofactor and the protein is shown in panel b. The dashed lines represent potential hydrogen bonds.

protein and the dinucleotide in the *E. coli* protein. Specifically, in the *E. coli* epimerase, there are 19 hydrogen bonds between the dinucleotide and the protein as opposed to eleven in the human enzyme.

Perhaps the best characterized epimerase to date is that from *E. coli* which has been the subject of numerous X-ray crystallographic analyses (*11*−*15*). Previously, it was postulated that substrate binding to epimerase leads to a large conformational change in the enzyme which activates the cofactor toward reduction (*32*, *33*). In every structure solved thus far for the bacterial enzyme, however, there has always been either UDP or some type of UDP-sugar analogue bound in the substrate-binding pocket. To address the conformational changes that occur upon substrate binding, it is necessary to solve the structure of the enzyme in its resting state. All attempts to produce crystals of the epimerase from *E. coli* that are amenable for such a study have thus far been unsuccessful. Consequently, the structure of the human enzyme described here is particularly informative in that it represents the first glimpse of an epimerase in its resting state, i.e., without bound UDP-sugar or UDP.

A superposition of the α-carbon traces for the models of the *E. coli* epimerase and the human protein/NAD^+ complex is displayed in Figure 5. The two molecules superimpose with a root-mean-square deviation of 0.98 Å for 299 structurally equivalent α-carbon atoms. Note that the X-ray coordinates for the *E. coli* epimerase employed for this comparison were those from the previously reported enzyme/NAD^+/UDP complex (*12*). As can be seen, there is a six-residue insertion in the human enzyme relative to the

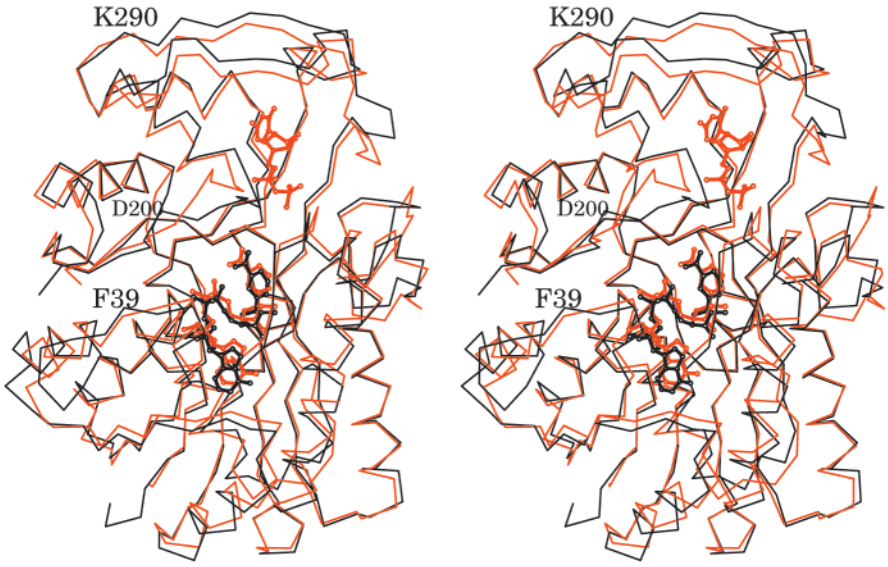


FIGURE 5: Superposition of the polypeptide chain backbones for the epimerases from human and *E. coli*. The human epimerase polypeptide chain is displayed in black while the bacterial enzyme is shown in red. Note that the model for the bacterial enzyme contained, in addition to NAD^+, UDP as indicated by the ball-and-stick representation.

bacterial protein starting at Phe 39. The biological role of this insertion is not immediately obvious. What is clear, however, is that the C-terminal motifs of the bacterial and human enzymes are in slightly different orientations with respect to their N-terminal domains. As would be predicted, the C-terminal domain of the *E. coli* enzyme, which contains bound UDP, is closed down relative to that observed for the C-terminal domain of the human resting enzyme. From this superposition, it can be postulated that the major conformational changes that occur upon UDP or UDP-sugar binding to epimerase are confined primarily within the regions defined by Asp 200 to Asp 231 and Lys 290 to Val 304 (human enzyme numbering).

Molecular Structure of the Epimerase/NADH/UDP-Glucose Ternary Complex. A superposition of the polypeptide chains for the epimerase/NAD^+ and the epimerase/NADH/UDP-glucose (subunit II of the asymmetric unit) complexes is displayed in Figure 6a. As can be seen, the net effect of UDP-glucose binding to human epimerase is the clamping down of the C-domain over the active-site region. Indeed, the α-carbons for the two polypeptide chains superimpose with a remarkably high root-mean-square deviation of 2.5 Å. Again, as discussed above, the major conformational changes are confined to the regions defined by Glu 199 to Asp 240 and Gly 274 to Tyr 308. If these regions are omitted from the calculation, however, the two polypeptide chains correspond with a root-mean-square deviation of 0.92 Å.

Crystals of the epimerase/NADH/UDP-glucose ternary complex employed for this structural analysis contained a complete dimer in the asymmetric unit. A superposition of the α-carbon traces for subunits I and II in the asymmetric unit is displayed in Figure 6b. It was anticipated that the two subunits would be nearly identical in their three-dimensional folds. As can be seen, however, there are several very significant differences between the two monomers contained within the asymmetric unit. Indeed, the α-carbons for these two polypeptide chains superimpose with a root-mean-square deviation of 1.0 Å. Again, the differences between the two monomers in the dimer are limited primarily to those regions defined by Glu 199 to Asp 240 and Gly 274 to Tyr 308.

Shown in Figure 7 is a close-up view of the active site for subunit II. As can be seen, Phe 226 forms a striking stacking interaction with the uracil ring of the substrate. In addition, the uracil ring is linked to the protein via the backbone amide group of Phe 226 and the carbonyl group of Asn 224. The 2′-hydroxyl group of the uridine ribose lies with hydrogen-bonding distance to the carboxylate group of Asp 303. Asn 187, Asn 207, Arg 239, and Arg 300 form electrostatic interactions with the phosphoryl oxygens of the UDP-glucose. With respect to the glucosyl moiety, four amino acid side chains provide key electrostatic interactions to the sugar hydroxyl groups: Ser 132, Tyr 157, Asn 187, and Asn 207.

What are the structural consequences resulting from the movement of the polypeptide chain in subunit II relative to subunit I? As can be seen in Figure 8, the hydrogen-bonding pattern around the 4′-hydroxyl group of the glucosyl moiety has been modified. Recall that, during the reaction mechanism, the hydride from C4 of the sugar is transferred to C4 of the nicotinamide ring. In subunit I, C4 of the sugar lies within 3.6 Å of C4 of the nicotinamide ring. This distance is comparable to that observed in subunit II (3.5 Å). Likewise O^{γ} of Ser 132 is positioned at 2.4 Å from the 4′-hydroxyl group of the sugar in subunit I and is located at 2.5 Å from the sugar hydroxyl group in subunit II.

The significant changes occur, however, at Tyr 157. In subunit I, O^{η} of Tyr 157 is hydrogen-bonded to the 2′-hydroxyl group of the nicotinamide ribose. The distance between these two atoms is 2.7 Å. With respect to the glucosyl moiety, O^{η} of Tyr 157 is positioned at 3.2 and 4.1 Å from the 3′-hydroxyl and 4′-hydroxyl groups, respectively. These distances are comparable to those observed for the *E. coli* abortive complex model (3.1 and 4.3 Å, respectively). In subunit II of human epimerase, where the C-terminal domain has clamped down more tightly, O^{η} of Tyr 157 now lies within hydrogen-bonding distance not only to the

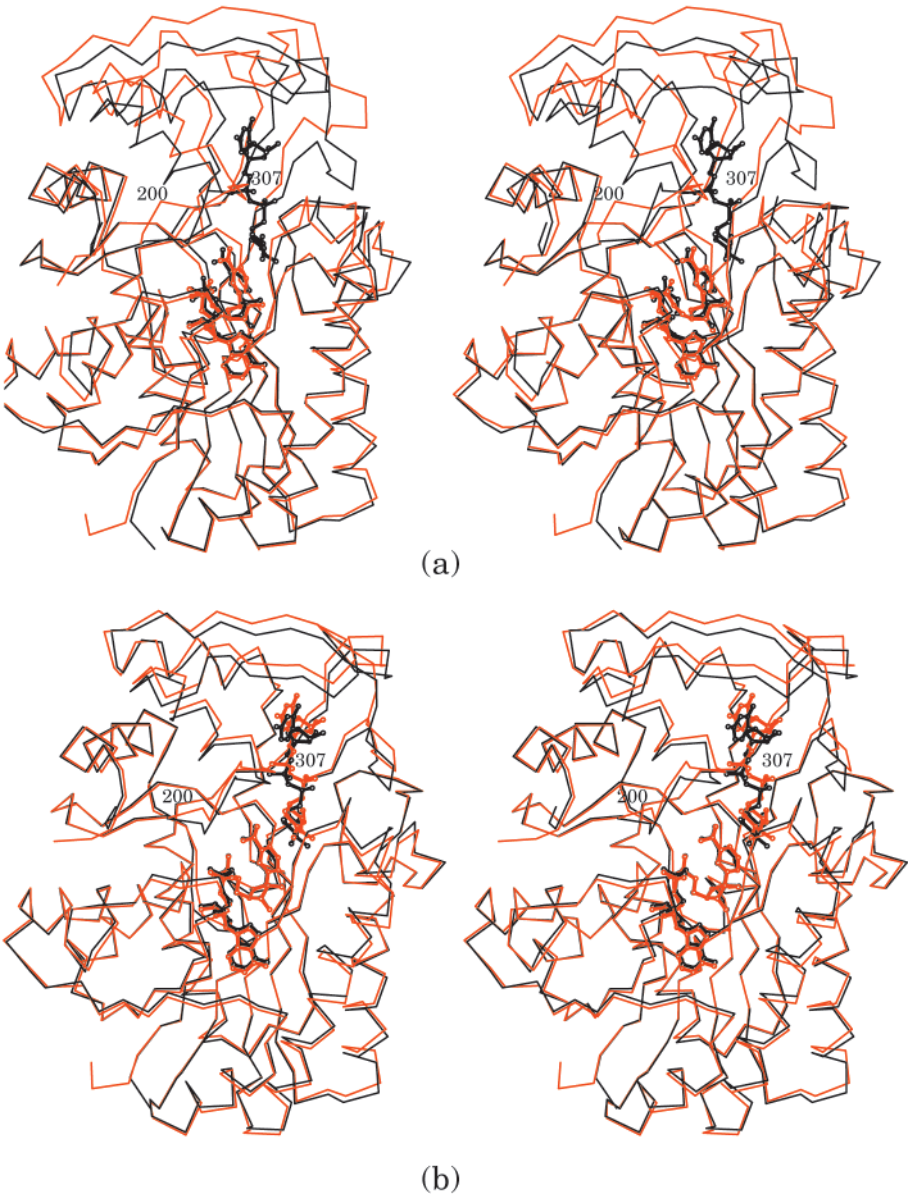


FIGURE 6: Comparison of the open and closed forms of human epimerase. Shown in panel a is a superposition of the models for the resting enzyme, depicted in red, and for the enzyme in a ternary complex (subunit II of the asymmetric unit), displayed in black. Note the movement of the C-terminal domain upon substrate binding. Shown in panel b is a superposition of the two subunits contained within the asymmetric unit of the epimerase/NADH/UDP-glucose abortive complex. Subunit I is displayed in red while subunit II is drawn in black.

nicotinamide ribose 2′-hydroxyl group (2.8 Å) and the 3′-hydroxyl group of the glucosyl moiety (2.9 Å), but also to the 4′-hydroxyl group of the UDP-sugar (3.1 Å). Note that it is the 4′-hydroxyl hydrogen that is presumably removed by an active-site base during catalysis. Not surprisingly, the average temperature factor for the UDP-glucose in subunit I is 22.3 Å^2, while in subunit II it is 16.7 Å^2.

Until this investigation, all X-ray crystallographic models of the *E. coli* epimerase indicated that O$^\eta$ of Tyr 149 (Tyr 157 in the human enzyme) was invariably located at approximately 4.3 Å from the 4′-hydroxyl group of the glucosyl substrate. This distance seemed to preclude the role of Tyr 149 interacting directly with the 4′-hydroxyl group of the glucosyl moiety as the active-site base. Accordingly, a proposal was put forth whereby Ser 124 (Ser 132 in the human enzyme), which is located at 2.6 Å from the 4′-hydroxyl group of the glucose in the *E. coli* epimerase, acts as a proton shuttle between the sugar and the presumed phenolic side chain of Tyr 149 (*11*). Indeed, there was nothing unusual in the active site to suggest that the side-chain hydroxyl group of Ser 124 would be ionized, and hence, the proton shuttle hypothesis was a reasonable assumption given the available structural evidence at the time. The one caveat of this hypothesis, however, was that the observed distance between the side chain hydroxyls of Ser 124 and Tyr 149 was quite long at 4.6 Å. It was argued, however, that this distance was based on dead-end complexes and might be significantly different in the epimerase/NAD$^+$/UDP-glucose complex.

The results described here for the human epimerase are especially revealing with regard to the catalytic mechanism of the enzyme. From subunit II, it is now known that the C-terminal domain moves down upon the active site more than originally envisioned or observed and this three-dimensional shift results in O$^\eta$ of Tyr 157 lying within hydrogen-bonding distance to the 4′-hydroxyl group of the

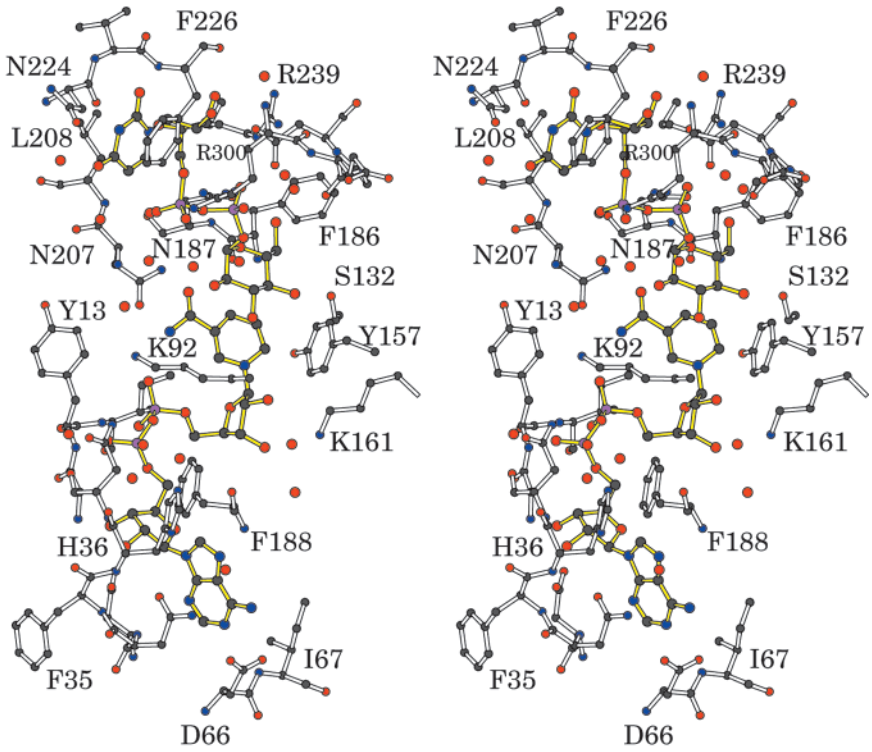


FIGURE 7: Close-up view of the human epimerase active site. Those residues that are located within approximately 3.0 Å of atoms of the NADH and UDP-glucose are displayed.

glucosyl substrate. Additionally, its side-chain hydroxyl group is positioned within 3.7 Å of the nicotinamide nitrogen that would carry a positive charge if the cofactor was in the oxidized state. This type of interaction could clearly aid in lowering the pK_a of the side chain of Tyr 157 to allow it to function directly as the active-site base. Indeed, O^{η} of Tyr 157 is still positioned at 4.1 Å from the hydroxyl side chain of Ser 132, indicating that, at least in the case of the human epimerase, the proton shuttle mechanism most likely does not occur.

What is the role of Ser 132 in human epimerase? The short distance between O^{γ} of Ser 132 and the 4′-hydroxyl group of the UDP-glucose substrate (2.5 Å), even in light of X-ray coordinate errors, suggests the formation of a low-barrier hydrogen bond (*34*−*36*) whereby the pK_{α} values of both alcoholic groups are similar leading to a proton being shared between the two oxygen atoms. Indeed, this close interaction would facilitate removal of the 4′-hydroxyl hydrogen of the sugar by the phenolic side chain of Tyr 157 and the transfer of the hydride from C4 of the sugar to C4 of the dinucleotide cofactor. The energy released by forming the low-barrier hydrogen bond would permit transfer of the proton from the 4′-hydroxyl group of the sugar to Tyr 157, which would normally be a thermodynamically unfavorable event. A similar type of scenario has been recently described for horse liver alcohol dehydrogenase whereby a zinc-bound alkoxide forms a low-barrier hydrogen bond with the hydroxyl group of a serine residue (*37*). Site-directed mutagensis experiments are underway to further test the roles of Ser 132 and Tyr 157 in the catalytic mechanism of the human epimerase.

The present study has revealed both the conformational changes that occur upon substrate binding to human epimerase and the identity of the active-site base, namely Tyr 157. Perhaps more importantly, however, these investigations reveal the necessity for studying not just one member of a superfamily but rather a series of proteins. In today's scientific climate, it could have been argued that since the epimerases from human and *E. coli* are 55% identical, there was no compelling reason or need to study the human enzyme, given the previous extensive structural work on the bacterial protein. The investigation of the human enzyme, however, revealed important molecular features that were elusive in the X-ray crystallographic studies of the *E. coli* epimerase. Contrary to what is being purported by those involved in structural genomics, knowledge of the three-dimensional architecture for only one, or even a few, members of a superfamily may not be sufficient to fully define either the structures or functions of other members.

How successful were the previous homology-modeling studies of the human epimerase sequence onto the *E. coli* protein? As described in the introductory portion of this paper, one of the first patient hGALE alleles to be identified encoded an N34S substitution (*4*). On the basis of the *E. coli* structure, it was assumed that Asn 34 in the human enzyme formed a hydrogen bond with the adenine ring of the bound NAD^+. In fact, while this residue abuts the adenine ring, no such hydrogen bond occurs. Rather the carboxamide functional group of Asn 34 participates in an electrostatic interaction with the carboxylate group of Asp 66, which in turn, interacts with the adenine ring of the NAD^+. Again, this points to the danger of a scientific philosophy whereby one molecular fold defines an entire family.

While the model described here answers several interesting structural/functional questions, many remain. For example, how are UDP−GalNAc and UDP-GlcNAc accommodated within the active site of the human enzyme? Why does the protein from *E. coli* not interconvert UDP−GalNAc and UDP-GlcNAc? How is UDP-galactose bound in the active

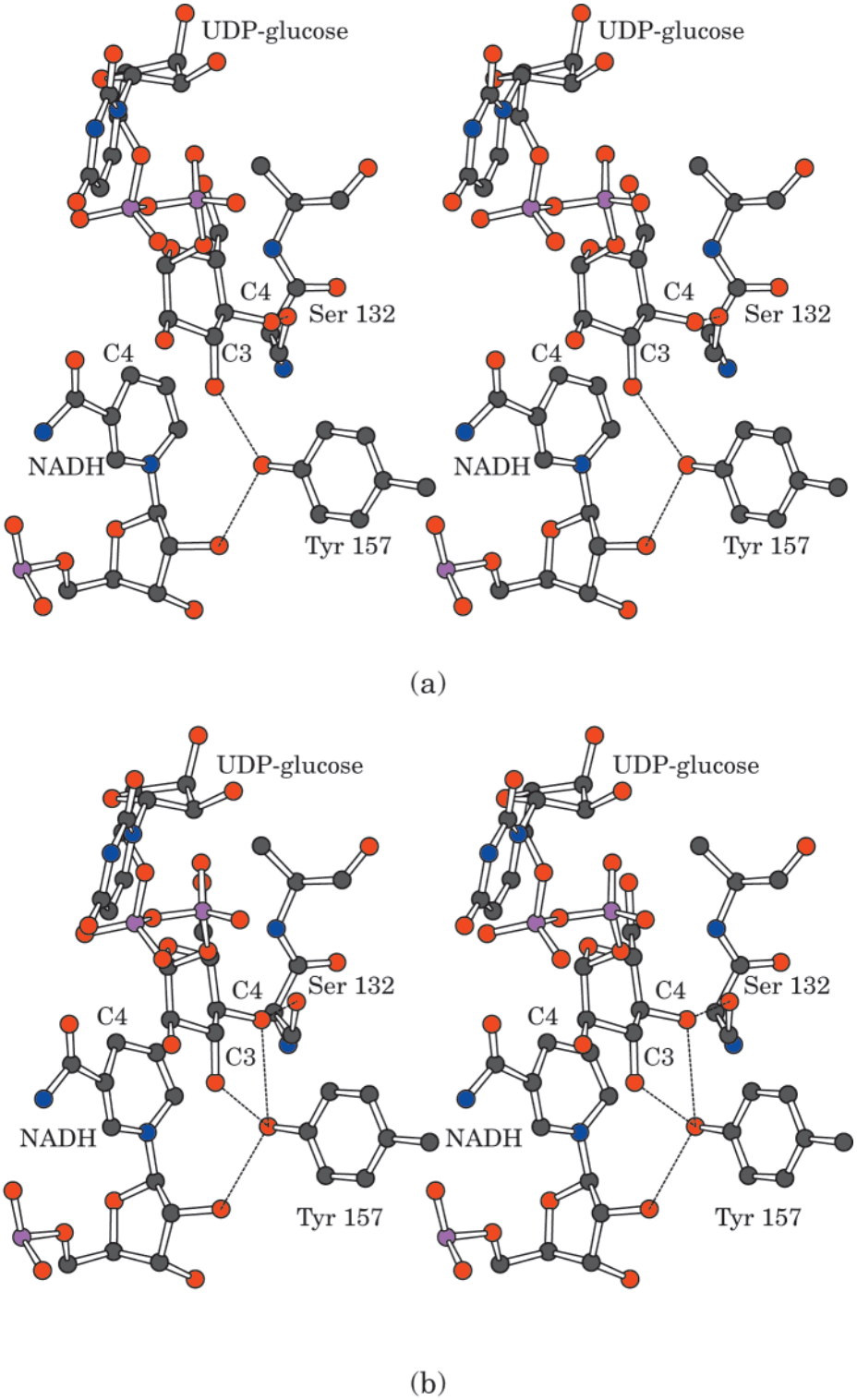


FIGURE 8: Differences in the hydrogen-bonding patterns around the 4′-hydroxyl groups of the glucosyl moieties in subunits I and II. The hydrogen-bonding patterns for subunits I and II are displayed in panels a and b, respectively.

site of the human enzyme? Both X-ray crystallographic studies and site-directed mutagenesis experiments designed to address these and other issues are presently underway.

ACKNOWLEDGMENT

We gratefully acknowledge the contribution of Dr. W. W. Cleland throughout the course of this investigation. We also thank Dr. Dale Edmondson and Ms. Paige Newton-Vinson for generously allowing and helping us to use their fermenter. Use of the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Energy Research, under Contract W-31-109-ENG-38.

REFERENCES

1. Alano, A., Almashanu, S., Maceratesi, P., Reichardt, J., Panny, S., and Cowan, T. M. (1997) *J. Invest. Med. 45*, 191A.
2. Gitzelmann, R., Steinmann B., Mitchell, B., and Haigis E., (1977) *Helv. Paediatr. Acta 31*, 441−452.

3. Alano, A., Almashanu, S., Chinsky, J. M., Costeas, P., Blitzer, M. G., Wulfsberg, E. A., and Cowan T. M. (1998) *J. Inherited Metab. Dis. 21*, 341−350.
4. Quimby, B. B., Alano, A., Almashanu, S., DeSandro, A. M., Cowan, T. M., and Fridovich-Keil, J. L. (1997) *Am. J. Hum. Genet. 61*, 590−598.
5. Segal, S., and Berry, G., (1995) in *The Metabolic and Molecular Bases of Inherited Disease* (Scriver, C., Beaudet, A., Sly, W., Valle, D., Eds.) pp 967−1000, McGraw-Hill, Inc., New York.
6. Shin, Y. S., Korenke, G.-C., Huppke, P., Knerr, I., and Podskarbi, T. (1999) *J. Inherited Metab. Dis. 22* (Suppl. 1), 8.
7. Walter, J. H., Roberts, R. E., Besley, G. T., Wraith, J. E., Cleary, M. A., Holton, J. B., and MacFaul, R., (1999) *Arch. Dis. Child. 80*, 374−376.
8. Wohlers, T. M., Christacos, N. C., Harreman, M. T., and Fridovich-Keil, J. L. (1999) *Am. J. Hum. Genet. 64*, 462−470.
9. Maceratesi, P., Dallapiccola, B., Novelli, G., Okano, Y., Isshiki, G., andReichardt, J. K. V. (1996) *Am. J. Hum. Genet. 59* (Suppl.), A204.
10. Maceratesi, P., Daude, N., Dallapiccola, B., Novelli, G., Allen, R., Okano, Y., and Reichardt, J. (1998) *Mol. Genet. Metab. 63*, 26−30.
11. Thoden, J. B., Frey, P. A., and Holden, H. M. (1996) *Protein Sci. 5*, 2149−2161.
12. Thoden, J. B., Frey, P. A., and Holden, H. M. (1996) *Biochemistry 35*, 2557−2566.
13. Thoden, J. B., Frey, P. A., and Holden, H. M. (1996) *Biochemistry 35*, 5137−5144.
14. Thoden, J. B., Gulick, A. M., and Holden, H. M. (1997) *Biochemistry 36*, 10685−10695.
15. Thoden, J. B., and Holden, H. M. (1998) *Biochemistry 37*, 11469−11477.
16. Wilson, D. B., and Hogness, D. S. (1969) *J. Biol. Chem. 244*, 2132−2136.
17. Lemaire, H. G., and Müller-Hill, B. (1986) *Nucleic Acids Res. 14*, 7705−7711.
18. Liu, Y., Thoden, J. B., Kim, J., Berger, E., Gulick, A. M., Ruzicka, F. J., Holden, H. M., and Frey, P. A. (1997) *Biochemistry 36*, 10675−10684.
19. Baker, M. E., and Blasco, R. (1992) *FEBS Lett. 301*, 89−93.
20. Holm, L., Sander, C., & Murzin, A. (1994) *Nat. Struct. Biol. 1*, 146−147.
21. Duax, W. L., Griffin, J. F., and Ghosh, D. (1996) *Curr. Opin. Struct. Biol. 6*, 813−823.
22. Maley, F., and Maley, G. F. (1959) *Biochim. Biophys. Acta 31*, 577−578.
23. Piller, F., Hanlon, M. H., and Hill, R. L. (1983) *J. Biol. Chem. 258*, 10774−10778.
24. Kingsley, D., Kozarsky, K. F., Hobbie, L., and Krieger, M. (1986) *Cell 44*, 749−759.
25. Terwilliger, T. C., and Eisenberg, D. (1983) *Acta Crystallogr., Sect. A 39*, 813−817.
26. Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987) *Acta Crystallogr., Sect. A. 43*, 489−501.
27. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
28. Navaza, J. (1994) *Acta Crystallogr., Sect. A 50*, 157−163.
29. Jones, T. A. (1992) in *Molecular Replacement* (Dodson, E. J., Gover, S., and Wolf, W., Eds.) pp 91−105, SERC Daresbury Laboratory, Warrington.
30. Kleywegt, G. J., and Jones, T. A. (1994) in *From First Map to Final Model* (Bailey, S., Hubbard, R., and Waller, D., Eds.) pp 56−66, SERC Daresbury Laboratory, Warrington.
31. Langer, R., and Glaser, L. (1974) *J. Biol. Chem. 249*, 1126−1132.
32. Kang, U. G., Nolan, L. D., and Frey, P. A. (1975) *J. Biol. Chem. 250*, 7099−7105.
33. Bertland, A. U., II, and Kalckar, H. M. (1968) *Proc. Natl. Acad. Sci. U.S.A. 61*, 629−635.
34. Cleland, W. W., and Kreevoy, M. M. (1994) *Science 264*, 1887−1890.
35. Gerlt, J. A., and Gassman, P. G. (1993) *Biochemistry 32*, 11943−11952.
36. Gerlt, J. A., and Gassman, P. G. (1993) *J. Am. Chem. Soc. 115*, 11552−11568.
37. Ramaswamy, S., Park, D.-H., and Plapp, B. V. (1999) *Biochemistry 38*, 13951−13959.

BI000215L