Escherichia coli β-Galactosidase Recognizes a High-Energy Conformation of C-Lactose, a Nonhydrolizable Substrate Analogue. NMR and Modeling Studies of the Molecular Complex

J. F. Espinosa,[†] E. Montero,[†] A. Vian,[‡] J. L. García,[‡] H. Dietrich,[†] R. R. Schmidt,[§] M. Martín-Lomas,[⊥] A. Imberty,[||] F. J. Cañada,*[,†] and J. Jiménez-Barbero*[,†]

Contribution from the Departmento Química Orgánica Biológica, Instituto Química Orgánica, CSIC, Juan de la Cierva 3, 28006 Madrid, Spain, CIB−CSIC, Velazquez, 28006 Madrid, Spain, Fak. Chemie, University Konstanz, Germany, Instituto de Investigaciones Químicas, CSIC, Isla de la Cartuja, Sevilla, Spain, and CERMAV−CNRS, Grenoble, France

Received July 10, 1997. Revised Manuscript Received November 5, 1997

Abstract: The enzyme-bound conformation of C-lactose, an *Escherichia coli* β-galactosidase inhibitor has been determined by NMR spectroscopy. It is demonstrated that the enzyme selects a high-energy conformation of this closely related structural analogue of the natural substrate, lactose. In addition, a molecular modeling protocol has been performed in order to obtain a detailed three-dimensional structure of the complex that can explain, in structural terms, the role that the key amino acid residues play in the catalytic mechanism. The implications of the recognition of a high-energy conformation of the analogue are also outlined.

Introduction

E. coli β-galactosidase has been widely employed as model enzyme for studies of glycosidases which act on disaccharides.[1,2] It is a retaining glycosidase enzyme with tetrameric structure and one active site per monomer, and its X-ray structure has been recently determined.[3] Each monomer consists of 1023 amino acid residues and its natural substrate is lactose. The mechanism of action has not been firmly established, but it is generally accepted that it involves a double-displacement reaction in which the enzyme first forms and then hydrolyzes a glycosyl−enzyme intermediate via oxocarbonium ion-like transition state, similar to that described for lysozyme,[4] and other glycosidases.[1,2,5−7]

Over the past years a number of studies have been reported to identify the residues that are responsible for catalysis. Specifically, Glu-461, Glu-537, and Tyr-503 have been confirmed to be essential for enzymatic activity.[8−10] The importance of Glu-461 has been shown by site-directed mutagenesis, and it is involved in the binding of Mg^{2+} ion. This residue was first thought to be the enzymic nucleophile; however, the finding of covalent attachment of an inhibitor to Glu-537 indicated that, in fact, Glu-537 is the active nucleophile. With respect to Tyr-503, inactivation has suggested that this residue is probably acting as the acid and base catalyst in the mechanism. A recent finding has also suggested the involvement of His-540 in the stabilization of the transition state by forming a hydrogen bond with galactose.[11]

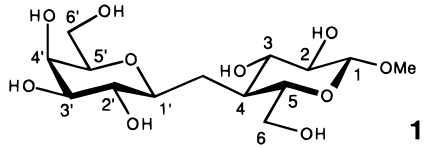
The X-ray three-dimensional structure of β-galactosidase has been recently solved[3] showing that Glu-461, Tyr-503, and Glu-537 are found close together and located in a deep pocket within a distorted "TIM" barrel. His-540 is located in the wall of the active-site cavity. Unfortunately, only the structure of the free enzyme is available to date, and the position and conformation of lactose within the binding pocket remains unknown. This information may be essential for a full understanding of the catalytic mechanism, as well as for the determination of the exact role that each residue plays in the reaction.

Although the high molecular weight of β-galactosidase precludes direct ^{1}H NMR observations using current strategies,[12] information about the conformation of complexed oligosaccharides can be derived from transferred NOE studies (TRNOE) provided that the exchange between the complexed and uncomplexed states is sufficiently fast.[13] Obviously, the hydrolysis of the natural substrate does not allow the use of lactose for this sort of experiments. Nevertheless, another approach could be the use of a structurally related inhibitor to obtain 3D structural information. In this context, the search for new inhibitors has led to a group of oligosaccharide analogues with

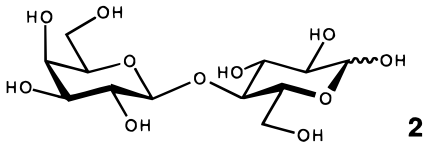
* Correspondence to Dr. J. Jiménez-Barbero and Dr. F. J. Cañada, Instituto Química Orgánica, CSIC, Juan de la Cierva 3, 28006 Madrid, Spain. E-mail: iqojj01@pinar.csic.es. Fax: 34-1-5644853. Phone: 34-1-5622900.
† Instituto Química Orgánica.
‡ CIB−CSIC, Velazquez.
§ University Konstanz.
⊥ Instituto de Investigaciones Químicas.
|| CERMAV−CNRS.

(1) Sinnott, M. L. *Chem. Rev.* **1990** *90*, 1171−1202.
(2) Huber, R. E.; Gupta, M. N.; Khare, S. K. *Int. J. Biochem.* **1994** *26*, 309−318.
(3) Jacobson, R. H.; Zhang, X.-J.; DuBose, R. F.; Matthews, B. W *Nature* **1994**, *369*, 761−766.
(4) Phillips, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 484−495.
(5) Legler, G. *Adv. Carbohydr. Chem. Biochem.* **1990**, *48*, 319−385.
(6) McCarter, J. D.; Withers, S. G. *Curr. Opin. Struct. Biol.* **1994**, *4*, 885−892.
(7) Davies, G.; Henrissat, B. H. *Structure* **1995**, *3*, 853−859.
(8) Huber, R. E.; Chivers, P. T. *Carbohydr. Res.* **1993** *250*, 9−18.
(9) Gebler, J. C.; Aebersold, R.; Withers, S. G. *J. Biol. Chem.* **1992**, *267*, 11126−11130.
(10) Ring, M.; Huber, R. E. *Arch. Biochem. Biophys.* **1990** *283*, 342−350.
(11) Roth, N. J.; Huber, R. E. *J. Biol. Chem.* **1996**, *271*, 14296−14301. It has to be noticed that, in this publication, it is mentioned the observation of this interaction in the X-ray structure of one enzyme−inhibitor complex.
(12) James, T. L.; Oppenheimer, N. J. Nuclear magnetic resonance. Part C. *Methods in Enzymology* **1994**, *239*, 1−813.
(13) Ni, F. *Prog. NMR Spectrosc.* **1994**, *26*, 517−606.

S0002-7863(97)02291-9 CCC: $15.00
Published on Web 02/03/1998



1



2

Figure 1. Schematic view of lactose **1** and C-lactose **2** showing the atomic numbering.

the glycosidic oxygen substituted by carbon (C-glycosides) that are resistant to hydrolysis.[14] Therefore, the determination of the conformation and orientation of C-lactose (the C-glycoside analogue of lactose) within the binding pocket would be of great interest.

On this basis, we here report on the use of transferred NOE experiments to determine the enzyme-bound conformation of the ligand, assisted by molecular modeling to get a detailed 3D structure of the complex that can explain the role that each residue plays in the mechanism.

To the best of our knowledge this strategy represents a novel and different approach to the understanding of the properties of this important enzyme, which, in turn, could be extended to related systems.

Results and Discussion

Kinetic Studies. The first step in our analysis was to show that the C analogue (**1**, Figure 1) is an inhibitor of β-galactosidase. Consequently, experiments to determine the inhibition constant K_i were carried out. Although the modified enzyme[15] (see Methods) should work similarly to its native counterpart, the K_i for isopropyl β-D-thiogalactoside (IPTG), a known competitive inhibitor of β-galactosidases was calculated. The hydrolysis of *o*-nitrophenyl-β-D-galactopyranoside (0.5 mM) was measured by HPLC (see methods) in the presence of different concentrations of IPTG. The K_i value obtained (0.1 mM) is comparable to that of the wild-type enzyme ($K_i = 0.085$ mM).[11] The same method was applied to C-lactose, giving a K_i of 3.3 mM for a competitive inhibitor model, of the same order as that described for lactose ($K_i = 1.2$ mM).

NMR Studies. TR-NOE experiments were performed to deduce the bound conformation of **1**. For ligands which are not bound tightly and exchange with the free ligand at reasonably fast rate, the transferred nuclear Overhauser enhancement (TR-NOE) experiment provides an adequate means to determine the conformation of the bound ligand.[13] In complexes involving large molecules, cross relaxation rates of the bound compound are opposite in sign to those of the free ligand and produce negative NOEs. The conditions for the applicability of this approach are well established:

$$p_b\sigma^B > p_f\sigma^F \quad (1)$$

$$K_{-1} \gg \sigma^B \quad (2)$$

where p_b and p_f are the fractions of bound and free ligand and σ^B and σ^F the cross relaxation rates for the bound and the free ligand, respectively. K_{-1} is the off-rate constant. This approach has been recently applied to several studies of lectin- and antibody-bound oligosaccharides.[16]

TR-NOESY Experiments (Figure 2) were performed at different mixing times giving rise to strong and negative NOEs, as expected for ligand binding. Our previous NMR and modeling studies on free C-lactose demonstrated that three different conformational families coexist in solution.[17] Approximately, 55% of population adopts the anti conformation (ϕ/ψ 36/180, where the glycosidic torsion angles ϕ and ψ are defined as H1′-C1′-Ca−C4 and C1′-Ca−C4−H4, respectively) while about 40% presents the syn conformation (ϕ/ψ 54/18), which is the conformation displayed by the previously reported structures for different $\beta(1\rightarrow4)$ equatorial-linked disaccharides. The experimental data also indicate that a high energy minimum (ϕ/ψ, 180/0) is also slightly populated (∼5%). This last conformation has been called gauche−gauche since displays the Cα−C4 linkage gauche with respect to both C1′-O5′ and C1′-C2′ bonds and represents a conformation never observed for free $\beta(1\rightarrow4)$ natural disaccharides.[18]

The existence of these three conformational families could be detected by the presence of NOEs that unequivocally characterize the different regions of the conformational map. These NOEs have been dubbed exclusive NOEs. For C-lactose, H1′-H4, H1′-H3, and H4−H2′ are exclusive NOEs for the syn, anti, and gauche−gauche conformations, respectively. Consequently, the corresponding NOE intensities will be sensitive to their respective populations. Therefore, and at least qualitatively, a first indication of the bound conformation can be obtained by focusing on these key NOEs.

The comparison between the NOESY spectra of C-lactose (Figure 2) recorded in the absence and in the presence of the enzyme shows important and clear differences. Some of the cross peaks in the NOESY spectrum of the free ligand are no longer displayed in the TR-NOESY spectrum of the complex. It is important to stress the disappearance of both H1′-H3 and H1′-H4 NOEs evidencing that neither the anti conformation nor the syn conformation are recognized by the enzyme. By contrast, the H4/H2′ NOE which display a weak intensity for the free ligand is now the strongest one in the spectrum displaying an intensity of about 12%. These findings clearly indicate that the bound conformation belongs to the high-energy gauche−gauche family. Fortunately, the presence of two additional methylene protons for C-lactose, in contrast to the natural glycosides, permits detection of more experimental constraints. Provided that the gauche−gauche conformation is indeed bound by the enzyme, additional cross peaks involving these two protons would be expected. In fact, HproR/H3, HproR/H5, HproS/H3, HproS/H5, and HproS/H6 cross peaks (Figure 2B) are also observed in the TR−NOESY spectra series, indicating the recognition of the gauche−gauche conformation (Figure 3). TR-ROESY experiments[16] were also carried out to exclude spin difussion effects (Supporting Information). The above-mentioned cross peaks showed different sign to the diagonal peaks thus excluding the possibility of protein-relayed or spin difussion-mediated correlations.

It is noteworthy to point out that these experimental results indicate that β-galactosidase recognizes a high energy conformer of C-lactose. This conformation has a small population for free

(14) Levy, D. E.; Tang, C. *The Chemistry of C-glycosides*; Pergamon: New York, 1995.

(15) Sánchez-Puelles, J. M.; Sanz, J. M.; García, J. L.; García, E. *Eur. J. Biochem.* **1992**, *203*, 153−159.

(16) Peters, T.; Pinto, B. M *Curr. Opin. Struct. Biol.* **1996**, *6*, 710−720.

(17) Espinosa, J.-F.; Asensio, J. L.; Cañada, F. J.; Jimenez-Barbero, J.; et al. *J. Am. Chem. Soc.* **1996**, *118*, 10862−10871.

(18) Asensio, J. L.; Cañada, F. J.; Jimenez-Barbero, J. *Eur. J. Biochem.* **1995**, *233*, 618−630.

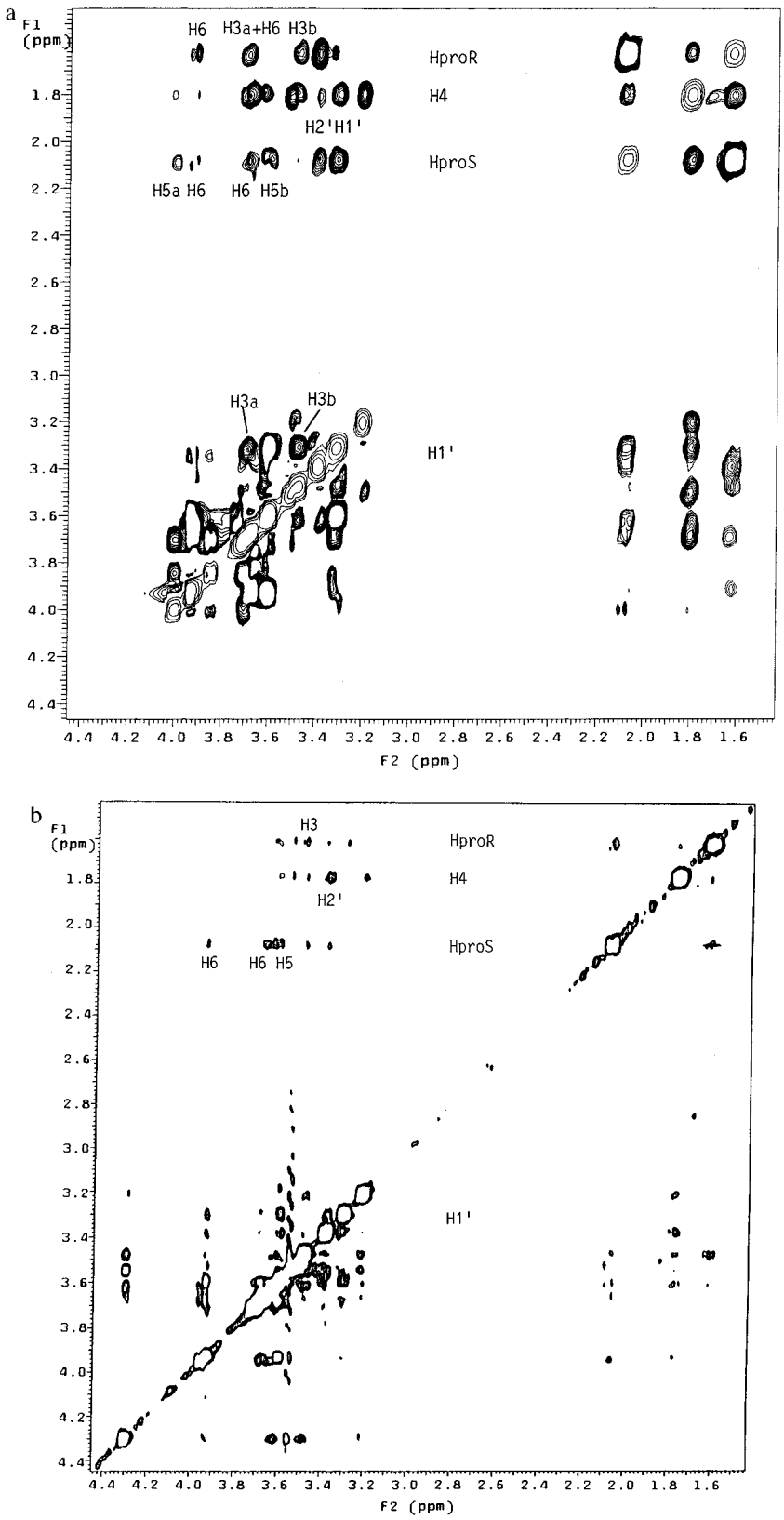


Figure 2. (A) NOESY spectrum for free C-lactose at 303 K and (B) TR-NOESY spectrum for C-lactose bound to *E. coli* β-galactosidase molar ratio 24:1 under the same experimental conditions. Relevant cross peaks are indicated.

1 (<5%), while it is not present for natural lactose in water solution. In addition, this conformation has not been detected upon binding of either lactose or C-lactose to ricin, a toxic galactose-binding protein.[19] Taking into account the energetic differences between the high energy and the global minimum conformer of C-lactose in solution, which amounts to about 9 kJ/mol, it seems that the enzyme shows a intrinsic binding energy around this magnitude. Thus, it is interesting to evaluate

(19) Espinosa, J.-F.; et al. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 303−306.

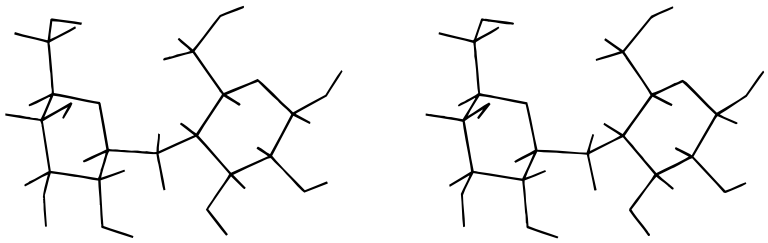


Figure 3. Stereoview of the gauche−gauche conformer of C-lactose.

up to which point the recognition of a highly distabilized confomer may have implications in the initial recognition of the natural substrate by the enzyme. First, a competitive TR-NOE experiment was performed by adding increasing amounts of IPTG to a NMR tube containing C-lactose and the enzyme, to confirm that the binding of C-lactose was active site selective. It was observed that the TR-NOE signals of C-lactose disappear at equimolecular concentrations of IPTG with concomitant appearance of strong NOEs for this molecule.

Second, it should be stressed, that although the natural glycoside and the C-glycoside are indeed fairly structurally related analogues, the recognized conformation of the synthetic analogue does not necessarily represent that of the natural compound. Indeed, we have previously shown that the ricin β-bound conformations of lactose and C-lactose significantly differ in terms of the glycosidic Ψ torsion angle.[17−19] For *E. coli* β-galactosidase, the direct comparison of both compounds is not possible, since obviously, lactose, the natural substrate, after being recognized by the protein, is readily transformed into products. Nevertheless, we feel that it is not appropriate to draw too close an analogy between ricin and the β-galactosidase. The lectin binds primarily to the galactose, whereas β-galactosidase will need to bind to both the glycon and the aglycon. It would be most surprising if a C-glycoside analogue binds to a glycosidase in a substantially different conformation than the natural substrate. In relation with other results published in the literature for different protein-bound oligosaccharides, there are cases in which the protein bind oligosaccharides near their global minimum conformation, although there are examples of major conformational variations upon binding.[16−19]

Docking Studies. The next step was to dock C-lactose within the enzyme binding pocket to determine the 3D structure of the complex, using the putative binding site of the *E. coli* enzyme as deduced from the recent X-ray crystallographic analysis.[3] The Tripos force field within the Sybyl package was used for this purpose. This represents a suitable force field to perform conformational analysis of oligosaccharides, since new parameters have been developed for them,[20] which are compatible with those for proteins. In most docking studies a substrate is manually docked within the receptor binding site, but this protocol should be taken with caution, due to possible bias.[20] We chose a different procedure, by employing the GRID method.[21] In this protocol, the interaction of a probe group with a protein of known structure is computed at sampled positions throughout and around the macromolecule, giving an array of energy values. This method is less time-consuming than other protocols (i.e., the "crankshaft" method[20]), and it has been widely applied to ligand binding studies. In addition, the presence of both hydrophilic and lipophilic surfaces on the carbohydrate makes this protocol suitable for our proposal since both OH (hydroxyl) and CH_2 (methylene) probes can be used for the energy calculations. As a first approximation, and since there is not any published structure of a complex of the enzyme with either an inhibitor or substrate analogue, we have considered that the enzyme does not undergo major conformational changes upon complexation.[22]

The analysis of the GRID calculations indicated that in the binding region the most favorable zones of interaction for the OH and the CH_2 probes appear to be clustered in one site located between amino acids Asn-102, Glu-461, Tyr-503, Glu-537, and His-540 and just above the aromatic Trp-568. The shape of the cluster is fairly reminiscent of a carbohydrate ring shape, with the OH on the outside, associated with CH_2 inside. Different possibilities to fit a galactose residue in the GRID isocontours were tested. For the best agreement, all the hydroxyl groups (with the exception of the anomeric one) fit the OH-favorable region while the C1, C2, C3, and C6 carbons fit the methylene favored area. (It is noteworthy to mention that a 60° rotation would bring O1 between the Mg^{2+} and Glu-537, which would be still compatible with the obtained isocontours.) Nevertheless, in addition to a less perfect fit for the galactose moiety, it is not possible to dock C-lactose when using this second orientation. Energy minimizations of this complex yielded the orientation shown in Figure 4. Minor movements are observed during the geometry optimization, and, indeed, the galactose remained fairly close to its starting orientation. Seven intermolecular hydrogen bonds are established between the monosaccharide and the protein: Tyr-503·OH→O1, Glu-537·OD1→O1, Glu-537·OD2→O2, Asn-460·ND2→O2, Asn-102·ND2→O4, Tyr-503·OH→O5, and His-540·NE2→O6. All the hydroxyl atoms with the exception of O-3 are involved in at least one hydrogen bond. Hydroxyl oxygens O-1 and O-2 are involved in cooperative and bidentate hydrogen bonds, which have been shown to be classical in protein−carbohydrate interactions.[23] The interactions of Trp rings with apolar carbohydrate faces are also widely spread in protein/carbohydrate complexes, both in the solid state[23] and in solution.[24] Indeed, in the model shown here, the galactose methine hydrogen atoms H1, H3, and H5 creates a flat hydrophobic surface which display an excellent stacking interaction with Trp-568.

Next, the pseudodisaccharide C-lactose was docked within the binding site. The galactose moiety was superimposed onto the corresponding geometry of the previously optimized enzyme/monosaccharide complex. Then, a systematic search of the possible conformations around the glycosidic linkages was performed by variations of ϕ and ψ angles (see Methods). Since the binding site is located in a deep pocket of the structure,

(20) Imberty, A.; Hardman, K. D.; Carver, J. P.; Perez, S. *Glycobiology* **1991**, *1*, 631−642.

(21) Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849−857.

(22) Nevertheless, some sort of conformational changes have been especulated to occur in the second step of the reaction upon formation of the galactosyl-enzyme intermediate.

(23) Vyas, N. K *Curr. Opin. Struct. Biol.* **1991**, *1*, 732−740.

(24) Asensio, J. L.; Cañada, F. J.; Bruix, M.; Rodríguez-Romero, A.; Jiménez-Barbero, J. *Eur. J. Biochem.* **1995**, *230*, 621−633.

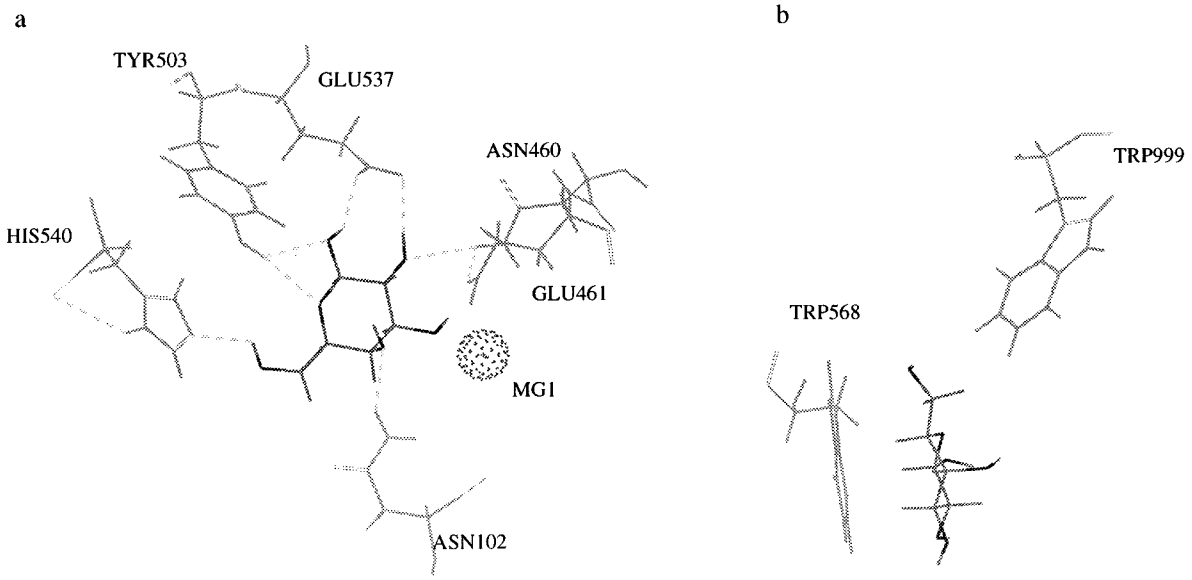


Figure 4. Different views of the three-dimensional structure of the β-galactosidase/galactose complex after docking studies: (A) relevant intermolecular hydrogen bonds and (B) stacking between Trp-568 and the nonpolar face of galactose.

Table 1. Hydrogen Bonds and Hydrophobic Contacts Deduced for the *E. coli* β-Galactosidase/C-Lactose Complex

solution 1: $\Phi = 57.2$, $\Psi = -117.6$, $\Phi_H = 175.9$, $\Psi_H = -0.5$		solution 2: $\Phi = 57.2$, $\Psi = -117.6$, $\Phi_H = 172.8$, $\Psi_H = 1.5$	
hydrogen bonds	hydrophobic contacts	hydrogen bonds	hydrophobic contacts
intramolecular	stacking	intramolecular	stacking
Glc·O3→Gal·O2	Gal·H1···Trp568	Glc·O3→Gal·O2	Gal·H3···Trp568
Gal·O4→Glc·O6	Gal·H3···Trp568		Gal·H4···Trp568
intermolecular	Gal·H5···Trp568	intermolecular	Gal·H5···Trp568
Asn460·ND2→Gal·O2		Gal·O2→Glu537·OD1	Gal·H62···Trp568
Asn102·ND2→Gal·O4	Glc·H1···Trp999	Gal·O3→Glu537·OE1	
Gal·O6→His540·NE2	Glc·H61···Trp999	Gal·O6→Asn102·OD1	Glc·H5···Trp999
Glc·O2→Met502·SD	Glc·H62···Trp999		Glc·H61···Trp999
Glc·O3→Glu537·OE1	others	Glc·O3→Met502·SD	Glc·H62···Trp999
	Glc·H71···Tyr503·CE2	Asn102·ND2→Glc·O6	Glc·H71···Trp999
	Glc·H72···Tyr503·CE2		Glc·H72···Trp999
	Gal·H4···Asp102·CB		others
	Glc·H2···His418·CE		Gal·H1···Tyr503·CE2
	Glc·H2···Glu461·CG		Glc·H71···Tyr503·CE2
	Glc·H3···Met502·CB		Glc·H72···Tyr503
			Glc·H2···His418·CE
			Glc·H4···His418·CE

indeed only one conformational family was allowed within the binding site, which corresponds to the gauche−gauche conformation. Neither the syn nor the anti conformations can exist in the binding site without severe steric conflicts. As discussed above, the obtained geometry is completely in agreement with the experimental NMR results.

Two protocols were then taken to get a energy-refined complex. In the first one, this complex was refined while keeping the coordinates of the galactose ring fixed. The resulting complex optimized in this way has been named solution 1 and its characteristics are listed in Table 1. Obviously, since the position of galactose had been constrained, this moiety exhibits the same hydrogen bonds and the same stacking mode with Trp-568 previously described (Figure 5). The glucose residue also displays stacking with a second aromatic residue, Trp-999. Glucose O-2 also presents a polar contact with the sulfur atom of Met-502, and indeed, this residue has been covalently modified[25] with an active site-directed reagent, (β-galactopyranosylmethyl)(*p*-nitrophenyl)triazene.

Finally, all the constraints affecting the galactose moiety were removed during the energy minimization. A global translation of the disaccharide was observed to yield a second docking mode, called solution 2 (Figure 6). It can be observed that the galactose moiety still presents van der Waals contacts with Trp-568. The hydrogen-bond network is also slightly different (Table 1). Additional hydrophobic contacts are observed through the interaction of the methylene group at the glycosidic linkage with Trp-999. This fact may explain why the ligand changes its orientation in the binding site when completely optimized. From an energetic point of view, solution 2 is favored, the reason being the carbohydrate structure, and particularly a distortion of the O5−C1−Cα valence angles in solution 1. Therefore we can infer that solution 2 represents the most likely binding mode of C-lactose by *E. coli* β-galactosidase. However, it is not possible to discriminate wether the binding mode of the natural substrate is closer to any solution, since the bulky CH_2 at the pseudoglycosidic linkage causes certain steric hindrance with the protein. Nevertheless, it has to be stressed, with regard to the bound conformation that for both 1 and 2 solutions, only the gauche−gauche conformation can be docked and it is impossible to fit either the syn or the anti within the binding pocket.

With a model of the bound conformation at hands, theoretical calculations of the expected NOEs using solution 2 were

(25) Fowler, A. V.; Zabin, I.; Sinnott, M. L.; Zabin, I. *J. Biol. Chem.* **1978**, *253*, 5283−5285.

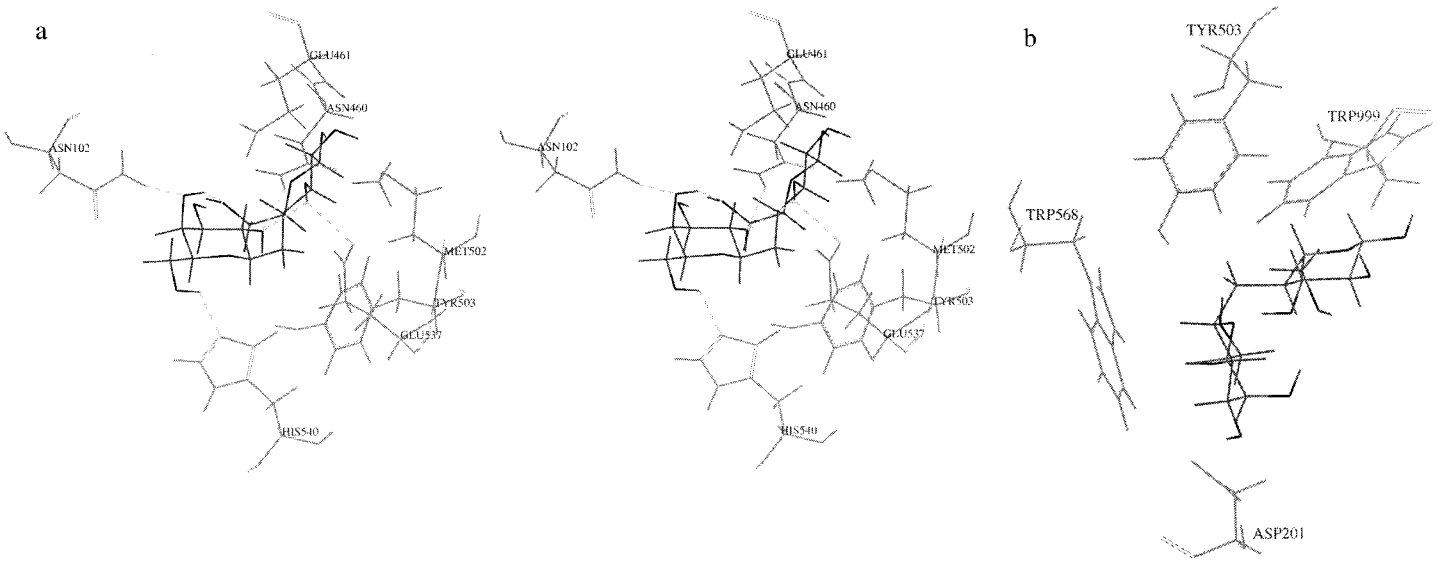


Figure 5. (A) Stereoview of solution 1 for the three-dimensional structure of the β-galactosidase/C-lactose complex after docking studies. The relevant intermolecular hydrogen bonds are indicated. (B) Stacking between Trp-568 and Trp-999 and the nonpolar faces of the galactose and glucose residues, respectively. The orientation of Tyr-503 close to the glycosidic linkage is also indicated.

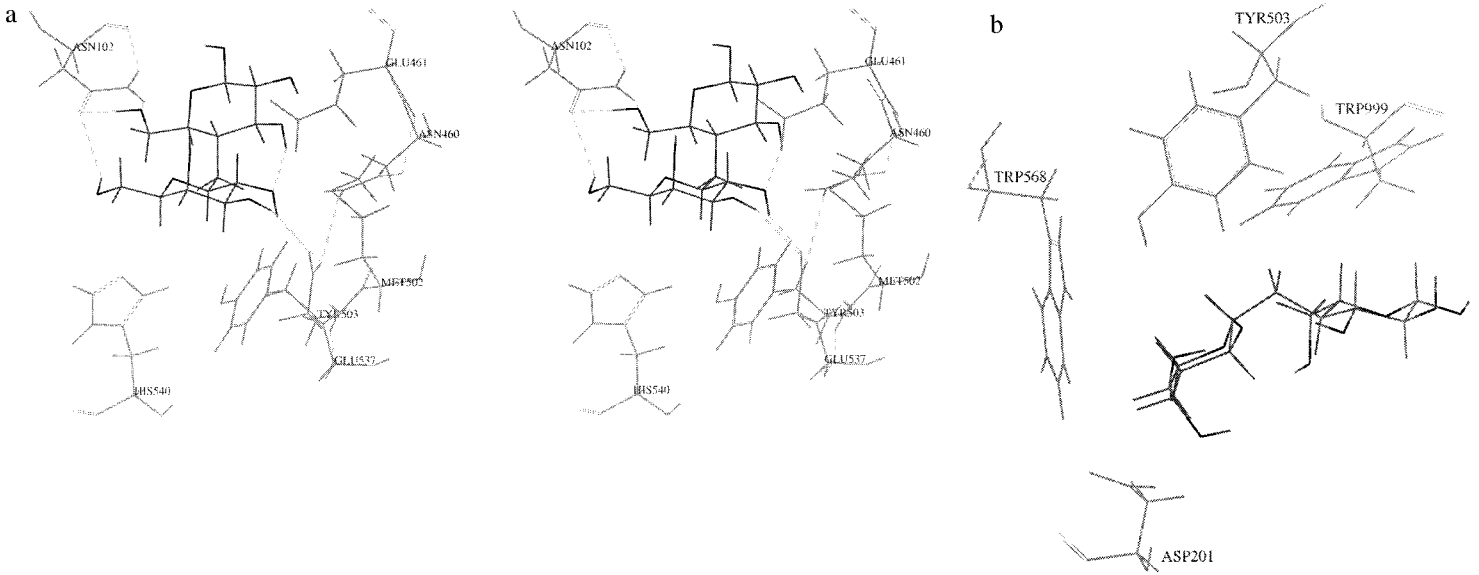


Figure 6. (A) Stereoview of solution 2 for the three-dimensional structure of the β-galactosidase/C-lactose complex after docking studies. The relevant intermolecular hydrogen bonds are indicated. (B) Stacking between Trp-568 and Trp-999 and the nonpolar faces of the galactose and glucose residues, respectively. The orientation of Tyr-503 close to the glycosidic linkage is also indicated.

Table 2

	intensity %							
	H1′/H3′+H5′	H4/H2′	HpR/H3	HpR/H5	HpS/H3	HpS/H5	HpS/H6pR	HpS/H6pS
calcd	17.7	9.3	12.2	7.0	8.1	9.5	8.6	6.0
exp	17.7	11.8	9.8	8.3	6.6	11.3	10.0	6.7

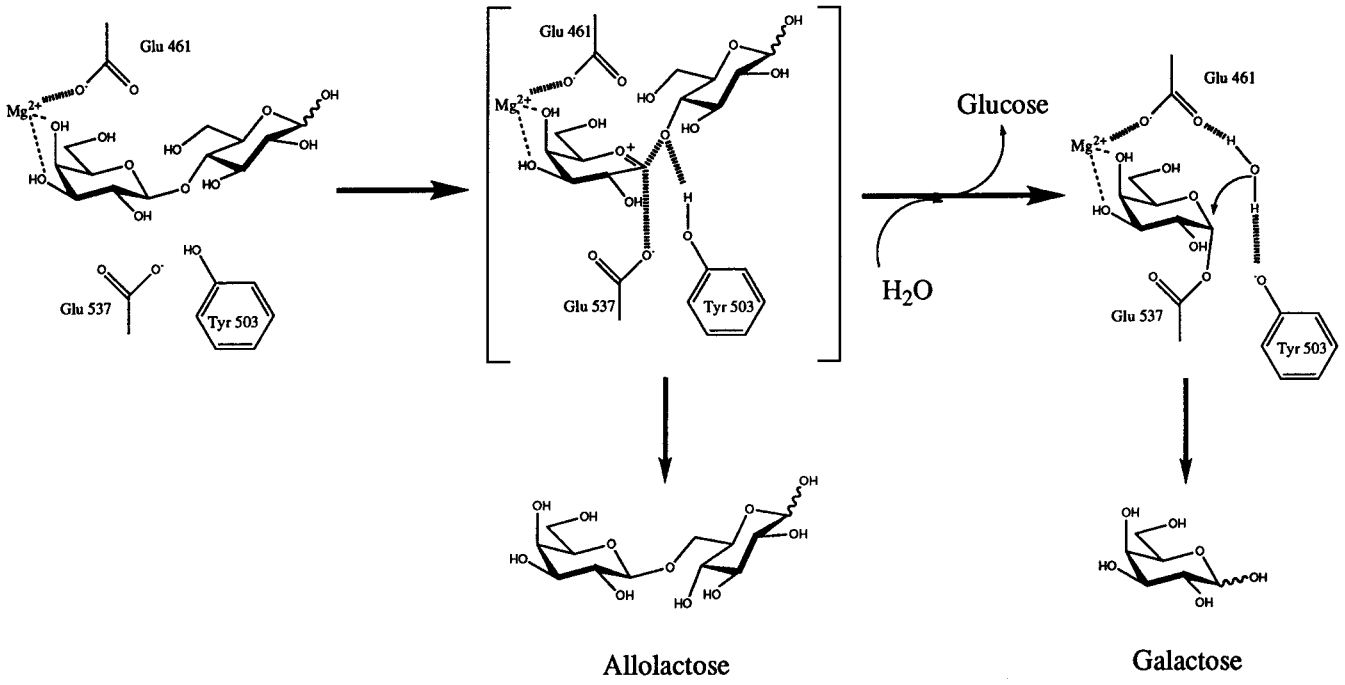


Figure 7. Scheme of the proposed mechanism for the hydrolysis of lactose in the gauche−gauche conformation by *E. coli β*-galactosidase. The role and relative orientation of the key amino acid residues as well as that of the Mg^{2+} ion is indicated. The formation of allolactose is also indicated.

performed following the protocol described in Materials and Methods. The results are shown in Table 2, producing a good match between experimental and theoretical NMR data. This analysis supports solution 2 as being that present in the complex.

Discussion

If it is assumed that the experimentally and theoretically deduced *E. coli β*-galactosidase-bound conformation of C-lactose is indeed significant with respect to the initial binding mode of lactose, the recognition of the gauche−gauche conformer may have important implications on the reaction mechanism.

As it has been mentioned above, the gauche−gauche conformer is a high-energy conformer, barely present in solution for C-lactose, and absent for lactose.[17−19] The preferential recognition of a high-energy structure of the substrate, whether close or in the pathway toward the transition state of the reaction, may decrease the activation energy of the reaction, thus accelerating the catalytic process.[26−28]

Nevertheless, using both solutions 1 and 2, it is possible to infer the role that each key amino acid residue plays in the hydrolytic mechanism, according to its orientation and distance with respect to the ligand (Figure 7). Indeed, the geometry of the substrate−enzyme complex, at least qualitatively, is consistent with multiple biochemical data described in the literature for *E. coli β*-galactosidase, and with the proposed roles for the different amino acid residues located around the binding site:[2] (a) Glu-537, which has been shown to be the nucleophile residue of this retaining enzyme,[9,29] is located in such a way that can approach the anomeric center from the α face of the galactose ring, as expected for a retention mechanism (Figure 6). In addition, it is positioned at hydrogen bond distance of O-2 (∼3.0 Å). In this context, it has been suggested that galactose HO-2 is important at some catalytic event and that could participate in hydrogen bonding with the nucleophilic carboxylate.[30,31] (b) His-540 is located at hydrogen-bond distance of galactose O6 (Figure 5, ∼2.0 Å), in agreement with recent data proposing this interaction by studying His-540 mutated enzymes.[11] (c) Glu-461 is a rather complicated case, since although it shows an important function in the catalytic process,[2,8,32] its exact role is not clear. The crystallographic structure shows Glu-461 to be coordinated to the magnesium ion.[3] After the docking protocol of galactose, Glu-461 appeared close to the anomeric region (∼4.5 Å), but approaching the *β* face of the galactose moiety. This location is not compatible with early proposals of this residue as the nucleophile.[33] In fact, the role of Glu-461 as the nucleophile in the hydrolysis mechanism was not completely clear from mutagenesis studies[34,35] and, according

(26) Jencks, W. P. *Adv. Enzymol.* **1975**, *43*, 219−410.
(27) Lightstone, F. C.; Bruice, T. C. *J. Am. Chem. Soc.* **1996**, *118*, 2595−2605.
(28) Cannon, W. R.; Singleton, S. F.; Benkovic, S. J. *Nature Struct. Biol.* **1996**, *3*, 821−833.
(29) Yuan, J.; Martinez-Bilbao, M.; Huber, R. E. *Biochem. J.* **1994**, *299*, 527−531.
(30) Bock, K.; Adelhorst, K. *Carbohydr. Res.* **1990**, *202*, 131−149.
(31) McCarter, J. D.; Adam, M. J.; Withers, S. G. *Biochem. J.* **1992** 721−727.
(32) Richard, J. P.; Huber, R. E.; Heo, C.; Amyes, T. L.; Lin, S. *Biochemistry* **1996**, *35*, 12387−12401.
(33) Herrchen, M.; Legler, G. *Eur. J. Biochem.* **1984**, *138*, 527−531.
(34) Bader, D. E.; Huber, R. E. *Biochem. Biophys. Res. Commun.* **1988**, *153*, 301−306.

to the newer data[9] it is now established that Glu-537 is the actual nucleophile. In this context, the role of Glu-461 has been recently revised by studying the properties of mutants at this position.[8,32] Thus, it has been shown that the E461G mutant is able to produce β-galactosyl azide or acetate adducts in the presence of high concentrations of those anionic nucleophiles. This result can be explained in structural terms, if the carboxyl side chain of Glu-461 is close to the β face of the galactose moiety in the galactosyl−enzyme intermediate, as deduced from our docking studies. In the E461G mutant, the external anionic nucleophiles can occupy the Glu-461 side chain location without significant repulsion and then attack the galactosyl residue from the β face. Thus, Glu-461 is probably involved in electrostatic stabilization of the cationic transition state as suggested by Huber[32] and confirmed by the loss of binding of positively charged transition state analogues after substitution of this amino acid.[2] (d) Tyr-503 appears in the crystallographyc structure[3] at hydrogen bond distance of Glu-537. It has been previously postulated that Tyr-503 could either modulate the p*K* of Glu-537 or act as the acid/base catalyst for the hydrolysis process.[2,10] In the docking protocol (Figures 5, and 6), the Tyr-503 hydroxyl is close to the interglycosidic atom (~3.2 Å), thus allowing protonation of the interglycosidic oxygen of the disaccharide. Recently, a similar role for a tyrosine residue has been proposed for the endocellulase E2 from *Thermomonospora fusca*, after modeling of the corresponding enzyme−substrate complex.[36] Moreover, in our 3D model, Tyr-503 is located at the same face as Glu-537 relative to the galactose ring, while in other glycosidases, the acid/base-catalysts, frequently Asp or Glu residues, are on the opposite face of the nucleophile/base amino acid.[6,7] Interestingly, for *E. coli* β-galactosidase, the binding of the substrate in the gauche−gauche conformation (Figures 5 and 6) would allow the interaction of the interglycosidic oxygen with the hydroxyl group of Tyr-503 without any steric hindrance caused by the presence of the aglycon moiety. As a matter of fact, a gauche−gauche-type conformation has also been proposed for the scissile bond in the binding of a pentaglucoside substrate to the endoglucanase CelA, based on the crystallographic structures of cellobiose-enzyme complexes.[37] In this inverting glycosidase, the best candidate for acting as the base that activates the nucleophilic water is Asp-152, which approaches the substrate from the same side as the proton donor Glu-95.[37] (e) An additional structurally relevant observation may be extracted from the modeling of C-lactose at the enzyme active site, and concerns to Trp-568 and Trp-999 (Figures 5 and 6). Both galactose and glucose pyranose rings, in the gauche−gauche orientation, expose their hydrophobic α-faces to the indole rings of Trp-568 and Trp-999, respectively. The stacking of hydrophobic surfaces of carbohydrates to Trp residues is a recurrent motif for oligosaccharide-binding proteins, both in the solid state and in solution.[23,24] (f) Finally, a Mg^{2+} ion has been located in the putative active site pocket,[3] coordinated to Glu-461 and His-418. Although this cation is not essential and is not directly implicated in the catalytic events,[1] it has been shown to be relevant for optimal performance of the enzyme. In this context, TR-NOE experiments recorded for samples prepared in a Mg^{2+}-containing buffer did not show any difference to those carried out without added Mg^{2+}. In both cases, only the exclusive NOEs for the gauche−gauche conformation were observed. Early NMR studies of complexes of substrate analogues and β-galactosidase, in the presence of Mn^{2+} (instead of Mg^{2+}), allowed the estimation that the cation was located 8−9 Å far from the aglycon moiety.[38] Our model structure displays the aglycon protons at a distance over 7 Å from the Mg^{2+} ion, and additionally, presents interactions between this cation and the axial Gal 4-OH and equatorial 3-OH hydroxyl groups. Both interactions are clearly compatible to those early studies[38] and, in addition, could explain the enzyme selectivity for terminal galactose-containing disaccharides.

The observed bound gauche−gauche conformation of C-lactose could also explain the transglycosidation properties of *E. coli* β-galactosidase. It is well established that the retaining mechanism goes through a glycosyl-enzyme reaction intermediate,[1] which allows the transfer of glycosyl residues to other nucleophiles different than water. *E. coli* β-galactosidase in the presence of lactose synthesizes allolactose (Figure 7), the natural inducer of the Lac operon.[39] However, it has been shown that this secondary reaction occurs without the need for the aglycon to leave the active site of the enzyme.[40,41] Interestingly, only the gauche−gauche conformation of the disaccharide places the OH-6 group of glucose on the β face of galactose approaching the anomeric C-1 carbon (<4 Å). This is indeed a favorable situation for the formation of allolactose. Thus, it is tempting to speculate that the binding of the gauche−gauche conformation would facilitate the internal transfer of the galactosyl residue from O-4 to O-6 of glucose with just minor movements of the glucose aglycon inside the active site.

As indicated above, there is not a direct way to demonstrate that the conformations of C-lactose and lactose itself when bound to the enzyme are the same. However, this possibility deserves some comment, since structural, biochemical, and modeling studies point to this possibility. C-lactose and lactose display minor geometric and structural differences according to MM3* calculations (C1′−Cα, 1.538 vs C′1−O1′, 1.426 Å and O5′−C1′−Cα 107.4° vs O5′−C1′−O1′, 108.3°). In addition, the inhibition constant determined for β-methyl C-lactoside (3.3 mM) and C-lactose (2 mM) are of the same order of the K_i for lactose itself (1.2 mM).[11] Besides, preliminary modeling studies with lactose at the *E. coli* β-galactosidase active site indicates that only the same gauche−gauche conformation may be bound without severe steric conflicts. However, it should be kept in mind that modeling of a substrate molecule at the enzyme active site is a difficult task, since enzymatic catalysis is a dynamic process that may involve substrate deformation.[26−28]

It is noteworthy to consider whether the observed conformation of C-lactose is reminiscent of the deformations in the reaction pathway. Generally speaking, glycosyl oxocarbonium-like transition state structures have been proposed for glycosidases.[1,5] For disaccharides, the variation from a normal chair conformation at the nonreducing end toward the expected half chair or sofa conformations for a glycosyl oxocarbonium, produces a significant upward shift of the aglycon moiety with respect to the plane defined by C-5, O-5, C-2, and C-3 in the initial $^4C_1$ chair of the nonreducing end. Several studies have recently addressed this topic in structural terms. Sofa conformations at the glycon part have been already observed in

(35) Cupples, C. G.; Miller, J. H.; Huber, R. E. *J. Biol. Chem.* **1990**, *265*, 5512−5518.
(36) Taylor, J. S.; Teo, B.; Wilson, D. B.; Brady, J. W. *Protein Eng.* **1995**, *8*, 1145−1152.
(37) Alzari, P. M.; Souchon, H.; Dominguez, R. *Structure* **1996**, *4*, 265−275.
(38) Loeffler, R. S.; Sinnott, M. L.; Sykes, B. D.; Withers, S. G. *Biochem. J.* **1979**, *177*, 145−152.
(39) Müller-Hill, B.; Rickenberg, H. V.; Wallenfels, K. *J. Mol. Biol.* **1964**, *10*, 303−318.
(40) Huber, R. E.; Kurz, G.; Wallenfels, K. *Biochemistry* **1976**, *15*, 1994−2001.
(41) Adelhorst, K.; Bock, K. *Acta Chem. Scand.* **1992**, *46*, 1114−1121.

crystallographic studies of HEWL mutant lysozyme complexed with oligosaccharide reaction products.[42] A similar case occurred in a covalent adduct of a mutated T4 lysozyme.[43] More interestingly, this conformational deformation has also been directly observed in the X-ray three-dimensional structure recently determined for a bacterial chitobiase complexed with its substrate, diacetyl chitobiose.[44] A related binding-induced conformational chair to boat change has also been observed for *Fusarium oxysporum* endoglucanase I complexed with a nonhydrolyzable substrate analogue.[45] In both cases, the conformational changes of the residues attached to the scissile bond place the aglycon moiety of the substrate in a pseudoaxial orientation, as preferred for a leaving group.[44,45] Going back to our case, β-galactosidase/C-lactose, if the galactose moiety would adopt a half chair or sofa conformations, H-1′ proton should move away from H-3′ and H-5′ atoms with the consequent decrease in the intensities of the corresponding H-1′/H-3′ or H-1′/H-5′ intraresidue NOEs. However, since the experimental intraresidue NOEs are strong, it can be excluded that the galactopyranose ring of C-lactose suffers a significant conformational change. However, it cannot be excluded that the observed gauche−gauche conformation of C-lactose could be a gross mimic of the natural substrate geometry at the enzyme active site.

Conclusions

As a final comment we stress that the methodology described herein (based on TRNOE and molecular modeling) represents a novel way to study the conformation and orientation of nonhydrolizable substrate analogues at the enzyme active site. Whether the results obtained in this way may be directly extrapolated to the natural substrate still remains an open question. Nevertheless, this approach can give new clues to the understanding of enzyme mechanism, by complementing other structural studies where the structures of other enzyme−substrate or enzyme−inhibitor complexes are not easily accessible. The TRNOE technique has an applicability window limited by the kinetics underneath the receptor−ligand equilibrium.[13] However, it seems that the study of carbohydrate−enzyme interactions is an area where this methodology can have a remarkable application. The reason of this favorable situation probably lays in different facts: These interactions are not extremely strong, there is fast exchange between the free and the bound states of the ligand, and the perturbations of the conformational equilibrium of a given oligosaccharide upon binding to a protein are accessible to observation by TRNOE. We have applied this approach for the first time to study a glycosidase enzyme. To know if this methodology can be of general use, studies with other glycosidase enzymes and similar nonhydrolizable substrates, together with the reverse approach of studying the conformation of natural substrates in the presence of mutated and nonactive enzymes are now in progress.

Methods

Source of the Enzyme. The engineered enzyme used in this work was obtained by overproduction of a fusion protein of β-galactosidase and the C-terminal choline binding domain (C-LYTA) of the amidase LYTA from *Streptococcus pneumoniae* coded in plasmid pEG40.[46] The protein was purified as described.[15] The concentration of the chimeric protein was estimated by measuring the optical density at 280 nm (0.4 mg/mL). The obtained enzyme was over 95% pure, its activity assayed using ONPG as substrate, and gave a specific activity of 300 mmol min^{-1} of *o*-nitrophenol per milligram of protein at 301 K and pH 7. The obtained K_m for ONPG (0.1 mM) and K_i for IPTG (0.1 mM) were the same as those obtained when the wild-type enzyme was used.

Inhibition Assays. The engineered enzyme (0.05 mg/mL) was incubated in sodium phosphate buffer (50 mM, pH = 7.2) with the substrate ONPG (0.5 mM) in the presence of different concentrations (0, 4, 8, 16, 32 mM) of inhibitor (methyl β-C-lactoside). The reaction was monitored by HPLC. After 5 min, incubation at 310 K, the samples (20 mL) were diluted with 200 mL of cold HPLC elution buffer (acetonitrile 34%, TFA 0.1% in water). The diluted samples were injected in the HPLC system (reverse-phase C18 column, UV detection). The *o*-nitrophenol produced was quantified from the UV absorbance at 320 nm. Kinetic data were adjusted to a competitive inhibition model in order to obtain the corresponding K_i. Similar inhibition experiments were performed with isopropyl β-thiogalactoside (IPTG) and methyl β-lactoside.

Molecular Modeling. Protein coordinates were taken from the published crystal structure of *E. coli* β-galactosidase.[3] C-Lactose was built using the SYBYL 6.2 program (Tripos Associates, St. Louis, MO) from the 3D structures of galactose and glucose residues. Atomic charges were calculated using the MNDO semiempirical program.

The GRID program[21] was used to predict the most favorable anchoring position for a galactose residue. Both a OH (hydroxyl) probe and a CH_2 (methylene) probe were used in the calculations. The grid spacing was set to 0.5 Å, and the calculations were limited to a box with sides of 40 Å around the binding site. Within the binding site, four oxygen and three carbons atoms corresponded to the most favorable sugar/protein interaction. Several starting orientations of the galactose residue were calculated by fitting the carbohydrate atoms with the GRID atoms.

Energy Calculations. A hot region, close to the protein active site, was defined by 15 amino acids, namely Asn-102, Asp-201, Asn-355, His-357, His-391, His-418, Asn-460, Glu-461, Met-502, Tyr-503, Glu-537, His-540, Trp-568, Asn-604, and Trp-999. The magnesium ion was also included in the hot region. A cold region was also defined, including 171 amino acids in a 8 Å sphere around the hot region. Hydrogen atoms were added on the amino acids of both hot and cold regions and atomic charges were calculated using the Pullman method.[47] All energy calculations were done using the Tripos force field with inclusion of parameters specially developed for carbohydrates.[20] For C-lactose, systematic conformational analysis was conducted by varying the torsional angles Φ and Ψ while keeping the galactose moiety fixed. Energy minimizations were conducted in several cycles. All the atoms of the cold and hot regions and of the carbohydrate were taken into account. In the first run, only the hydrogen atoms were allowed to optimized, then the carbohydrate atoms and the side chains of the hot region, and finally all the hot region atoms. Final minimizations were carried on until a rms gradient of 0.005 was achieved.

NMR Experiments. NMR spectra were recorded at 30 °C in D_2O, on a Varian Unity 500 spectrometer. Enzymes samples in sodium phosphate buffer (50 mM, pH = 7.2) were exposed

(42) Hadfield, A. T.; et al. *J. Mol. Biol.* **1994**, *243*, 856−872.
(43) Kuroki, R.; Weaver, L. H.; Matthews, B. W. *Science* **1993**, *262*, 2030−2033.
(44) Tews, I.; et al. *Nature Struct. Biol.* **1996**, *3*, 638−648.
(45) Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schülein, M.; Davies, G. J. *Biochemistry* **1996**, *35*, 15280−15287.
(46) Garcia, E.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 914−918.
(47) Bertod, H.; Pullman, A. *J. Chem. Phys.* **1965**, *62*, 942−945.

to repeated cycles of freeze-drying with D_2O, and transferred to the NMR tube to give a final concentration of 0.037 mM. The stability of the enzyme solutions was checked by measuring protein concentration and enzymatic activity before and after the NMR experiments. TR-NOESY experiments were performed with mixing times of 200 and 300 ms, for a 24/1 molar ratio of β-methyl C-lactoside/enzyme. Additional experiments were carried out in the presence of 1 mM $MgCl_2$. Independent TR-NOESYs were also recorded after adding equimolecular amounts of IPTG (isopropyl β-D-thiogalactopiranoside). In all cases, line broadening of the sugar protons was monitored after the addition of the ligand. The theoretical analysis of the TRNOEs was performed according to the full relaxation matrix as described.[48] Different exchange-rate constants, k, defined as pf $\times$ $k = K_{-1}$ (where pf is the fraction of the free ligand) and leakage relaxation times were employed to get the best match between experimental and theoretical results for the given protein/ligand ratio. Normalized intensity values were used since they allow for correction of spin relaxation effects. The overall correlation time τ_c for the free state was always set to 0.15 ns[17] and the τ_c for the bound state was estimated as 500 ns according to the high molecular weight of the enzyme ($\tau_c = 10^{-12}$ W_M). The association constant was approximated as the inverse of K_i ($K_a = 303$ M^{-1}). To fit the experimental NOE intensities, exchange-rate constants between 200 and 1000 s^{-1} and external relaxation times for the bound state of 0.5, 1, and 2 s were tested. The best agreement was achieved when using $k = 400$ s^{-1} and $r^* = 1$ s^{-1}.

TR-ROESY experiments[16] were also carried out to exclude spin difussion effects. A continuous wave spin lock pulse was used during the 250 ms mixing time. Key NOEs were shown to be direct cross peaks, since they showed a sign different to diagonal peaks.

Acknowledgment. We thank DGICYT (PB96-0833) and the Mizutani Glycoscience Foundation for financial support. Part of this work has been funded by the Human Capital and Mobility Program, CARENET 1. J.F.E. thanks MEC for a fellowship.

Supporting Information Available: Expansion of the key region of the NOESY spectrum for free C-lactose at 303 K, and for the TR-NOESY spectrum for C-lactose bound to *E. coli* β-galactosidase molar ratio 24:1 under the same experimental conditions (relevant cross peaks are indicated); stereoviews of the *syn* and *anti* conformers of C-lactose; exclusive NOEs for the different conformers of C-lactose; and TR-ROESY spectrum for C-lactose bound to *E. coli* β-galactosidase molar ratio 24:1 (relevant cross peaks are indicated) (6 pages). See any current masthead page for ordering information and Web access instructions.

JA972291Q

(48) London, R. E.; Perlman, M. E.; Davis, D. G. *J. Magn. Reson.* **1992**, *97*, 79−98.