A Structural View of the Action of *Escherichia coli* (*lac*Z) β-Galactosidase[†,‡]

Douglas H. Juers,[§] Tom D. Heightman,[||,⊥] Andrea Vasella,[||] John D. McCarter,[#,○] Lloyd Mackenzie,[#,△] Stephen G. Withers,[#] and Brian W. Matthews*[,§]

Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, University of Oregon, Eugene, Oregon 97403-1229, Laboratorium für Organische Chemie, ETH-Zentrum, Universitätstrasse 16, CH-8092 Zürich, Switzerland, and Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1

Received August 27, 2001; Revised Manuscript Received October 9, 2001

ABSTRACT: The structures of a series of complexes designed to mimic intermediates along the reaction coordinate for β-galactosidase are presented. These complexes clarify and enhance previous proposals regarding the catalytic mechanism. The nucleophile, Glu537, is seen to covalently bind to the galactosyl moiety. Of the two potential acids, Mg^{2+} and Glu461, the latter is in better position to directly assist in leaving group departure, suggesting that the metal ion acts in a secondary role. A sodium ion plays a part in substrate binding by directly ligating the galactosyl 6-hydroxyl. The proposed reaction coordinate involves the movement of the galactosyl moiety deep into the active site pocket. For those ligands that do bind deeply there is an associated conformational change in which residues within loop 794−804 move up to 10 Å closer to the site of binding. In some cases this can be inhibited by the binding of additional ligands. The resulting restricted access to the intermediate helps to explain why allolactose, the natural inducer for the *lac* operon, is the preferred product of transglycosylation.

β-Galactosidase, the product of the Z gene of the *lac* operon of *Escherichia coli*, has a rich history in molecular biology and biochemistry (*1*−*3*). It was through the study of the appearance of this enzyme as an adaptive response to a new food source that Jacob and Monod formulated their operon model for gene expression (*4*). Subsequent investigation of β-galactosidase has been extensive. It has become an invaluable tool for molecular biology, and detailed proposals have been made for its catalytic mechanism (*5*−*13*).

Formally, the role of β-galactosidase in *E. coli* is to hydrolyze the disaccharide lactose to galactose and glucose as well as to convert lactose to another disaccharide, allolactose, which is the natural inducer for the *lac* operon. Operationally, it has fairly strict specificity at the galactosyl position but is adept at hydrolyzing β-D-galactopyranosides with a wide variety of aglycons. Such promiscuity is part of the reason it has been so useful. In particular, it allows the use of simple, colorimetric assays based on substrates such as X-gal[1] and ONPG which have chromogenic aglycons.

β-Galactosidase is a retaining glycosidase, catalyzing hydrolysis of its substrates in a double displacement reaction where the product retains the same stereochemistry as the starting state (Figure 1a). The two-step nature of such a reaction was proposed by Koshland (*14*) and was later demonstrated experimentally in β-galactosidase using methanol competition (*15*) and also, under certain conditions, by a burst of aglycon in the pre steady state (*8*).

In addition to the participation of the nucleophile, acid/base catalysis is also important. Structure/reactivity relationships suggest that there is at least partial acid assistance of leaving group departure, especially with more basic leaving groups such as glucose (*16*−*18*). A magnesium ion appears to be necessary for such acid assistance, although the details are somewhat controversial (*12*, *17*, *19*−*22*). All of the kinetic data are consistent with either Lewis catalysis by the magnesium or Bronsted catalysis by Glu461. In the former case, cleavage of the glycosidic bond is facilitated by a direct electrophilic attack on the glycosidic oxygen by the magnesium ion leading to an Mg−OR complex. Release of this complex via protonation of the leaving group is suggested to be rate-determining for the first step, masking prior events, including transfer of the galactosyl group to the enzyme. In the latter case, glycosidic bond cleavage is itself rate-limiting and is facilitated by proton donation to the glycosidic oxygen

[†] This work was supported in part by NIH Grant GM20066.

[‡] The coordinates and structure factors have been deposited with the Protein Data Bank. Access codes are given in Table 2.

* To whom correspondence should be addressed: Telephone: (541) 346-2572. Facsimile: (541) 346-5870. E-mail: brian@uoxray.uoregon.edu.

[§] University of Oregon.

[||] ETH Zürich.

[⊥] Current address: GlaxoSmithKline Pharmaceuticals, New Frontiers Science Park (N), Third Ave., Harlow, Essex CM195AW, U.K.

[#] University of British Columbia.

[○] Current address: Mail Stop 29-2-MA, AMGEN Inc., One Amgen Center Drive, Thousand Oaks, CA 91320-1799.

[△] Current address: Suite 425, 5600 Parkwood Way, Richmond, British Columbia, Canada V6V 2M2.

[1] Abbreviations: X-gal, 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside; ONPG, *o*-nitrophenyl β-D-galactopyranoside; PNPG, *p*-nitrophenyl β-D-galactopyranoside; IPTG, isopropyl β-D-thiogalactopyranoside; 2-F-gal-E, 2-fluoro-α-D-galactosyl enzyme; 2-d-gal-E, 2-deoxy-α-D-galactosyl enzyme; galactal, 1,5-anhydro-2-deoxy-D-*lyxo*-hex-1-enitol; PEG, poly(ethylene glycol); DMSO, dimethyl sulfoxide; DTT, dithiothreitol; lactose, galactopyranosyl-1−4-β-D-glucopyranose; allolactose, galactopyranosyl-1−6-β-D-glucopyranose; SSRL, Stanford Synchrotron Radiation Laboratory; ALS, Advanced Light Source.

10.1021/bi011727i CCC: $20.00
Published on Web 11/15/2001

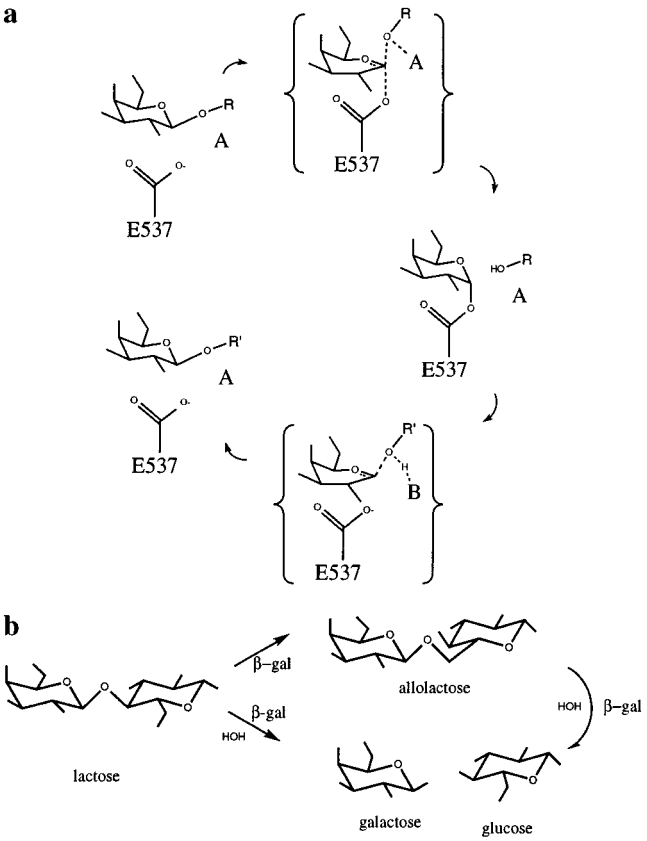

FIGURE 1: (a) Generalized outline for a double-displacement reaction catalyzed by β-galactosidase. In the first step (top), the substrate, a β-D-galactopyranoside with OR as the aglycon, forms a covalent α-D-galactosyl enzyme intermediate with the nucleophile Glu537 and with assistance from an acid, A (either Glu461 or a magnesium ion). Galactosyl transfer to the nucleophile is shown here as concerted with glycosidic bond cleavage, although this is controversial and may depend on the nature of the leaving group (see text). In the second step (bottom), release of the intermediate is facilitated by a base, B (probably Glu461), which abstracts a proton from the acceptor molecule, R′OH. Galactosyl transfer from the enzyme is shown as stepwise, because there is substantial evidence that all substrates have a trigonal anomeric center in the transition state for this step. (b) General scheme for the action of β-galactosidase on the natural substrate, lactose. The enzyme can either perform hydrolysis (lower path) or transglycosylation (upper path).

by Glu461. In this case, galactosyl transfer is kinetically important. Knowledge of the relative positions of the magnesium, Glu461, and the glycosidic oxygen at different points in the reaction should shed light on the mode of acid catalysis. Regardless of the exact mode of acid catalysis, base catalysis is probably important for degalactosylation in activating an acceptor molecule to attack the enzyme-bound intermediate.

The nature of the transition states for both steps has also been probed using isotope effects (*16*, *19*, *20*, *22*). These experiments suggest that the transition state for step 2 is more planar (sp^2 hybridized) than the starting state (the intermediate), which is consistent with an oxocarbenium ion transition state, as has been proposed for many other glycosidases. The dependence of the rate on acceptor concentration, however, suggests that this transition state involves the acceptor molecule and is thus not a pure S_N1 process. Indeed, displacement reactions at acetal centers are generally thought to occur via a mechanism that is intermediate between S_N1 and S_N2, with a transition state that has substantial oxocarbenium ion character, but still involves bonding to both the incoming and departing groups.

Uncertainty in the mode of acid catalysis makes interpretation of isotope effects for step 1 more ambiguous. Substrates for which step 1 is rate-limiting have relatively small secondary deuterium isotope effects. This could be because galactosylation is concerted (S_N2) with glycosidic bond cleavage or because galactosylation is masked by another event (such as cleavage of the Mg−OR complex in the case of the Mg^{2+} acting as a direct electrophile), in which case galactosylation could still proceed in a stepwise manner (S_N1). Thus, knowledge of the mode of acid catalysis may resolve ambiguities in the geometry of the transition state for step 1.

The enzyme is dependent on mono- and divalent cations for full activity (*6*, *23*). Magnesium activates 5−100-fold, depending on the substrate, and can be substituted with manganese with full activity (*24*). Monovalent cations also affect activity, activating 0.3−6-fold depending on the ion and the substrate (*23*). Sodium and potassium are the most commonly used for kinetic assays.

In the transglycosylation reaction with lactose as substrate (Figure 1b), allolactose is the preferred product (by 97%) and, as the natural inducer for the lac operon, is the trigger for β-galactosidase expression (*25*, *26*). Except in the presence of high concentrations of glucose, this occurs via an internal return process in which the 1−6 glycosidic bond with galactose is formed by the same glucose whose 1−4 glycosidic bond was just cleaved (rather than another glucose molecule from solution) (*25*). In addition to allolactose, other di- and polysaccharides can be formed at low levels, especially when glucose is added to the reaction. With allolactose as substrate, lactose is not produced in measurable quantities (*27*). Thus, although allolactose and lactose are hydrolyzed with equal efficiency ($k_{cat} \simeq 60\ s^{-1}$ and $K_m \simeq 1$ mM for both substrates), only allolactose is produced from the transglycosylation reaction. Additionally, the value of k_{cat} for allolactose production is about the same as for hydrolysis, but this balance is altered by pH and the absence of magnesium (*24*, *26*, *28*).

Some details of enzyme−substrate interactions have been determined previously. Labeling studies initially located Met502, Glu461, and Glu537 to the active site. Further investigation suggested that Glu537 is the nucleophile while Glu461 may act as an acid (*29*−*31*). Mutagenesis has been carried out on several active site residues, including Glu461, Glu537, Tyr503, Asn460, His357, His391, and His540, suggesting specific roles for these residues and, in some cases, their interactions with specific hydroxyls on the substrate (*32*−*39*). A model of a substrate analogue bound to the active site based on NMR has been proposed (*40*). Additionally, some enzyme−substrate interactions have been inferred from crystal structures of related enzymes (*41*, *42*). Nevertheless, detailed three-dimensional structural information about ligand binding to β-galactosidase has lagged behind the biochemical studies, due in part to the large size of the molecule.

Using a monoclinic crystal form it was possible to determine the structure to 2.5 Å resolution (*43*). Subsequently, this was extended to 1.7 Å using an orthorhombic crystal form with a smaller unit cell (*44*). These structures

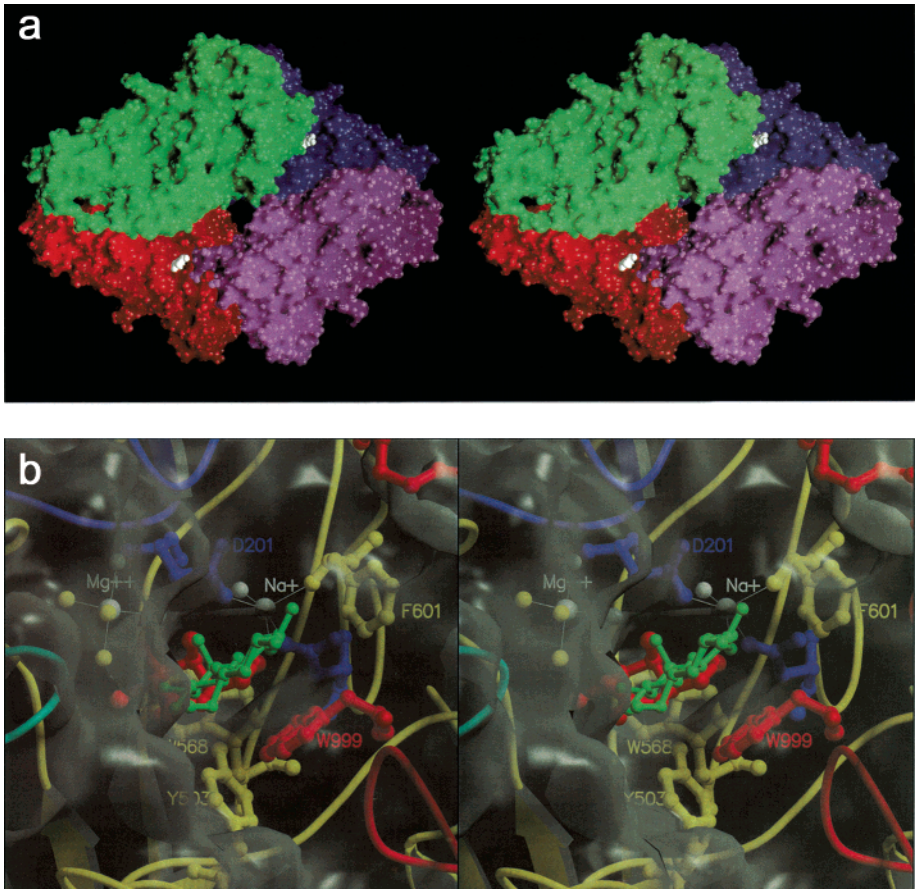

FIGURE 2: Overall views of the β-galactosidase active site. (a) Stereoview of a surface representation of the tetramer in which each subunit is shown in a different color. Two lactose molecules (shown in white) are shown binding in the active site pockets. The deepest parts of the active sites are not visible in the figure. The other two active sites are on the "back" side of the tetramer. The figure was calculated using GRASP (*77*). (b) Closer stereoview of a single active site. Protein atoms are colored by domain with domain 1 in blue, domain 2 in cyan, domain 3 in yellow, and domain 5 in red (there are no atoms from domain 4). Solvent molecules and metals are in white. The yellow balls are metal ligands which are protein atoms. The semitransparent surface was calculated using protein atoms plus solvent atoms with $B < 25$ Å^2. Two ligands are shown: lactose (green) binds in the "shallow" mode while galactonolactone (red) binds in the "deeper" mode.

showed that the enzyme is a homotetramer with 222 point symmetry. Each subunit is composed of five domains, and the active site is a deep pocket (Figure 2) built around the central $(\alpha/\beta)_8$ barrel with other domains apparently recruited to confer specificity for a disaccharide substrate (*45*). All of the residues identified by biochemical experiments as being catalytically important are positioned in or near this pocket. Several divalent and monovalent cation binding sites have been identified, including a magnesium site and a sodium/potassium site in the active site (*44*).

We now report the high-resolution structures of a series of ligands bound to the active site of β-galactosidase. They include substrate analogues, transition state analogues, products, and trapped covalent intermediates. Together, these complexes make it possible to visualize the reaction coordinate, clarifying and enhancing previous proposals for the catalytic mechanism of the enzyme.

EXPERIMENTAL PROCEDURES

Protein Purification and Mutagenesis. E. coli β-galactosidase was expressed and purified as previously reported (*44*, *46*). The E537Q and F601A mutations were made with the Stratagene quick-change kit.

Crystallography. Crystals in space groups $P2_1$ and $P2_12_12_1$ were grown as previously reported (*44*, *46*) and stored in their respective mother liquors (15% PEG 8K, 60 mM cacodylate, pH 5.9, 50 mM NaCl, and 80 mM $MgSO_4$; 10% PEG 8K, 100 mM Bis-Tris, pH 6.5, 100 mM NaCl, 200 mM $MgCl_2$, and 10 mM DTT). Most of the inhibitor soaks were done in stepwise fashion, gradually increasing the concentration up to 100 times K_i. In the case of the covalent intermediates, the soaks were for about 30 min. For the 2-F-galactosyl enzyme 7 mM substrate 2,4-dinitrophenyl 2-F-β-galactopyranoside was added directly to the mother liquor. In the case of the 2-deoxygalactosyl enzyme 10 or 50 mM galactal was used. For the low-temperature structures, the crystals were equilibrated with the cryoprotectant solution (30% DMSO, 70% mother liquor) (*44*, *47*) prior to introducing the inhibitor. Flash cooling was done with a polyethylene loop (Hampton Research) directly into the cold stream.

Diffraction data for different complexes were collected at the Photon Factory, SSRL, or ALS or using a home source (Table 1). For the space group $P2_1$ crystals, processing was done with Weis/Agrovata/Rotavata (*48*, *49*), and the PDB code for the starting model for refinement was 1F49. For the crystals in space group $P2_12_12_1$, data were processed using Denzo/Scalepack ($P2_12_12_1$) (*50*) and Mosflm/Scala ($P2_12_12_1$) (*51*−*53*), and the starting model for refinement was 1DP0 (*45*). Rigid-body refinement was usually sufficient to position the molecule, although molecular replacement

Table 1: Data Collection Statistics[a]

protein	ligand (mM)	X-ray source	space group	cell dimensions (Å) *a*	*b*	*c*	resolution (Å)	$\langle I\rangle/\langle\sigma\rangle$	R_{sym} (%)	unique reflections	redundancy	completeness (%)
early complexes												
E537Q	lactose (100)	9-1	$P2_12_12_1$	149.6	168.6	200.9	1.80	12.5 (4.1)	6.6 (19)	433 670	2.6	92
E537Q	ONPG (100)	9-2	$P2_12_12_1$	149.5	167.5	200.3	1.75	9.0 (2.6)	4.7 (28)	491 766	3.0	97
E537Q	PNPG (150)	9-2	$P2_12_12_1$	149.7	168.6	201.1	1.55	10.0 (4.0)	4.2 (28)	715 525	2.7	98
native*	IPTG (7)	5.0.2	$P2_12_12_1$	151.8	161.2	202.9	1.75	12.0 (2.0)	7.0 (41)	444 288	3.8	90
native	2-F-lactose (89)	6A2	$P2_1$	107.6	207.3	510.3	2.70	6.0 (1.7)	9.0 (41)	409 961	1.8	66
intermediate complexes												
native	2-F-gal-E (7)	6A2	$P2_1$	107.5	207.2	510.2	2.60	7.0 (2.0)	7.1 (36)	477 611	1.9	70
native	2-F-gal-E (7) and glucose (1000)	7-1	$P2_12_12_1$	151.5	167.7	201.6	2.10	10.3 (2.8)	6.0 (34)	273 807	2.6	91
native*	2-d-gal-E (50)	5.0.2	$P2_12_12_1$	149.6	168.2	200.7	1.75	14.0 (2.2)	6.1 (46)	470 782	3.3	94
native*	2-d-gal-E* (10)	RA	$P2_12_12_1$	149.5	169.0	200.8	2.10	7.5 (2.3)	8.1 (27)	264 902	2.7	90
native*	lactone (100)	5.0.2	$P2_12_12_1$	149.7	168.0	201.0	1.80	12.7 (2.0)	6.8 (42)	417 689	3.5	91
native*	tetrazole (0.1)	5.0.2	$P2_12_12_1$	149.7	166.8	200.9	2.10	12.4 (2.7)	8.8 (38)	248 629	4.0	86
product complexes												
native*	galactose (400)	5.0.2	$P2_12_12_1$	149.6	166.5	200.6	1.50	9.1 (2.5)	7.4 (23)	750 734	4.2	90
E537Q	allolactose (30)	9-1	$P2_12_12_1$	149.4	168.7	200.9	1.50	12.4 (3.2)	5.4 (25)	777 574	2.6	99

[a] The variant native* includes an eight-residue substitution on the N-terminus (*44*). Data were collected at beamlines 7-1, 9-1, and 9-2, Stanford Synchrotron Laboratory; 5.0.2, Advanced Light Source; 6A2, Photon Factory; and RA, rotating anode. $P2_1$ data were collected at 288 K, while $P2_12_12_1$ data were collected at ~100 K. For space group $P2_1$, $\beta = 95.0°$. $\langle I\rangle/\langle\sigma\rangle$ gives the ratio of the average intensity to the standard deviation with the value for the highest resolution shell of data given in parentheses. R_{sym} gives the agreement between equivalent reflections. All $P2_12_12_1$ complexes were determined at 100 mM Bis-Tris, with the exception of 2-D-gal-E*, which was determined at 1 mM Bis-Tris.

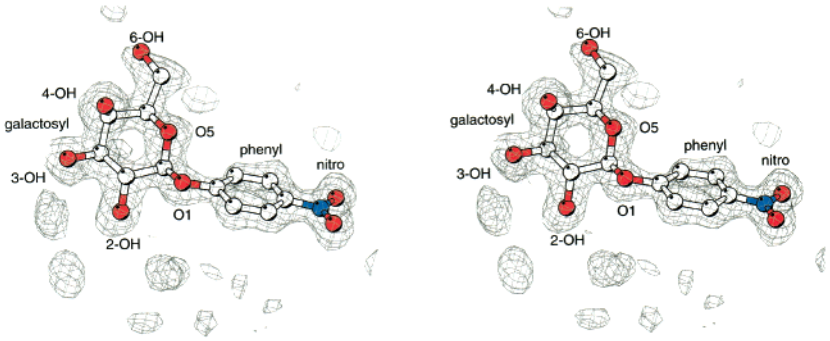

FIGURE 3: Stereoview of a 1.7 Å resolution difference electron density for PNPG bound to the E537Q variant. Coefficients are F_o(E537Q:PNPG) − F_o(native) with phases from the refined native structure. The density shown is for monomer A. The electron density is contoured at $\pm 4\sigma$ (0.24 electron/Å^3). A ball-and-stick model based on the refined coordinates is also shown. Oxygens are in red, carbons are in white, and nitrogens are in blue.

using AMoRe was required for the IPTG complex (*54*). Ligand geometries (bond length, angle, and planarity restraints) were defined using the Cambridge Small Molecule Database. Refinement proceeded using TNT (*55*) with rigid-body refinement at the tetramer, monomer, domain, and secondary structural element levels, followed by positional and *B*-factor refinement. Successive rounds of model building and inspection, including water addition and deletion with the help of Arp (*56*), yielded the final model. An example of ligand electron density is shown in Figure 3.

Synthesis of Allolactose. Allolactose was synthesized enzymatically after Huber et al. (*27*). To 50 mL of 50 mM ONPG and 170 mM glucose in 50 mM Na_2HPO_4, pH 7.0, and 1 mM $MgCl_2$, 1 mg of β-galactosidase was added at room temperature. This was allowed to stir for about 45 min, and the reaction was halted by heating to 80 °C. The nitrophenol was extracted (twice) with 50 mL of chloroform by stirring at room temperature for 30 min. The resulting aqueous solution was filtered (0.2 μm) and lyophilized. This was then redissolved and run on a gel filtration column (Bio-Rad P3) in water, and the fractions were identified using circular dichroism and capillary electrophoresis (*57*). The allolactose peak was identified as allolactose with mass spectrometry (the mass being identical with lactose) and thin-layer chromatography (allolactose running slower than lactose) (*25*). As judged by capillary electrophoresis, the allolactose was 98% pure.

RESULTS

Early Points in the Reaction. The presumed initial steps on the β-galactosidase reaction pathway (Figure 1a) were mimicked in two different ways. First, natural substrates were bound to the altered enzyme, E537Q, which is catalytically incompetent (activity reduced at least 10000-fold, data not shown; *35*). Second, the nonhydrolyzable substrate analogues isopropyl thiogalactoside and 2-F-lactose (*58*) were bound to the native enzyme. In each of these approaches a single atom has been changed to prevent the substrate from being hydrolyzed (either $O^{\epsilon 2} \rightarrow$ N on Glu537, O4 → S on the substrate, or O2 → F on the substrate). All of these "early event" ligands bind similarly, in what may be described as a "shallow" mode (Figure 4). The ligand stacks on the face of Trp999, and the galactosyl hydroxyls 2, 3, and 4 make specific contacts to the enzyme and to bound water, while

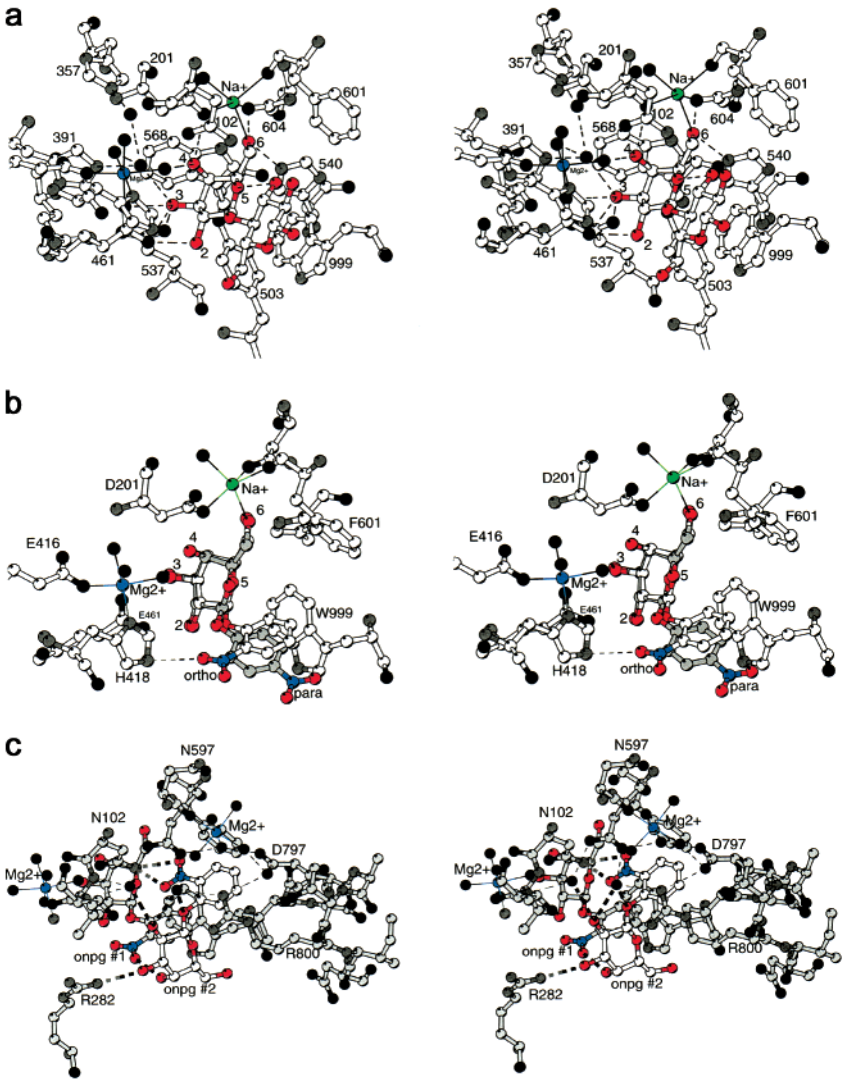

FIGURE 4: "Shallow" binding mode. (a) Stereoview showing lactose bound to the E537Q variant. The galactosyl moiety makes many hydrogen bonds (shown as dashed lines) with the galactosyl hydroxyls, which are numbered, while the glucosyl moiety makes relatively few specific contacts. The sodium ion (green) directly ligands the lactose 6-hydroxyl while the magnesium ion (blue) makes no direct contacts to the lactose. Putative water molecules are shown as black spheres. (b) Stereoview comparing the binding of ONPG and PNPG. The PNPG nitro group makes no specific contacts whereas the ONPG nitro group contacts His418. (c) Stereoview showing the second ONPG molecule binding near the active site. This molecule stabilizes the domain 5 loop (residues 794−804), as well as a second magnesium binding site (upper right of figure). Polar contacts made by the second ONPG molecule are shown as dashed lines.

the 6-hydroxyl contacts the enzyme and a sodium ion. A summary of these interactions is given in Table 3. In the case of ONPG, which was soaked at a concentration of 100 mM (Table 1), a second molecule was seen to bind (Figure 4c). This molecule interacts with the first, as well as with a loop from domain 5, and also appears to stabilize the binding of a presumed Mg^{2+} ion to the carbonyl of Asn597. The mean *B*-factor of the second ONPG molecule is about 35 $Å^2$ while the first is about 26 $Å^2$. Although PNPG and lactose were bound at concentrations equal to or higher than that of ONPG (Table 1), in these complexes there was no evidence for the binding of a second molecule. Complexes of the native enzyme with the ground-state analogues 2-F-lactose and IPTG have geometry very similar to that of ONPG, PNPG, and lactose (Tables 1−3).

Attempts to use low-temperature crystallography to image the natural substrate, lactose, bound to the native enzyme were unsuccessful (not shown). Such attempts resulted in an image of galactose bound in the active site with some residual electron density at the aglycon position.

Intermediate Points along the Reaction Path. True intermediates were isolated using two different substrates for which the breakdown of the enzyme−substrate adduct is much slower than its formation (Figure 1), allowing a buildup of the intermediate in the crystal. First, a 2-F-α-D-galactosyl enzyme intermediate was isolated using dinitrophenyl 2-F-galactoside (*58*). The resulting covalent complex has a half-life for release by water of 45 h. Second, a 2-deoxy-α-D-galactosyl enzyme intermediate was isolated using galactal. This slow-binding inhibitor is predicted to form an intermediate with a half-life of 10 min (*59*), which is long enough to soak the galactal into the crystal and flash-freeze, trapping the intermediate.

In both of the intermediates, the galactosyl ring is in a normal chair conformation and has the expected covalent linkage to Glu537 (Figure 5a). In comparison with the

Table 2: X-ray Refinement Statistics

complex	PDB code	atoms		*R*-factor (%)	*R*-free (%)	deviations from ideal values			average *B*-factor (Å^2)			
		protein	solvent			bonds (Å)	angles (deg)	*B*-factor (Å^2)	protein	solvent	glyc	aglyc
early complexes												
E537Q/lactose	1JYN	32 500	4942	15.5	21.9	0.018	2.9	8.0	17	28	9	23
E537Q/ONPG	1JYV	32 500	4542	18.0	24.0	0.017	2.8	4.9	23	35	16	39
E537Q/PNPG	1JYW	32 500	4929	18.0	22.9	0.020	2.9	7.4	19	32	12	20
native*/IPTG	1JYX	32 500	3926	16.8	24.4	0.015	2.8	5.5	25	34	20	23
native/2-F-lactose[a]	1JYY/1JYZ	131 504[a]	2656[a]	23.4		0.015	2.4	2.5	40	44	61	71
intermediate complexes												
native/2-F-gal-E[a]	1JZ0/1JZ1	131 504[a]	2304[a]	23.0		0.017	2.6	5.7	37	41	31	
native/2-F-gal-E	1JZ2	32 500	3253	16.7	27.0	0.014	2.8	5.1	32	40	24	61
native*/2-d-gal-E	1JZ3	32 506	4767	15.8	22.2	0.016	2.7	8.4	22	34	13	27
native*/2-d-gal-E*	1JZ4	32 500	4350	16.4	26.7	0.016	2.9	5.4	19	28	17	
native*/lactone	1JZ5	32 260	4656	16.0	23.1	0.017	2.9	4.6	26	38	21	
native*/tetrazole	1JZ6	32 500	3325	15.7	26.8	0.017	2.8	6.1	34	41	29	
product complexes												
native*/galactose	1JZ7	32 508	4609	17.6	21.9	0.020	3.0	7.8	19	31	19	
E537Q/allolactose	1JZ8	32 512	5063	16.5	20.8	0.019	2.7	6.6	17	29	16	66

[a] The structures native/2-F-lactose and native/2-F-gal-E were determined using the $P2_1$ crystal form (Table 1) with 16-fold noncrystallographic symmetry. The number of independent atoms is $^1/_{16}$th that listed. For all structures the deviations of the parameters are from a mean small molecule database. Targets are 0.020 Å for bond lengths, 3.0° for bond angles, and 6.0 Å^2 for *B*-factor correlations. The average *B*-factors are for the atoms within each group listed. glyc refers to the glycon (galactosyl) and aglyc refers to the aglycon.

Table 3: Distances between Polar Enzyme Groups and the Galactosyl Substituents in the Various Complexes[a]

interacting partners		early complexes					intermediate complexes		transition state analogues		product complexes	
		E537Q plus lactose	E537Q plus ONPG	E537Q plus PNPG	native plus IPTG	native plus 2-F-lactose	native plus 2-F-gal	native plus 2-deoxy-gal	native plus galactono-lactone	native plus galacto-tetrazole	native plus galac-tose	E537Q plus allo-lactose
2-OH(F)	N460 ($O^{\delta1}$)						3.2					
	N460 ($N^{\delta2}$)						3.0		3.0	3.1	3.0	
	E461 ($O^{\epsilon1}$)								3.1	3.0	3.0	
	E461 ($O^{\epsilon2}$)	2.6	2.6	2.6	2.5	2.8			3.2			2.6
	E537 ($O^{\epsilon1}$)						2.7		2.6	2.6	2.7	
	E537 ($O^{\epsilon2}$)								3.1	3.1	3.1	
	HOH	3.1	3.0	3.1								
3-OH	H391 ($N^{\epsilon2}$)						2.7	2.7	2.7	2.6	2.8	
	H357−HOH						2.7	2.7	2.8	3.1	2.8	
	E461 ($O^{\epsilon1}$)	3.2										3.1
	E461 ($O^{\epsilon2}$)	3.0	3.1	3.1								3.1
	E(Q)537(1)	3.2	3.2	3.0								3.2
	E537 ($O^{\epsilon2}$)				2.7	2.8						
	Mg−HOH(1)						2.6	2.7	2.7	2.8	2.8	
	H391−HOH	2.8	2.9	2.8	3.0							2.8
4-OH	D201 ($O^{\delta2}$)	2.7	2.6	2.6	2.6	3.2	2.7	2.6	2.5	2.6	2.5	2.6
	Mg−HOH(1)	2.5	2.5	2.6	2.8		2.6	2.6	2.8	2.4	2.6	2.6
	H391−HOH				3.0							
6-OH	D201 ($O^{\delta2}$)	3.2	3.1	3.2								
	H540 ($N^{\epsilon2}$)	2.8	2.9	2.8	2.6	2.8	2.8	2.8	2.8	2.6	2.8	2.8
	N604 ($O^{\delta1}$)	2.9	3.0	2.9	3.1	3.2	3.1	2.9	2.9	3.0	2.9	2.9
	Na^+	2.3	2.4	2.3	2.4	2.6	2.5	2.5	2.5	2.6	2.4	2.3
	W568 ($N^{\epsilon1}$)								3.1	3.0		
O(N)1	E461 ($O^{\epsilon2}$)								2.5	2.4	2.5	
	M502 (S^{δ})								3.0		3.2	
O5	N102 ($N^{\delta2}$)	3.1	3.2	3.1	3.1							
	E537 ($O^{\epsilon2}$)						2.5	2.4	3.1			
	Y503 (O^{η})							2.9	3.1			
	HOH						2.5	3.0				
C1	E537 ($O^{\epsilon2}$)								2.9		3.0	

[a] Distances were determined on the basis of monomer A from each complex. Standard deviations for these distances, based on variation between all four monomers, are typically ±0.1. Distances greater than 3.2 Å are not shown.

shallow binding mode, the galactosyl moiety has rotated ~90° and moved deeper into the active site to rest on Trp568 (Figure 5a). The 2- and 3-hydroxyls form new interactions with His391, while the 4- and 6-substituents have also moved deeper but retain the same interaction partners as in the shallow mode. Phe601 has swung into a different rotamer conformation to pack against the hydroxymethyl of the galactosyl moiety. Additionally, a loop from domain 5

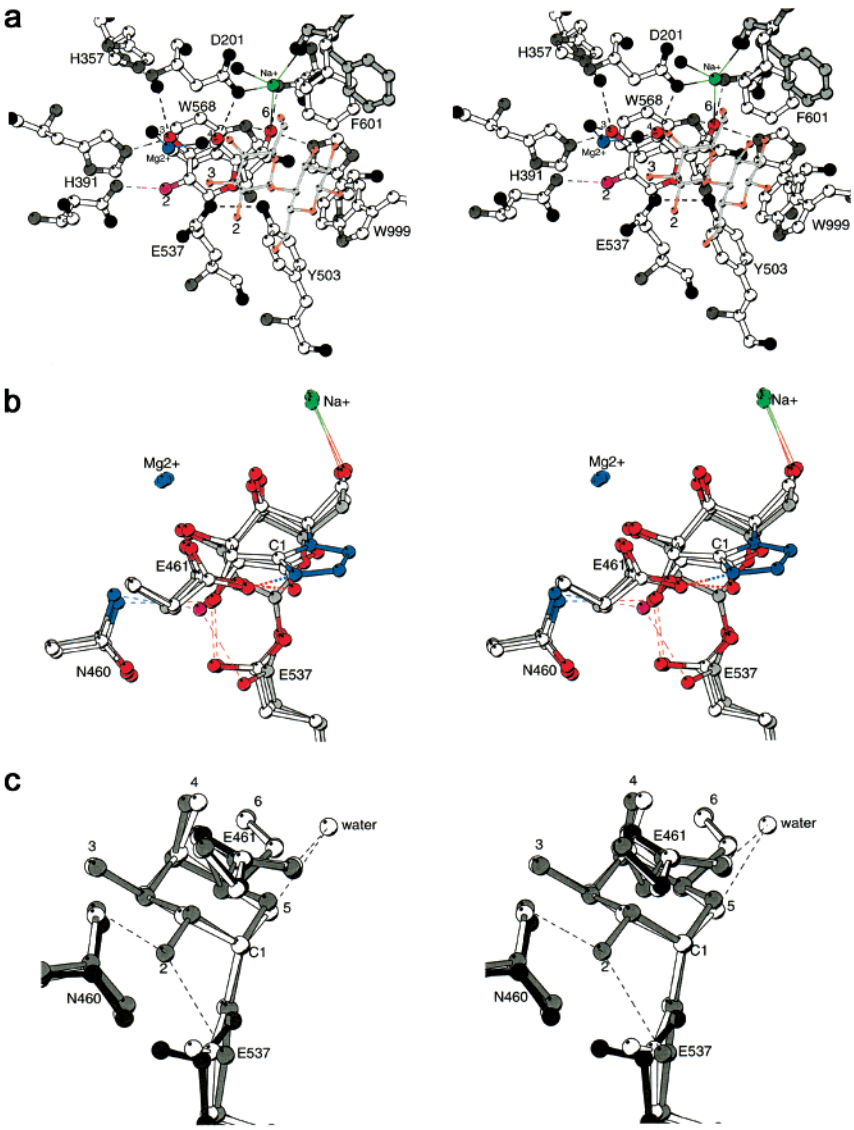

FIGURE 5: "Deep" binding mode. (a) Stereoview comparing the 2-F-galactosyl intermediate (open bonds) to the complex of lactose with mutant E537Q (gray bonds). (b) Stereoview comparing the transition state analogues galactonolactone and galactotetrazole. Binding is similar to the intermediate, which is also shown. Glu461 donates a hydrogen bond to the glycosidic oxygen mimic. This hydrogen bond is donated from the side (in the plane of the ring), rather than above as implied in early work on retaining glycosidases. (c) Stereoview showing the 2-deoxygalactosyl intermediate (gray), the 2-F-galactosyl intermediate (open bonds), and unligated enzyme (black). The galactosyl ring is placed almost identically in the two complexes, suggesting that its binding is dominated by the covalent bond to Glu537 and noncovalent interactions made by hydroxyls 3, 4, and 6. The major difference among the three structures is the rotation of Glu537. This residue (χ_3) rotates 25° in the 2-deoxy intermediate, presumably in association with formation of the covalent bond. The rotation increases by 15° in the 2-F intermediate, probably from repulsion of the Glu537 oxygen by the fluorine at the 2-position.

extending from Gly794 to Pro804 also swings toward the active site (see Discussion).

Several attempts were made to determine the structure of the intermediate for transglycosylation by soaking either the 2-fluoro or 2-deoxy intermediate in high concentrations of glucose (500−1000 mM) (Figure 6). This and other maps suggested that glucose binds adjacent to Trp999, possibly interacting with Glu461, Asn102, and His418. However, a convincing model could not be refined.

Transition State Analogues. The prediction of oxocarbenium ion-like transition states suggests that galactosyl derivatives that are planar at C1 and have a positive charge should mimic transition state binding. Galactotetrazole and galactonolactone are both trigonal at C1 and bind tightly (*60*, *61*). Although they are uncharged, they have an electric dipole moment with its positive end near C1. Both of these transition state analogues bind in the "deep" mode, with Glu461 contacting the atom that corresponds to the glycosidic oxygen (Figure 5b). The lactone binds in the 1−5 form, although this is by far the least populated of the three possible isomers (*60*), illustrating the specificity of the enzyme.

Complexes with Products. The complex with galactose shows that it binds in the deep mode, in common with the intermediates and the transition state analogues (Figure 7a). Data were also collected for the crystals soaked in 500 mM glucose. No significant density could, however, be seen (data not shown) presumably due to its very weak binding ($K_i \sim$ 400 mM).

A complex between allolactose and the E537Q variant was used as a mimic for the product following transglycosylation.

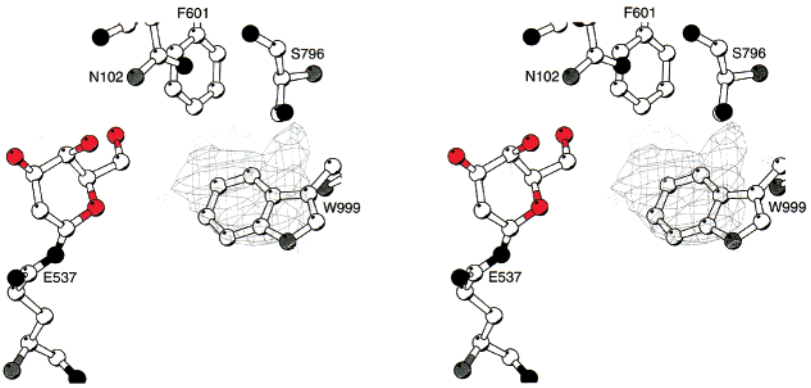

FIGURE 6: Result of an attempt to detect glucose binding to the intermediate (see text). The figure is a stereoview showing a 4-fold averaged 3.0 Å resolution "F_o(glucose) − F_o" electron density contoured at ±4σ (0.08 electron/Å^3). The data sets were collected at room temperature for the 2-deoxygalactosyl intermediate in the presence and absence of 500 mM glucose.

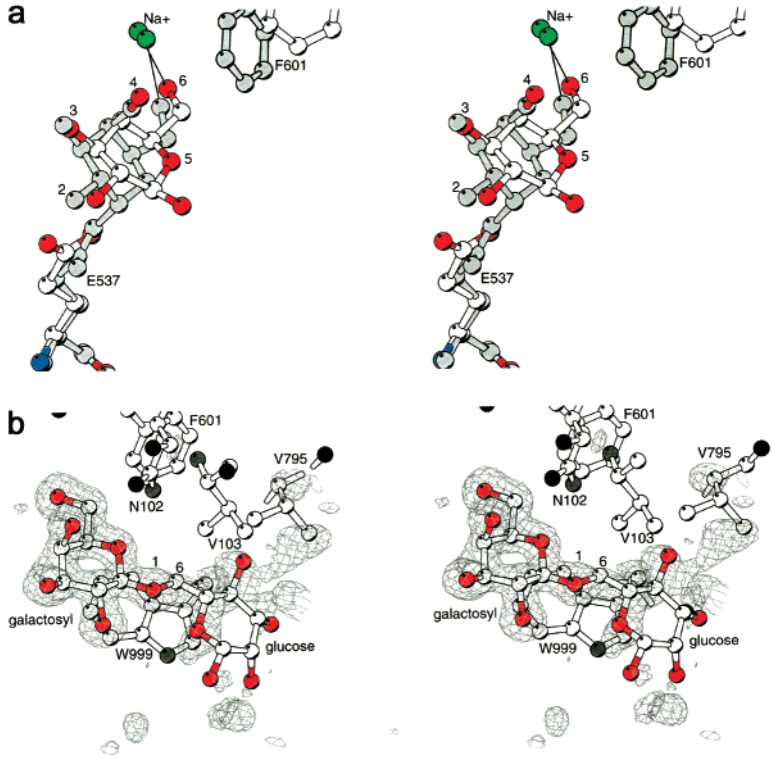

FIGURE 7: Product complexes. (a) Stereoview comparing the binding of galactose (white and red) to the 2-F-galactosyl intermediate (gray). Although overall the galactose binds in the "deep" mode characteristic of intermediates and transition state analogues, the combination of van der Waals repulsion between C1 and Glu537 and the pucker in the sugar ring causes the 6-hydroxyl to be positioned toward the "early" conformation. This causes Phe601 to flip back to the native conformation. (b) Stereoview showing the binding of allolactose to the E537Q variant. The electron density (1.5 Å resolution contoured at 3σ) has coefficients F_o − F_c, where F_c corresponds to the refined coordinates without the allolactose.

The native enzyme could not be used because allolactose is a substrate (*27*). In the complex, the galactosyl moiety occupies the "shallow" binding site. Emanating from the 1-position there is electron density (Figure 7b) that is consistent with the two-atom 1−6 linkage of allolactose but not with the one-atom 1−4 linkage of lactose. The electron density for the remainder of the glucose moiety is weak. Three glucose conformations can be modeled that are consistent with the branching density and have standard torsion angles (*trans*, *gauche*$^+$, and *gauche*$^-$). For all of these conformations, taken together, there are only three contacts to protein less than 3.5 Å and none less than 3.1 Å. These observations suggest that the glucose moiety does not have a clearly preferred mode of binding.

DISCUSSION

General Features. The complexes presented here show that there are two distinct modes for the binding of ligands to β-galactosidase. Substrate analogues and one product (allolactose) bind in a "shallow" site at the mouth of the active site (Figures 2, 4, 7b, and 8a), while intermediates, transition state analogues, and another product (galactose) penetrate 1−4 Å deeper (Figures 2, 5, 7a, and 8b). In both modes, the galactosyl 6-hydroxyl directly ligands the active site sodium ion. In contrast, a direct interaction is never seen between any ligand and the magnesium ion that is in the active site. Except for the galactose complex, the deeper mode of binding is accompanied by an enzyme conforma-

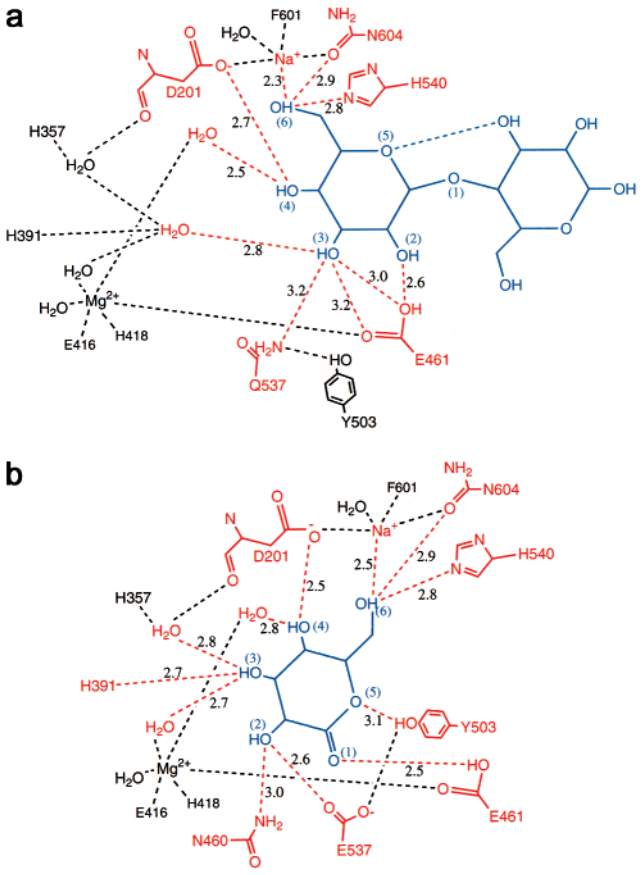

FIGURE 8: Schematic showing hydrogen bonds that are formed on ligand binding. The ligand is shown in blue, with contacting atoms in red. (a) Binding of lactose to the E537Q variant (shallow mode). (b) Binding of galactonolactone to native enzyme (deep mode).

tional change in which Phe601 rotates, and apparently in concert, the 794−804 loop from domain 5 moves up to 10 Å closer to the active site (Figure 9).

On the basis of kinetics experiments, it has been suggested that the substrate binding site is composed of two subsites, one for the galactosyl and the other for the glucosyl moiety (*62*−*64*). The structural data show that there are two overlapping but distinct sites for the galactosyl group, one shallow and one deep, with specific hydrogen bonds between the enzyme and the substrate in each site. In contrast, binding in the glucose subsite is relatively ill-defined, being specified mostly by a stacking interaction with Trp999. This is also reflected in the higher *B*-factors than for the galactosyl moiety (Table 2). This nonspecific binding is consistent with the relative promiscuity of the enzyme for various aglycons.

Galactosylation. The presumed first half of the reaction cycle (Figure 1a) involves cleavage of the glycosidic bond, formation of the galactosyl enzyme intermediate, and release of the aglycon (*8*, *9*). Because of the large distances (~6 Å) between the glycosidic oxygen and potential acid catalysts in the shallow binding mode (see below), the substrate probably needs to move toward the deeper mode before cleavage can occur. Additionally, due to the large nucleophile−anomeric center distance (~6 Å), cleavage in the shallow mode would need to be purely S_N1 and thus require diffusion of an oxocarbenium ion up to 3 Å, which is unlikely given the expected short lifetime. We therefore presume that the substrate moves at least partly and perhaps completely toward the deeper mode before bond cleavage can occur.

Many different galactosides have been analyzed, and while leaving group pK_a is a good predictor of k_{cat} within different aglycon structural classes, the geometry of the aglycon clearly affects substrate reactivity (*16*, *17*, *19*). The principal example of this is that k_{cat} for ONPG hydrolysis is about 10 times higher than for PNPG, although their leaving group pK_a values differ by only 0.02 (*16*), while K_s for ONPG is also about 10 times higher than for PNPG. The reason for this difference is not obvious from the structures. One possibility is that movement of the substrate toward the "deeper" mode could cause a steric clash which would be worse for an ortho than a para substituent. This could inhibit binding of the former (elevating K_s) but might also cause distortion to accelerate cleavage (elevating k_{cat}). Another possibility is that the ONPG nitro group has a polar interaction with His418, which is itself ligated to the active site magnesium ion (Figure 4b). This interaction, which it appears can be maintained at least part of the way to the deeper mode, could improve the electron-withdrawing character of the nitro group, thus making it easier to delocalize an electron on the nitrophenyl group and to break the glycosidic bond. In contrast, the PNPG nitro group interacts with bulk solvent and is more remote from the site of cleavage. Whether the binding of a second ONPG molecule (Figure 4c) might also contribute to the increase in k_{cat} is not known.

Some aspects of the initial catalytic events have been somewhat controversial. It is thought the enzyme uses magnesium-dependent acid catalysis to facilitate leaving group departure. The biochemical data are consistent with either Lewis catalysis by the magnesium ion or Bronsted catalysis by Glu461, as discussed in the introduction. While the structures presented here cannot provide an unequivocal resolution, in no case is there a direct contact between a ligand and the magnesium ion (Figures 4 and 5). Furthermore, a Mg−O−PNP or Mg−O−ONP group cannot be modeled into the active site without introducing steric clashes, even allowing for some side chain movements. This suggests that direct electrophilic attack by the magnesium ion is unlikely. Additionally, the transition state analogues suggest that Glu461 is in good position to donate a proton to the glycosidic oxygen, supporting its role as an acid catalyst for galactosylation (Figure 5b). The direct interaction between Glu461 and the magnesium ion does suggest, however, that the magnesium ion may play a role in the ability of Glu461 to act as an acid catalyst. Such a secondary role for the magnesium ion has been proposed earlier on the basis of Mn^{2+} NMR, which suggested a relatively large (8−9 Å) distance between the Mg^{2+} and the glycosidic oxygen (*65*). Whether, as has been suggested by Richard et al. (*21*), the magnesium plays a dynamic role in tuning the pK_a of Glu461 during the reaction is unclear, but we find that the magnesium−ligand distances change by less than 0.2 Å between the various complexes.

Most of the ambiguities regarding acid hydrolysis are based on studies of galactosylation with *aryl* galactosides, rather than the natural substrate, lactose. With lactose hydrolysis the leaving group is considerably more basic than with ONPG or PNPG, and it is to be expected that acid catalysis should play a larger role (*66*).

Degalactosylation. The second half of the reaction cycle, degalactosylation, is less controversial than the first. Isotope

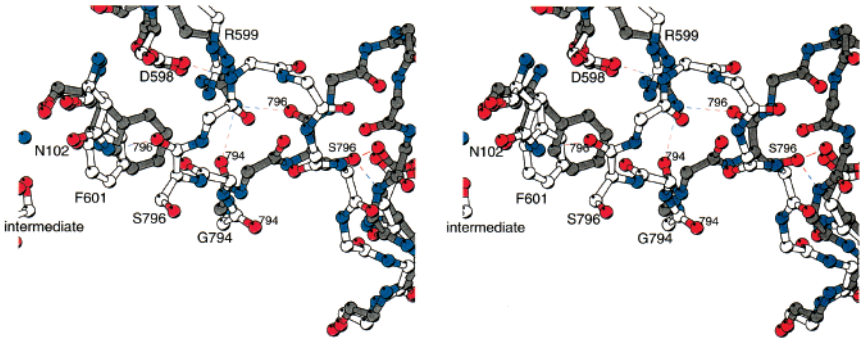

FIGURE 9: Stereoview showing the conformational switch of the loop that includes residues 794−804. The native, or open, conformation is shown in gray while the intermediate, or closed, conformation has open bonds. The Ser796 side chain, which forms inter- and intraloop hydrogen bonds in the open conformation, moves about 10 Å to pack against the side chain of Phe601.

effects suggest that the transition state is more sp^2 hybridized than the starting state (the covalent adduct), consistent with an S_N1 release (*67*). However, the dependence of the rate on acceptor concentration suggests that the acceptor is involved in the transition state (*67*). This again is consistent with an oxocarbenium ion-like transition state which forms only in the presence of both incoming and leaving groups. A molecule that could play the role of the incoming group can be seen in the covalent adducts, contacting Glu461 and the galactosyl ring oxygen (Figure 5c).

Transition State Stabilization. Although the intermediates and transition state analogues all bind in the deeper subsite, there are some subtle differences which suggest how β-galactosidase may stabilize the transition state. The chief difference between the two covalent intermediates (2-deoxygalactosyl and 2-F-galactosyl) is the orientation of Glu537 (Figure 5c). Relative to the rest of the protein the ligand moves very little, suggesting that its binding orientation is determined by the specific interactions with hydroxyls 3, 4, and 6. Overall, the lactone and tetrazole bind similarly but somewhat rotated relative to the intermediates, shifting the 2-substituent by ~0.7−0.8 Å and the other substituents by ~0.2−0.5 Å (Figure 5b). This shift is probably due to steric repulsion between Glu537 and the C1 mimic and has also been seen in an analysis of aza-sugar inhibitor binding to the xylanase Cex from *Cellulomonas fimi* (*68*).

Measurements on a series of deoxy and fluoro analogues of 2,4-dinitrophenyl galactoside (*58*) suggested that interactions between the enzyme and the 3-, 4-, and 6-hydroxyls provide at least 4 kcal/mol of transition state stabilization for galactosylation, whereas the 2-hydroxyl provides about 8 kcal/mol. Similar experiments with other ligands and with mutant enzymes suggest that about 4−8 kcal/mol of transition state stabilization is provided by His357, His391, and His540, which interact with the 3- and 6-hydroxyls (*36*, *37*, *39*). The importance of the 2-substituent correlates well with the structures presented here, which show that it is this group that changes its interactions the most. The large stabilization provided by the 2-hydroxyl may be reflected in its interactions with both Asn460 and Glu537 (Figure 5c). Between the lactone/tetrazole complex and the intermediates, the 2-substituent moves through the Glu537 carboxylate plane. Since a hydrogen bond to this carboxylate would be strongest in the plane, the true transition state may lie on the point in the reaction coordinate at which the 2-hydroxyl is in this plane. The role of the other hydroxyls is presumably to properly orient the galactosyl moiety.

The product galactose binds much less tightly than the transition state analogues galactonolactone and galactotetrazole. The reasons for this difference in binding affinity (about 11 kcal/mol) are, however, not obvious since the enzyme−ligand interactions are similar but by no means identical (Figure 7a). Lactone and tetrazole binding is thought to use cooperative interactions with Glu461 and Glu537, such that hydrogen bond donation by Glu461 to the glycosidic oxygen mimic increases the dipole moment of the inhibitor, which in turn strengthens the interaction with Glu537 (*69*). Calculations suggest this may account for a few kilocalories of binding energy (*69*). This proposal is consistent with the binding modes of both lactone and tetrazole (Figure 5b). Glu461 $O^{\epsilon 2}$ is about 2.5 Å from the glycosidic oxygen mimic, in good position to donate a hydrogen bond, while Glu537 is about 2.9−3.3 Å from C1. Although the distances are similar for bound galactose, the galactose molecule probably has a smaller dipole moment and is less polarizable along the C1−glycosidic oxygen mimic bond, preventing it from taking full advantage of the electrostatic environment of the active site.

The hydrogen bond from Glu461 to the glycosidic oxygen comes from the side, near the plane of the galactosyl ring, as predicted (*70*). Such in-plane protonation appears to be general and contrasts with earlier general models for retaining β-glycoside hydrolysis, in which protonation is described as coming from perpendicular to the pyranoside plane. The result is that, for retaining glycosidases, the two catalytic acids are not 180° apart, on opposite sides of the substrate, but 90° apart, in closer proximity to each other. This proximity has implications for cycling the acid pK_a through the reaction cycle (see below).

The Michaelis Complex and the Role of Strain. In the shallow binding mode (Figure 4a), C1 is about 6 Å from the nucleophile Glu537, while O4 is about 6 Å from the acid Glu461. Since these distances are presumably too large for any direct enzyme assistance to take place, the question arises as to whether this mode of binding is on the reaction coordinate or is nonproductive. Because the active site is at the bottom of a relatively deep pocket, it is difficult to imagine full penetration without the substrate going through a complex similar, if not identical, to the shallow mode. This is supported by mutagenesis of active site histidines, 357, 391, and 540, which showed that while all three of these residues are important for binding transition state analogue inhibitors, their effect on the binding of substrate analogues varies (*36*, *37*, *39*). Figure 4a shows that this variability

mirrors the proximity to the ligand. His540 directly ligands the 6-hydroxyl, and its replacement has the largest effect on binding. His357 interacts indirectly via water molecules, and its replacement causes the smallest decrease in substrate analogue binding. His391 has intermediate proximity to the ligand, and its replacements have an intermediate effect.

Cleavage of lactose, the natural substrate, proceeds via a much slower first step than does cleavage of ONPG. Thus lactose can serve as a competitive inhibitor for ONPG hydrolysis. Since ground- state interactions are responsible for the binding events prior to the first step, the binding of lactose may be similar to substrate analogue inhibitors. This also suggests that the shallow mode lies along the reaction coordinate for action on lactose. It probably does not correspond exactly to the ground state, however, since the alkyl structure−reactivity parameters indicate that a moiety on the enzyme can partially neutralize a negative charge that develops on the leaving group oxygen (*21*). This moiety is most likely Glu461, which would require true ground-state binding to be slightly deeper than the shallow mode.

Glycon distortion to a boat conformation in retaining glycosidases is thought to aid in catalysis by opening access to C1 for nucleophilic attack and creating optimal orbital overlap between the ring oxygen and C1 for formation of an oxocarbenium ion (*12*, *71*). Indeed, such distortion has been observed in putative ground-state complexes in other systems (*72*, *73*). There is no evidence, however, for this distortion in the shallow mode of binding to β-galactosidase. This, of course, does not rule out distortion as part of the reaction coordinate. In fact, modeling the binding of lactose in the deeper mode suggests that distortion of the galactosyl group to a boat would be helpful by reducing the number of steric clashes between the aglycon and the enzyme.

One possible reason for the apparent lack of substrate distortion with the observed complexes is that the substrate for β-galactosidase, being a disaccharide, is much smaller than the polymeric substrates of the other enzymes. In such cases there is some evidence that the distortion of one sugar is paid for with the binding interactions of the others (*72*). In the case of β-galactosidase, the leaving group comprises only a single saccharide, and this has very few specific interactions that might drive distortion of the galactosyl group.

Despite the apparent lack of distortion in the observed complexes, it seems likely that flattening of the pyranose ring would improve the binding of galactose, given the 10^3−10^8-fold better binding of molecules such as the lactone and tetrazole. Such flattening would cause only subtle changes in the observed binding geometry, although it might be enough to trigger the conformational change seen with the transition state analogues (see below). We suggest, however, that this flattening occurs only on the way to the transition state, rather than in the Michaelis complex.

Another kind of distortion has been predicted for substrate binding to β-galactosidase. Using NMR and model building, Espinosa et al. (*40*) suggested that the inhibitor, C-lactose, adopts a high-energy gauche−gauche conformation. However, the crystallographic complexes of lactose with E537Q and 2-F-lactose with the native enzyme show no evidence for such a conformation.

Details of the Conformational Switch. As the substrate penetrates more deeply into the active site pocket, the side chain of Phe601 follows and, apparently in concert, the 794−804 loop swings toward the active site (Figure 9). This change appears to be accompanied by, and perhaps triggered by, a change in the coordination of the active site sodium ion. In the native, unligated structure, the sodium ion has square pyramidal geometry with two water ligands and three protein ligands. In the early complexes the galactosyl 6-hydroxyl replaces one of the water ligands (Figure 4a). This ligand prevents Phe601 from swinging into the new rotamer well. In the transition to the intermediate, the 6-hydroxyl remains ligated to the sodium but moves about 20° on the sodium sphere and 0.1−0.2 Å further away, providing room for Phe601 to swing in (Figure 5a). Repositioning Phe601 eliminates its packing with Arg599, which in turn apparently destabilizes the 794−804 loop to which Arg599 contributes two hydrogen bonds (Figure 9). A new loop conformation, with Ser796 packing on the repositioned Phe601 and alternative hydrogen bonds with Asn102 and Asp598, is then favored. The change in galactosyl position required to trigger this conformational change is very small, as the galactose complex is open, while the lactone complex is closed. Generally, the loop is not as well ordered as other parts of the protein in this region. Also, the binding of additional ligands to the intermediate, specifically Bis-Tris or DMSO, inhibits the loop switch (*47*). To test the importance of Phe601 in stabilizing the 794−804 loop, the variant F601A was constructed and crystallized. Difference electron density maps calculated with data collected at room temperature show that the loop becomes less well ordered, suggesting that Phe601 contributes directly to its stability (*47*).

A conformational change during the catalytic cycle of β-galactosidase has long been suggested on the basis of both physical and kinetic grounds. Deschavanne et al. (*62*) observed differences in the absorbance spectrum of the enzyme in the presence of IPTG and galactal. This is consistent with our observation that galactal binding causes a conformational change by forming a 2-deoxygalactosyl intermediate, whereas IPTG binding does not. Prior to this, Sinnott and co-workers proposed a magnesium-dependent rate-determining conformational change as a way of explaining both the lack of secondary deuterium isotope effects for slow substrates and the overall scatter in structure−reactivity plots (*16*, *17*). Subsequently, Rosenberg and Kirsch (*19*) measured oxygen-18 kinetic isotope effects for hydrolysis of PNPG, which suggested that a conformational change is not rate-determining for hydrolysis of slow substrates. They proposed instead that such substrates proceed via S_N2 hydrolysis, which explains the secondary deuterium kinetic isotope effects. This model has the disadvantage of requiring some substrates with poor leaving groups (the pyridinium salts) to proceed via S_N1 hydrolysis and other substrates, also with poor leaving groups (certain aryl galactosides), to proceed via S_N2. Subsequently, Selwood and Sinnott (*20*) pointed out that this contradiction could be avoided if the magnesium ion acts as a direct electrophile.

As described above, the structures presented here suggest that the magnesium ion would be sterically prevented from acting as a direct electrophile. This suggests then that aryl galactosides with poor leaving groups, such as PNPG, do proceed via S_N2. Some rough model building offers an explanation for the apparent paradox of the pyridinium salts

proceeding via S_N1. It is clear that for an S_N2-type reaction the ligand must move from the shallow mode toward the deep mode. Such movement would require some repositioning of the aglycon. With the aryl galactosides, there are two bonds about which to rotate for such repositioning, while only one for the pyridinium salts. Additionally, if galactosyl distortion toward a boat conformation is required for an S_N2 reaction, this would appear to force the pyridinium ring against the active site wall, while the extra degree of freedom with the aryl galactosides may allow the galactosyl distortion without this steric clash. Thus the correct aglycon positioning for an S_N2 reaction may be sterically hindered with the pyridinium salts.

Although the observed conformational change cannot be rate-determining, it may play a role in some of these results. It seems quite likely that the conformational change would be favored differently depending on the aglycon. Since the aglycon binding site would be altered, it seems reasonable that this would then perturb an otherwise linear structure−reactivity plot.

Transglycosylation and the Production of Allolactose. Both hydrolysis and transglycosylation are of physiological importance to the bacterium. The latter allows for production of the natural *lac* operon inducer, which is necessary for β-galactosidase production. The complexes presented here offer some insight into this reaction.

Generally, the molecule occupying the aglycon site has few specific interactions. Indeed, the glucose moiety is only well ordered in the E537Q/lactose complex. Well-defined binding is not seen in either the E537Q/allolactose complex or in the intermediates soaked with high-glucose concentrations. Some electron density can be seen in these complexes, suggesting that the glucose is bound but in multiple conformations. Relatively poor binding of glucose to the intermediate ($K_i \simeq 20$ mM) has been measured biochemically (*62*, *64*). Despite the absence of clear binding determinants, transglycosylation is relatively specific. Production of allolactose is favored by greater than 97% over other disaccharides (*26*). A possible explanation for this specificity lies in the transition from the shallow to the deeper binding modes. Lactose can bind in the shallow mode. Once galactosylation has taken place with the galactosyl group moving deeper into the active site, reattaching the glucose at the 1-, 2-, 3-, or 4-positions appears to be sterically difficult. In contrast, the 6-hydroxyl, which has an extra atom linking it to the galactose ring, can reach further into the active site pocket and more easily attack the enzyme-bound intermediate. In general terms, the relatively weak binding of glucose may reflect a balance between extremely poor binding, which would result in 100% hydrolysis, and extremely tight binding, which could, in principle, result in 100% transglycosylation.

Overall Mechanism of Catalysis. Synthesis of the extensive prior literature on β-galactosidase (see the introduction) with the present structural data suggests the following mechanism. The initial binding of substrate is characterized by stacking on Trp999, by a network of specific interactions made by the galactosyl hydroxyls and by relatively few specific interactions made by the aglycon. Assuming initial binding is similar to the shallow mode, the substrate will tend to move deeper into the active site pocket, improving interactions made between the enzyme and the 2- and 3-hydroxyls. These interactions, combined with those made by the 4- and 6-hydroxyls, appear to facilitate penetration of the substrate into the active site. As the substrate moves toward the deeper mode, Glu461 contacts the glycosidic oxygen and then the nucleophile contacts the anomeric center. Distortion of the galactosyl group to an inverted boat may be required for the intact substrate to proceed to the deeper mode.

Once the substrate is in the correct position, Glu461 is appropriately placed to donate a proton to the glycosidic oxygen in concert with formation of the intermediate with Glu537. Nearly concerted galactosylation explains the absence of secondary deuterium kinetic isotope effects for step 1 while proton donation by Glu461 would result in a solvent isotope effect.

The formation of the covalent intermediate effectively neutralizes the charge on the nucleophile, Glu537. Since this residue is in close proximity to the acid, Glu461, the charge neutralization may decrease the pK_a of Glu461, making it more likely to be deprotonated (*13*, *74*), which, in turn, allows it to act as a base during the second step. Alternatively, subtle changes in the interactions with the magnesium ion could modulate the Glu461 pK_a. This part of the mechanism is still uncertain.

As the substrate moves deeper into the active site, the side chain of Phe601 follows, as does the 794−804 loop. Although there is no direct evidence, this loop might help to restrict access to the intermediate in such a way as to ensure that allolactose is the preferred transglycosylation product.

The second half of the reaction involves galactosyl transfer from the nucleophile Glu537 to an acceptor molecule. This transfer involves restoration of the negative charge on Glu537, which may be facilitated by partial proton donation by Tyr503 (*38*). During this process, a trigonal oxocarbenium ion develops, which appears to be stabilized by interactions between Glu537, Tyr503, and the galactosyl ring oxygen. The trigonal coordination is consistent with secondary kinetic isotope effects (*8*). The tyrosine equivalent to Tyr503 has been proposed to play a similar role in two related systems, *Sinapis alba* myrosinase and a xylanase from *Bacillus circulans* (*75*, *76*). Glu461 facilitates attack of the oxocarbenium ion by abstracting a proton from the acceptor molecule. Since Glu537 is now negatively charged, Glu461 should be a stronger base, better able to fulfill this role.

In the case of hydrolysis, the leaving group diffuses away and appears to be replaced by a water molecule that hydrogen bonds to Glu461 and the ring oxygen of the galactosyl moiety. Since this water molecule does not have direct access to C1, the galactosyl moiety presumably must first form a trigonal species to be attacked by the water molecule, activated by Glu461. The move toward the transition state lengthens the O−C1 bond, forcing the sugar to rotate and causing the 2-hydroxyl to move across the carboxylate plane. As the galactose product becomes puckered, the sugar rotates farther, changing the interaction of the 6-hydroxyl with the sodium ion. Apparently in concert, Phe601 moves away, allowing the galactose to diffuse out of the active site.

During the first step, if the glucose molecule does not diffuse away, it can bind in several ways. These modes probably include interactions with Asn102 and possibly

His418. Since the covalent adduct is sequestered deep in the active site pocket, the hydroxyl with the longest connection to the pyranose ring (the 6-hydroxyl) has the greatest probability of releasing the intermediate. Thus galactosyl-1−6-glucose (allolactose) is the preferred transglycosylation product.

ACKNOWLEDGMENT

We thank the staffs of the Photon Factory, Advanced Light Source, and Stanford Synchrotron Radiation Laboratory for help with data collection, Leslie Gay and Drs. Todd Lowther and Walt Baase for technical assistance, and Dr. Gene Huber for helpful comments on the manuscript.

REFERENCES

1. Beckwith, J. R., and Zipser, D. (1970) *The Lactose Operon*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Miller, J. H., and Reznikoff, W. S. (1980) *The Operon*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
3. Müller-Hill, B. (1996) *The lac Operon*, Walter de Gruyter, New York.
4. Jacob, F., and Monod, J. (1961) *J. Mol. Biol. 3*, 318.
5. Ullmann, A., Perrin, D., Jacob, F., and Monod, J. (1965) *J. Mol. Biol. 12*, 918.
6. Wallenfels, K., and Weil, R. (1972) in *The Enzymes* (Boyer, P. D., Ed.) 3rd ed., Vol. 7, p 617, Academic Press, London.
7. Müller-Hill, B., and Kania, J. (1974) *Nature 249*, 561.
8. Sinnott, M. L. (1978) *FEBS Lett. 94*, 1.
9. Sinnott, M. L. (1990) *Chem. Rev. 90*, 1171.
10. Zabin, I. (1982) *Mol. Cell. Biochem. 49*, 87.
11. Huber, R. E., Gupta, M. N., and Khare, S. K. (1994) *Int. J. Biochem. 26*, 309.
12. Davies, G., Sinnott, M. L., and Withers, S. G. (1998) in *Comprehensive Biological Catalysis* (Sinnott, M. L., Ed.) p 136, Academic Press, San Diego.
13. Richard, J. P. (1998) *Biochemistry 37*, 4305.
14. Koshland, D. E. (1953) *Biol. Rev. 28*, 416.
15. Stokes, T. M., and Wilson, I. B. (1972) *Biochemistry 11*, 1061.
16. Sinnott, M. L., and Souchard, I. J. (1973) *Biochem. J. 133*, 89.
17. Sinnott, M. L., Withers, S. G., and Viratelle, O. M. (1978) *Biochem. J. 175*, 539.
18. Richard, J. P., Westerfield, J. G., and Lin, S. (1995) *Biochemistry 34*, 11703.
19. Rosenberg, S., and Kirsch, J. F. (1981) *Biochemistry 20*, 3189.
20. Selwood, T., and Sinnott, M. L. (1990) *Biochem. J. 268*, 317.
21. Richard, J. P., Huber, R. E., Lin, S., Heo, C., and Amyes, T. L. (1996) *Biochemistry 35*, 12377.
22. Richard, J. P., and McCall, D. A. (2000) *Bioorg. Chem. 8*, 49.
23. Cohn, M., and Monod, J. (1951) *Biochim. Biophys. Acta 7*, 153.
24. Huber, R. E., Parfett, C., Woulfe-Flanagan, H., and Thompson, D. J. (1979) *Biochemistry 18*, 4090.
25. Jobe, A., and Bourgeois, S. (1972) *J. Mol. Biol. 69*, 397.
26. Huber, R. E., Kurz, G., and Wallenfels, K. (1976) *Biochemistry 15*, 1994.
27. Huber, R. E., Wallenfels, K., and Kurz, G. (1975) *Can. J. Biochem. 53*, 1035.
28. Huber, R. E., Gaunt, M. T., Sept, R. L., and Babiak, M. J. (1983) *Can. J. Biochem. Cell Biol. 61*, 198.
29. Sinnott, M. L., and Smith, P. L. (1978) *Biochem. J. 175*, 525.
30. Legler, G., and Herrchen, M. (1981) *FEBS Lett. 135*, 139.
31. Gebler, J. C., Aebersold, R., and Withers, S. G. (1992) *J. Biol. Chem. 267*, 11126.
32. Cupples, C. G., Miller, J. H., and Huber, R. E. (1990) *J. Biol. Chem. 265*, 5512.
33. Edwards, R. A., Cupples, C. G., and Huber, R. E. (1990) *Biochem. Biophys. Res. Commun. 171*, 33.
34. Ring, M., and Huber, R. E. (1990) *Arch. Biochem. Biophys. 283*, 342.
35. Yuan, J., Martinez-Bilbao, M., and Huber, R. E. (1994) *Biochem. J. 299*, 527.
36. Roth, N. J., and Huber, R. E. (1996) *J. Biol. Chem. 271*, 14296.
37. Roth, N. J., Rob, B., and Huber, R. E. (1998) *Biochemistry 37*, 10099.
38. Penner, R. M., Roth, N. J., Rob, B., Lay, H., and Huber, R. E. (1999) *Biochem. Cell Biol. 77*, 229.
39. Huber, R. E., Hlede, I. Y., Roth, N. J., McKenzie, K. C., and Ghumman, K. K. (2001) *Biochem. Cell Biol. 79*, 183.
40. Espinosa, J. F., Montero, E., Vian, A., Garcia, J. L., Dietrich, H., Schmidt, R. R., Martin-Lomas, M., Imberty, A., Canada, F. J., and Jimenez-Barbero, J. (1998) *J. Am. Chem. Soc. 120*, 1309.
41. Jenkins, J., Leggio, L. L., Harris, G., and Pickersgill, R. (1995) *FEBS Lett. 362*, 281.
42. Sakon, J., Adney, W. S., Himmel, M. E., Thomas, S. R., and Karplus, P. A. (1996) *Biochemistry 35*, 10648.
43. Jacobson, R. H., Zhang, X. J., DuBose, R. F., and Matthews, B. W. (1994) *Nature 369*, 761.
44. Juers, D. H., Jacobson, R. H., Wigley, D., Zhang, X.-J., Huber, R. E., Tronrud, D. E., and Matthews, B. W. (2000) *Protein Sci. 9*, 1685.
45. Juers, D. H., Huber, R. E., and Matthews, B. W. (1999) *Protein Sci. 8*, 122.
46. Jacobson, R., and Matthews, B. (1992) *J. Mol. Biol. 223*, 1177.
47. Juers, D. H. (2000) Doctoral Dissertation, Physics Department, University of Oregon, Eugene, OR.
48. Higashi, T. (1989) *J. Appl. Crystallogr. 22*, 9.
49. Dodson, E. J., Winn, M., and Ralph, A. (1997) *Methods Enzymol. 277*, 620.
50. Otwinowski, Z., and Minor, W. (1997) in *Macromolecular Crystallography* (Carter, C. W., and Sweet, R. M., Eds.) Part A, p 307, Academic Press, New York..
51. Kabsch, W. (1988) *J. Appl. Crystallogr. 21*, 916.
52. Leslie, A. G. W. (1991) in *Crystallographic Computing 5, From Chemistry to Biology* (Moras, D., Podjarny, A. D., and Thierry, J. C., Eds.) p 50, Oxford University Press, Oxford.
53. Evans, P. R. (1993) in *Data Reduction*, Proceedings of a CCP4 Study Weekend on Data Collection and Processing, p 114.
54. Navaza, J. (1994) *Acta Crystallogr. A50*, 157.
55. Tronrud, D. E. (1997) *Methods Enzymol. 277*, 306.
56. Lamzin, V., and Wilson, K. (1993) *Acta Crystallogr. D49*, 129.
57. Zeleny, R., Altmann, F., and Praznik, W. (1997) *Anal. Biochem. 246*, 96.
58. McCarter, J. D., Adam, M. J., and Withers, S. G. (1992) *Biochem. J. 286*, 721.
59. Wentworth, D. F., and Wolfenden, R. (1974) *Biochemistry 13*, 4715.
60. Huber, R. E., and Brockbank, R. L. (1987) *Biochemistry 26*, 1526.
61. Heightman, T. D., Ermert, P., Klein, D., and Vasella, A. (1995) *Helv. Chim. Acta 78*, 514.
62. Deschavanne, P. J., Viratelle, O. M., and Yon, J. M. (1978) *J. Biol. Chem. 253*, 833.
63. Huber, R. E., and Gaunt, M. T. (1983) *Arch. Biochem. Biophys. 220*, 263.
64. Huber, R. E., Gaunt, M. T., and Hurlburt, K. L. (1984) *Arch. Biochem. Biophys. 234*, 151.
65. Loeffler, R. S. T., Sinnott, M. L., Sykes, B. D., and Withers, S. G. (1979) *Biochem. J. 177*, 145.
66. Richard, J. P., Westerfield, J. G., Lin, S., and Beard, J. (1995) *Biochemistry 34*, 11713.
67. Sinnott, M. L., and Viratelle, O. M. (1973) *Biochem. J. 133*, 81.
68. Notenboom, V., Williams, S. J., Hoos, R., Withers, S. G., and Rose, D. R. (2000) *Biochemistry 39*, 11553.
69. Heightman, T. D., Vasella, A., Tsitsanou, K. E., Zographos, S. E., Skamnaki, V. T., and Oikonomakos, N. (1998) *Helv. Chim. Acta 81*, 853.
70. Heightman, T., and Vasella, A. T. (1999) *Angew. Chem., Int. Ed. 38*, 750.

71. Rose, G., and White, A. (1997) *Curr. Opin. Struct. Biol. 7*, 645.
72. Sulzenbacher, G., Driguez, H., Henrissat, B., Schulein, M., and Davies, G. J. (1996) *Biochemistry 35*, 15280.
73. Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S., and Vorgias, C. E. (1996) *Nat. Struct. Biol. 3*, 638.
74. McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996) *Biochemistry 35*, 9958.
75. Burmeister, W. P., Cotlaz, S., Driguez, H., Tori, R., Palmieri, S., and Henrissat, B. (1997) *Structure 5*, 663.
76. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) *Biochemistry 38*, 5346.
77. Nicholls, A., Sharp, K., and Honig, B. (1991) *Proteins 11*, 281.

BI011727I