Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 339 (2004) 995–1006
Carbohydrate RESEARCH

Molecular dynamics simulations of glycosyltransferase LgtC ☆

Lenka Šnajdrová,[a,b] Petr Kulhánek,[a,c] Anne Imberty[b] and Jaroslav Koča[a,c,*]

[a]*National Centre for Biomolecular Research and Department of Organic Chemistry, Faculty of Science, Masaryk University Brno, CZ-611 37 Brno, Czech Republic*

[b]*Centre de Recherches sur les Macromolécules Végétales, CNRS and Université Joseph Fourier, IFR 2607, BP 53, F-38041 Grenoble, France*

[c]*Department of Organic Chemistry, Faculty of Science, Masaryk University Brno, CZ-611 37 Brno, Czech Republic*

Received 13 October 2003; accepted 17 December 2003

Abstract—Molecular dynamics simulations have been performed on fully solvated α-(1→4)-galactosyltransferase LgtC from *Neisseria meningitidis* with and without the donor substrate UDP-Gal and in the presence of the manganese ion. The analysis of the trajectories revealed a limited movement in the loop X (residues 75–80) and a larger conformational change in the loop Y (residues 246–251) in the simulation, when UDP-Gal was not present. In this case, the loops X and Y open by almost 10 Å, exposing the active site to the solvent. The 'hinge region' responsible for the opening is composed of residues 246–247. We have also analyzed the behavior of the manganese ion in the simulations. The coordination number is 6 when UDP-Gal is present and it increases to 7 when it is absent. In the latter case, three water molecules become coordinated to the ion. In both cases, the coordination is very stable implying that the manganese ion is tightly bound in the active site of the enzyme even if UDP-Gal is not present. Further analysis of the structural water molecules location confirmed that the mobility of water molecules in the active site and the accessibility of this site for solvent are higher in the absence of the substrate.

Keywords: Galactosyltransferase; Molecular dynamics; Loops opening; Structural water molecules

1. Introduction

The glycosyltransferases catalyze the transfer of glycosyl moieties from a donor sugar to an acceptor. In most cases, the donor is a nucleoside phosphosugar and the acceptor is a hydroxyl group of another sugar, a lipid, or another component of glycoconjugates. The glycosyltransferases are classified as either retaining or inverting, depending on the stereochemical outcome of the reaction catalyzed.

The galactosyltransferase LgtC from *Neisseria meningitidis* is a retaining α-(1→4)-glycosyltransferase, catalyzing a key step in the biosynthesis of lipooligosaccharide structures by transferring α-D-galactose from UDP-galactose to a terminal lactose (Scheme 1).[1] The resulting structure is a mimic of the human P^k blood group glycolipid and can therefore participate in the camouflage of the bacterial surface from recognition by the human immune system. Since *Neisseria meningitidis* can act as a human pathogen, causing an invasive systemic infection that can be fatal,[2] understanding of the substrate recognition and mechanism of LgtC is of interest for future design of inhibitors that could block lipooligosaccharide biosynthesis.

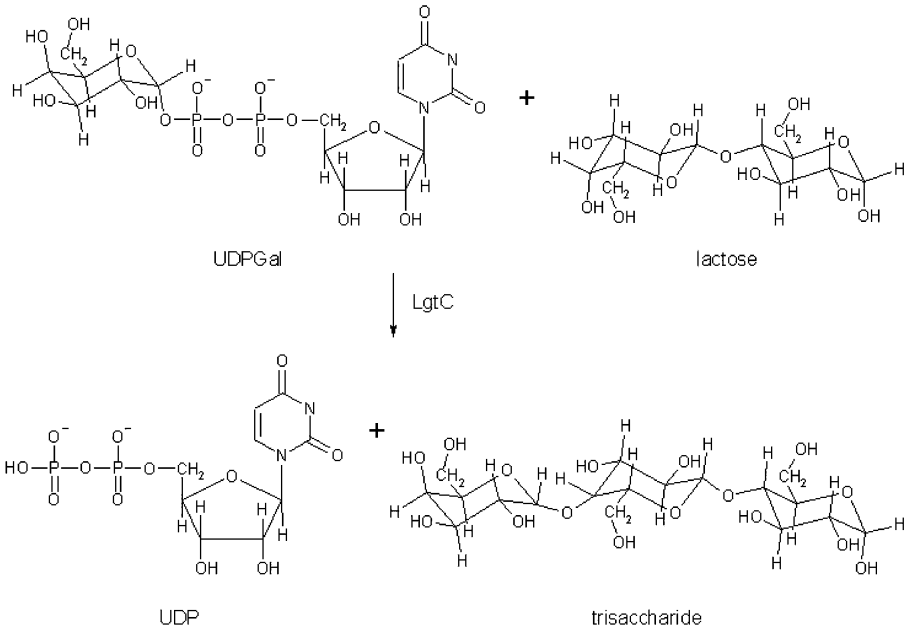
The crystal structure of LgtC has been recently solved as a complex with analogs of both donor and acceptor.[3] The enzyme is a monomer composed of 286 residues that create nine β-strands and 14 α-helices (Fig. 1). The active site is mostly buried under two loops that presumably play a key role in organizing the substrates before the reaction and probably also in removing the products

Abbreviations: RMSD, Root mean squares deviations; EDA, Essential dynamics analysis.

☆ Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.carres.2003.12.024

* Corresponding author. Tel.: +420-541-129310; fax: +420-541-129506; e-mail: jkoca@chemi.muni.cz

0008-6215/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2003.12.024



Scheme 1. Schematic representation of the LgtC catalyzed transfer of galactose from UDPGal to lactose. The product retains the configuration of the donor sugar glycosidic bond; LgtC is thus a retaining glycosyltransferase.

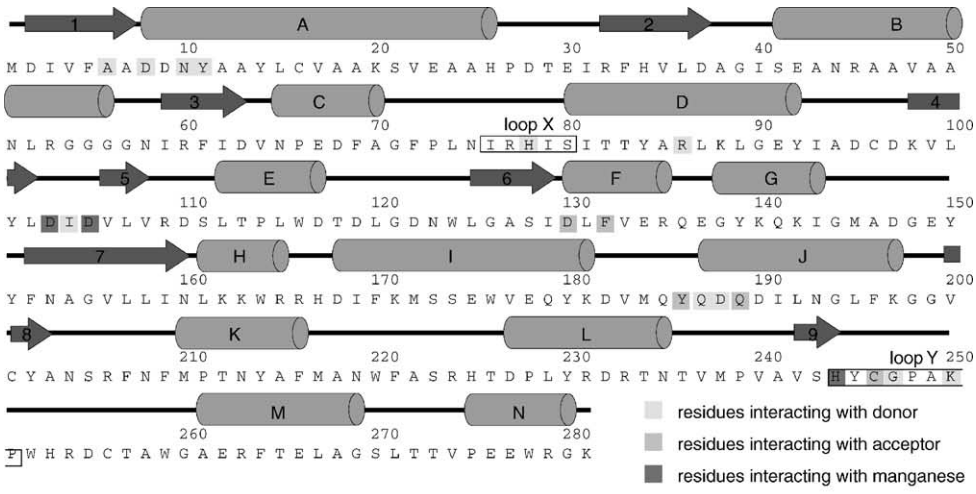
from the active site. Loop X (residues 75–80) makes a connection between two helices. Loop Y consists of residues 246–251 and is located in the C-terminus region of the enzyme. The two loops are on opposite sides of the active site groove. In the complex with substrate analogs, the two loops close above the donor sugar by a stacking interaction between Pro248 and His78.[3] There is no available structural data about the conformation of the loops in the native enzyme, since the absence of UDP-Gal results in inability to crystallize the enzyme.

The structure of LgtC is a key starting point to fully understand the mechanism of the enzymatic reaction. The understanding of the catalytic mechanism is an essential precondition for rational design of effective inhibitors of this enzyme. Although the structural details of the complex are well characterized,[3] a subsequent detailed mechanistic study could not clearly delineate the chemical mechanism of this enzyme.[4] Extensive structural studies have been performed on another family of retaining glycosyltransferases that contains bovine α-(1 $\rightarrow$ 3)-galactosyltransferase[5–7] and human blood group A and B glycosyltransferases.[8,9] From the comparison of the different structures, it could be established that in this family the presence of UDP-Gal is necessary for the conformational change of two loops, and the closed conformation could be modeled.[10] A first theoretical approach of catalytic reaction has also been proposed for α-(1 $\rightarrow$ 3)-galactosyltransferase,[11] but the proposed mechanism has not yet been validated by experimental work and is still discussed. Open questions are related not only to the enzymatic reaction itself, but also to the role of the enzyme in arranging the starting material to react and the products of the reaction to depart from the active site. For example, no detailed information is known about the flexibility of loops X and Y and their role in attaching the manganese ion as well as in the donor and acceptor substrate anchoring.

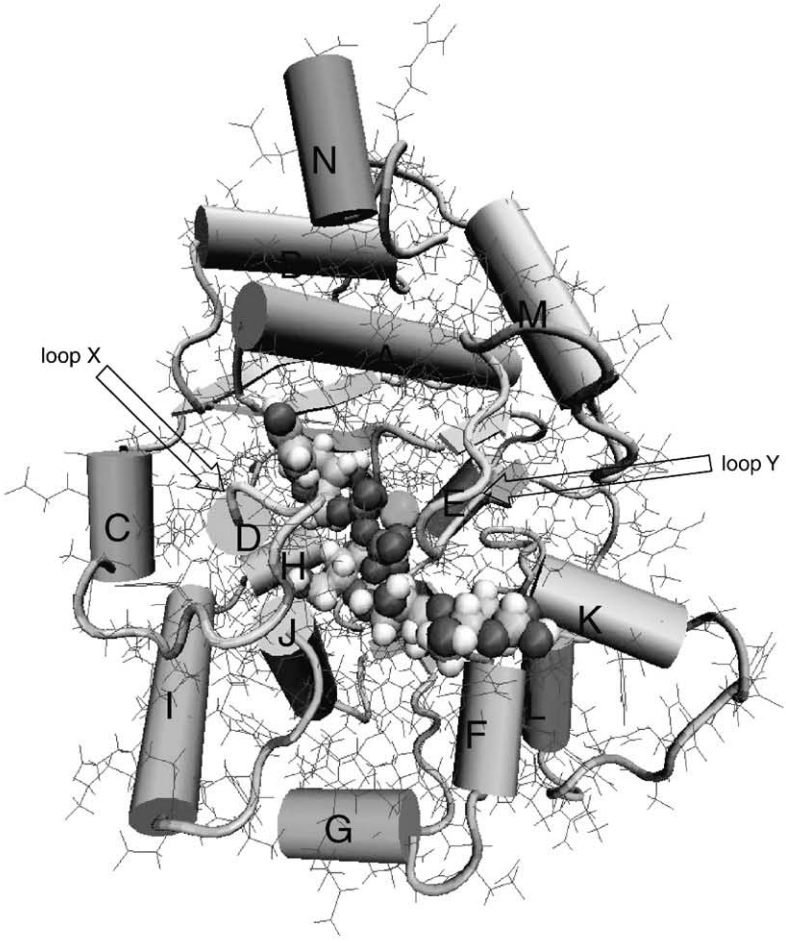
Because of the structural and mechanistic questions and problems mentioned above, we decided to use theoretical methods, namely molecular dynamics, to address the issue. Molecular dynamics has often proved to serve as a reliable computer simulation method that may usefully complement experimental observations. Very recently, it has also been used to analyze the backbone flexibility of an inverting glycosyltransferase.[12] We report here the first molecular dynamics simulations of a retaining glycosyltransferase, LgtC, in water environment and in the presence of its donor substrate. The analyses of the trajectories provide an insight into atomic interactions within the active site of this glycosyltransferase as well as into the flexibility of the active site. The aim of this work is to analyze the active site loops flexibility, how it is influenced by the presence of the UDP-Gal substrate, and to determine the role of Mn^{2+} ions. As a spin-off, we will show how the protonation state of His244 may influence the behavior of the loops and how this may produce artifacts in the calculations.

2. Results and discussion

Since kinetics studies suggested that UDP-Gal binds first to LgtC, followed by the acceptor lactose,[4] we focus here on this first step and analyze the stability of the complex without the acceptor. Three molecular



(a)



(b)

Figure 1. Structural informations from the crystal structure of LgtC complexed with substrate analogs.[3] Sequence information together with secondary structures (a), 3D-structure with schematic representation of α-helices and β-sheets (b). Donor, acceptor, and manganese ions are represented as CPK.

dynamics simulations of LgtC in explicit solvent were carried out in order to explore the dynamic behavior of LgtC (Table 1). Conditions of simulation were chosen in order to study the stability of the protein loops as a

Table 1. Molecular dynamics simulations on LgtC

Simulation	Composition	Purpose	Number of water molecules[a]
I	LgtC + Mn^{2+} + UDP-Gal, δ-coordinated His244	To explore the effect of different protonation states of His244	11,437
II	LgtC + Mn^{2+} + UDP-Gal, ε-coordinated His244	To analyze flexibility of the structure with the donor substrate	11,425
III	LgtC + Mn^{2+}	To identify flexible regions and the role of Mn^{2+} in binding	11,440

[a]Simulation box size before equilibration was $75\times78\times82$ Å^3 in all three cases.

function of the presence of UDP-Gal and the role of Mn^{2+} in UDP-Gal binding. Furthermore, the hydration was thoroughly analyzed.

2.1. Influence of different tautomeric forms of His244 on LgtC structure

The X-ray structure of LgtC was used as a source of initial coordinates for all molecular dynamics simulations. Since the resolution of the X-ray was not adequate enough to determine location of hydrogen atoms, they have to be added in theoretical positions. This may not be straightforward and simple as several residues can exist in more than one protonation state depending on pH, ionic strength etc. Following generally used convention for simulations in water solution, all Asp, Glu, Lys, and Arg residues were considered in ionic forms in all simulations and His residues were considered in neutral forms. These are represented by two tautomers in which the hydrogen atom is located either in ε- or in δ-position. In our calculations, each His residues were assigned such tautomeric form that the hydrogen bond network within the entire protein was optimal. This was only problematic for His244 as it is also the integral part of Mn^{2+} ion binding site and the structure of this site can be influenced by His244 isomeric form.

His244 coordinates Mn^{2+} ion via the nitrogen atom in ε-position in both X-ray structures of LgtC. However, the situation in free LgtC is not known because this structure has not been solved yet. It is therefore possible that His244 is δ-coordinated. Simulations II and I with differently coordinated His244 were run to bring more light into this problem. When the nitrogen in δ-position coordinates the manganese ion (simulation I), a strong repulsion between the hydrogen in ε-position and the ion occurs. As a consequence, the entire side chain of this amino acid residue turns around during the initial minimization so that the δ-nitrogen atom of His244 can coordinate the manganese ion. This causes a significant change of conformation of the part of the backbone related to this residue, which subsequently results in the opening of loops X and Y during the simulation. Such a situation is not in agreement with the crystal structure,[3] and furthermore would probably cause hydrolysis of UDP-Gal prior to the enzymatic reaction. On the contrary, when the nitrogen in ε-position coordinates the Mn^{2+} ion (simulation II), His244 remains in contact with the ion, as observed experimentally, throughout the simulation. This latter situation is therefore chosen as the most probable one, and is kept for the simulations analyzed in the present work (simulations II and III). It follows from both simulations that UDP-Gal probably binds to the active site with ε-coordinated His244. Nevertheless, the simulations do not exclude the second possibility (δ-coordinated His244) but with His244 coordination type changing during loops X and Y closing (see below). However, this later case is less probable.

2.2. Effect of the presence of UDP-Gal in the binding site

Stability of the trajectories—Simulations in the presence (II) and in the absence (III) of UDP-Gal were calculated with total production length of 3 ns for each of them. In both cases, potential energy of the system became stable during the first 100 ps of the simulation. When looking at the stabilities of the geometry, simulation II needed almost 2 ns to reach the full stability whereas trajectory III became fully stable after about 1 ns (Fig. 2a). Analysis of the flexibility along the peptide chain can be performed by calculating theoretical B-factors[13] and comparing them with the ones determined by X-ray crystallography. Both simulations II and III display an excellent agreement between calculated B-factors and the experimental ones (Fig. 2b). Small differences in some loop regions are discussed below.

Active site opening and the loops flexibility—Essential dynamics analysis (EDA) was used to identify correlated atomic motions occurring during simulations II and III. EDA is able to filter out large correlated motions in the simulated system from uninteresting local motions (mostly vibrations). Overall translational and rotational motions were removed before EDA was started. EDA itself consists of the following steps. First, a covariance matrix is built from the atomic fluctuation in the trajectory. Only C-alpha atoms have been included since it has been shown that these atoms contain all the information for a reasonable description of the concerted motion.[14] Upon matrix diagonalization, a set of eigenvalues and eigenvectors is obtained. The eigenvectors correspond to directions in space, and motions along single eigenvectors correspond to concerted fluctuation of atoms. The eigenvalues then represent the total mean

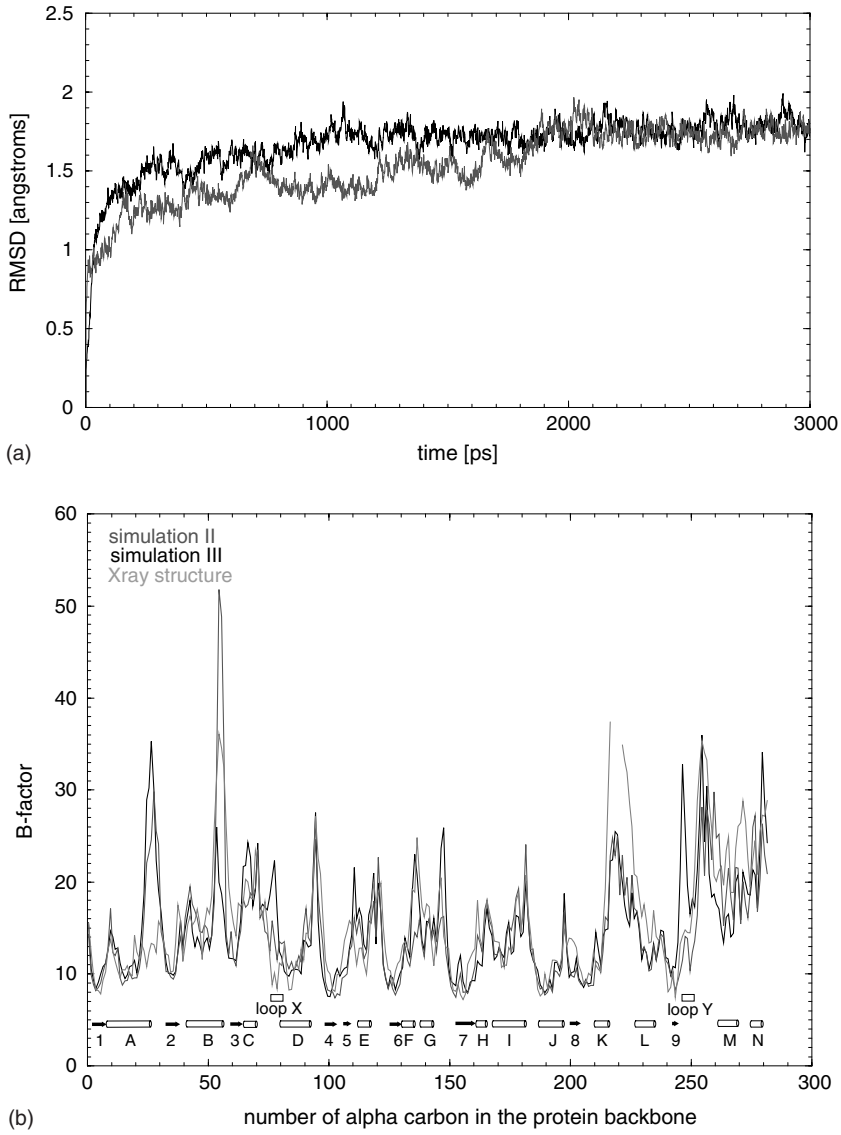


Figure 2. Mass-weighted RMSD (in Å) from the starting structure for simulations II and III (a), and B-factors of C_α-atoms of the backbone (b) taken from X-ray data or calculated from MD simulations.

square fluctuation of the system along the corresponding eigenvectors. In the last step, new trajectory containing only correlated atomic motions is calculated from the old trajectory by applying the most significant eigenvectors. These, as Berendsen and coworkers demonstrated,[14,15] correspond to only the early largest eigenvalues, usually between 5 and 10.

Root mean square deviation (RMSD) for each C-alpha atom calculated from such filtered trajectories (10 eigenvectors were used) are shown in Figure 3. In both simulations, as expected, loops display more motions than secondary structure elements do. As an overall observation, the simulation in absence of UDP-Gal displays larger motion of both loops X and Y than the simulation in the presence of this substrate. A significant difference is observed in the C-terminal region, and especially in the loop Y. Indeed, when looking at the crystal structure, this is one of the loops that cover the binding site, together with the loop X, located in the N-terminal region. The latter loop X also displays slightly increased motion for simulation III in Figure 3. Interestingly, there are some parts of the protein that exhibit larger motions when the UDP-Gal is present than when it is absent. All of the residues concerned are distant from the active site. This behavior is explained by a high mobility of side chains and by their tendency to form ionic pairs with another side chain (e.g., Lys20 and Glu23, Arg53 and Glu23, Arg224 and Glu137, and so on) and does not directly depend on the presence of donor substrate.

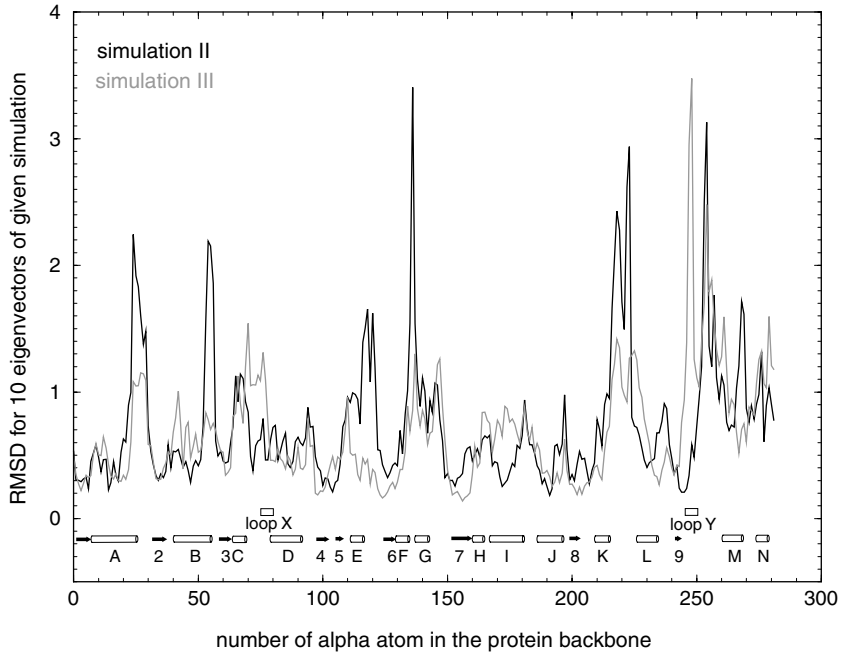


Figure 3. Results of essential dynamics analysis. RMSD (Å) calculated from the first 10 eigenvectors (simulation II in black and simulation III in light grey). Simulation III displays larger motion of both loops X and Y than simulation II.

Loops opening in simulation III—The history of these two loops has been analyzed. The inspection of the trajectories reveals that loops X and Y behave differently in simulations II and III. During simulation III, loops X and Y were observed to move away from each other, which results in opening of the binding pocket and exposing the active site to the solvent. This did not occur in simulation II, where the donor substrate was bound. This is illustrated by measuring the closest distance of both loops (Fig. 4). It is seen that the loops open by almost 10 Å measured by the distance of His78 and Pro248 alpha carbon atoms.

There are two key interactions that keep loops X and Y close to each other in simulation II: First, the presence of alternating hydrogen bonds between atom NE2 of His78 of the loop X and oxygen atoms OP2 and OP6 of pyrophosphate of UDP-Gal; and second, the stacking interaction between the rings of His78 and Pro248. During simulation III (UDP-Gal not being present in the active site), His78 cannot obviously form any hydrogen bond with donor substrate and the stacking interaction between His78 and Pro248 appears to be not strong enough to keep the loops together.

Dihedral angles over CA atoms in both loops X and Y and its surrounding were monitored and it was confirmed that relatively large changes occur in loop Y, whereas only small changes were observed in loop X. Further analysis revealed that there are two kinds of movement seen α in loop Y. The first is a 'hinge' movement of the entire loop. The hinge is composed of the residues Cys246-Gly247. The movement is best represented by pseudodihedral angle Cα245-Cα246-Cα247-Cα248 change. The second is an internal movement within loop Y itself (pseudodihedral Cα249-Cα250-Cα251-Cα252—see supplementary material, Fig. 1S).

Other flexible regions observed during the simulations—A certain number of loops that were stabilized by hydrogen bonds or ionic contact in the crystal structure exhibited a different conformational behavior during the molecular dynamics run in explicit solvent. The resulting local rearrangement explains the small differences in B-factors between simulations and X-ray that are displayed in Figure 2. Some of the observed movement present differences in simulation II and III: In the crystal structure, the C-terminal domain was bound to another part of the protein by a strong hydrogen bond involving the side chain of Ser270 and Lys20. In both simulations II and III, this contact is broken early during the MD simulation and the C-terminal domain does not keep its starting position. Furthermore, other flexible regions were observed in both simulations II and III. However, as it follows from the analyses of both trajectories, these changes do not influence behavior of the active site.

2.3. Coordination of the manganese ion

Manganese ion is required for both the stability and the activity of LgtC.[16] This is the case in glycosyltransferases adopting the same one domain fold (GT-A), in the sequence of which an acidic motif (DXD) is conserved and directly involved in coordination of the divalent ion.[17,18] The role of this divalent cation is to bridge the protein side chains with the diphosphate group of the

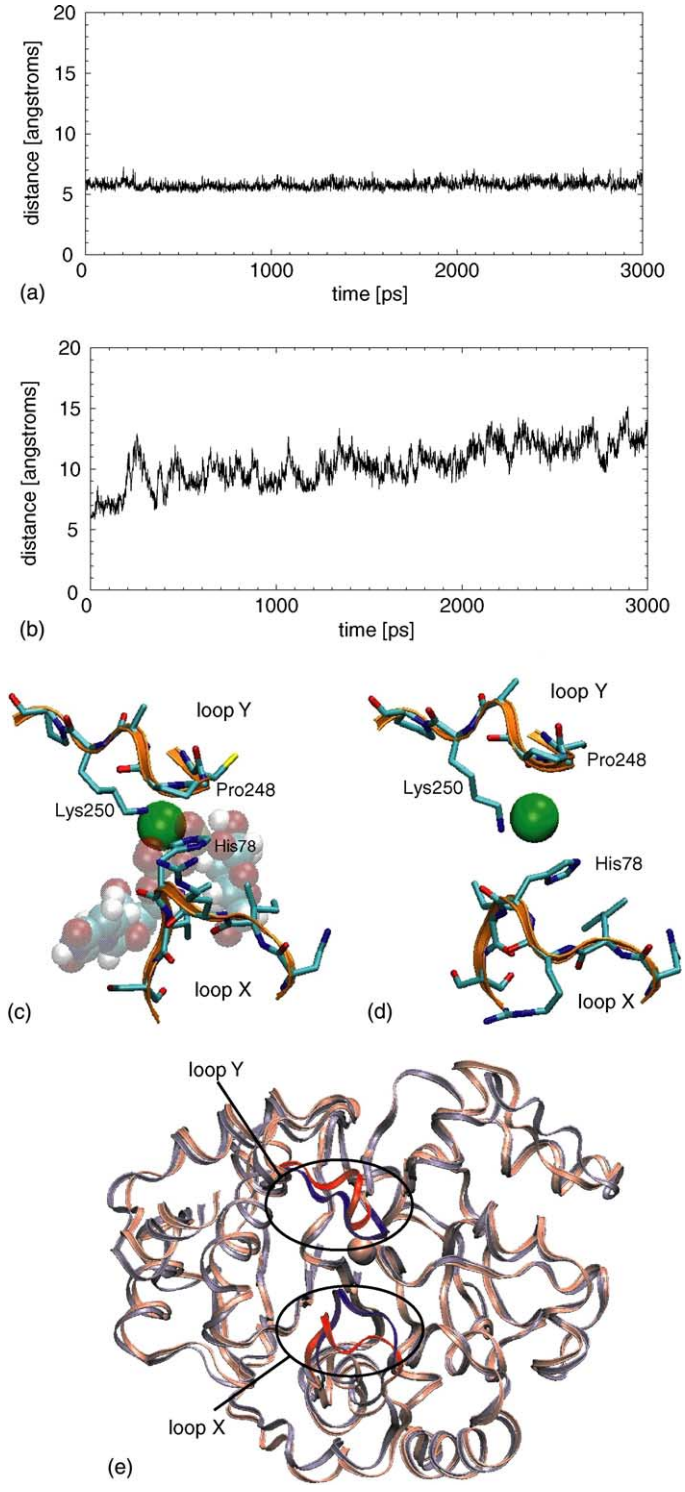


Figure 4. Distance between the α-carbons of His78 and Pro248 as a measure of the loops X and Y opening during simulations II (a) and III (b). Graphic representation of the system where the loops are closed (simulation II) (c) and open, with the active site exposed to solvent (end of simulation III) (d), and comparison of the two structures at the end of the MD simulations (e).

donor substrate. In the crystal structure of LgtC,[3] a Mn^{2+} ion is coordinated by two phosphate oxygen atoms of UDP as well as by the side chain atoms of Asp103 and Asp105, the two acidic residues forming the DXD motif. The question remains open as to whether the ion arrives into the active site prior to the UDP-Gal or together with it as a complex. We therefore ran the simulation III, where the UDP-Gal is not present, and compared the

behavior of the Mn^{2+} ion with that in simulation II. We focused our attention mainly on the coordination number of the ion during the two simulations.

In simulation II, the coordination number of manganese ion remains 6 over the whole trajectory involving both carboxylate oxygens of Aspl05, one carboxylate oxygen of Aspl03, imidazole nitrogen of His244 and two pyrophosphate oxygens of UDP-Gal (Fig. 5a). The average distance of the ion from its coordinating atoms oscillates between 2.1 and 2.4 Å (see Table 2), in a good agreement with the X-ray structure.

When the donor substrate is removed (simulation III), water molecules occupy the location of UDP and solvate Mn^{2+} ion. Radial distribution function (RDF) between the ion and water oxygens calculated from the last 1 ns of simulation III shows two solvation shells (Fig. 6). The first solvation shell, which is represented by sharp peak in the range from 2.1 to 2.6 Å and with the maximum at 2.3 Å, contains three water molecules. Zero value of RDF between the first and second peaks indicates that no exchange of water molecules between the first and second solvation shell occurred in the last 1 ns of simulation. Indeed, residence time of each of these three water molecules is 3 ns. The second solvation shell is represented by a broader peak of RDF in the range of 3.6–5.8 Å and with the maximum at 4.6 Å. It contains about 10 water molecules with residence times shorter than 1000 ps.

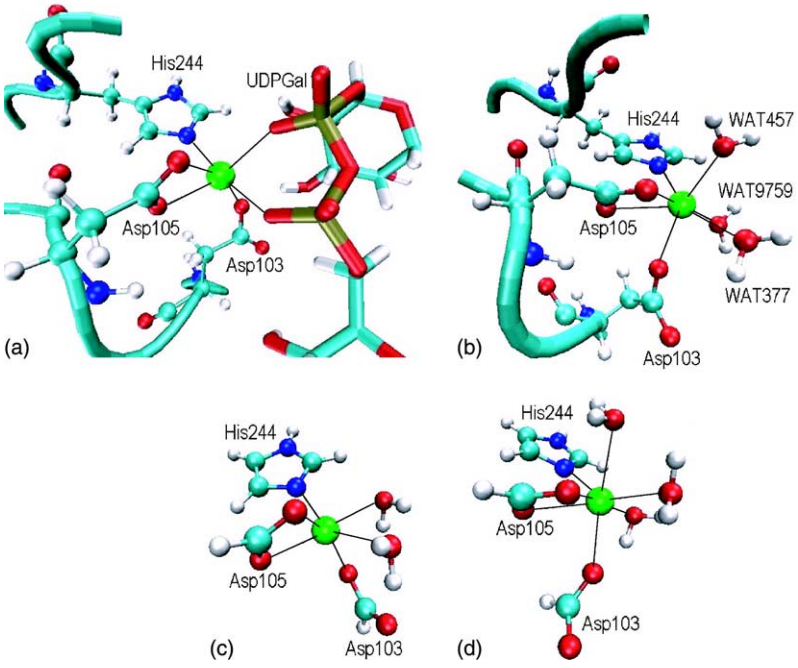


Figure 5. Coordination of manganese ion during MD (a and b) and QM (c and d) calculations. Manganese ion is hexa-coordinated in simulation II (a), and hepta-coordinated in simulation III (b). The energy-minimized structures using ab initio quantum mechanics show better agreement of the hepta-coordinated structure (d) with the MD data than with that of the hexacoordinated manganese ion (c).

Table 2. Coordination of the manganese ion in structure determined by X-ray, molecular dynamics simulations and quantum mechanics calculations (all distances in Å)

Atom	X-ray[a]	Simulation II		Simulation III		QM calculation	
	Distance from Mn^{2+}	Average distance from Mn^{2+}	Standard deviation	Average distance from Mn^{2+}	Standard deviation	Two water molecules	Three water molecules
NE2(His244)	2.28	2.37	0.07	2.44	0.09	2.19	2.48
OD1(Asp103)	2.24	2.17	0.05	2.18	0.05	2.08[b]	2.21
OD1(Asp105)	2.22	2.18	0.05	2.26	0.09	2.20	2.35
OD2(Asp105)	2.41	2.27	0.07	2.29	0.09	2.25	2.30
OP3(PYR285)	2.25	2.18	0.05	—	—	—	—
OP5(PYR285)	2.23	2.17	0.05	—	—	—	—
O(WAT377)	—	—	—	2.31	0.07	2.27	2.28
O(WAT457)	—	—	—	2.31	0.08	—	2.53
O(WAT9759)	—	—	—	2.33	0.08	2.43	2.20

[a] Data was taken from crystal structure of LgtC.[3]
[b] Distance to OD2(Asp103) atom, OD1(Asp103) atom forms hydrogen bond with water 377.

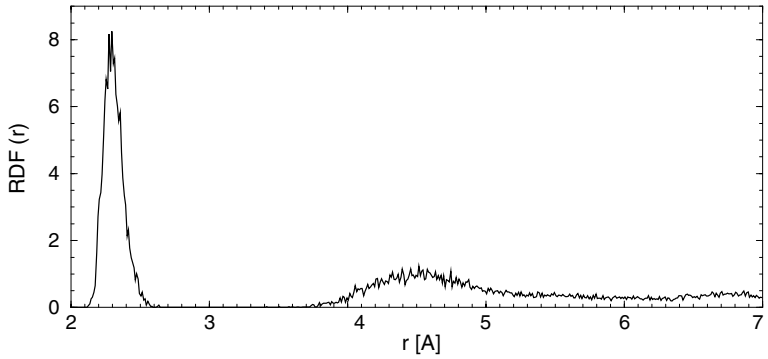


Figure 6. Radial distribution function calculated for the Mn^{2+} ion and water oxygens (data taken from the last 1 ns of simulation III). The first solvation shell (sharp peak) contains three water molecules, the second solvation shell (flat peak) contains about 10 water molecules.

The above analysis confirms that oxygen atoms of three water molecules coordinate the ion (Fig. 5b, Table 2), resulting in a coordination number of 7 for the manganese ion. This is not the most common stereochemistry of Mn^{2+} that is octahedral (coordination number 6). However, the Mn^{2+} stereochemistry is strongly influenced by the active site of the enzyme, especially by Asp105 that acts as a bidentate ligand. The average value of the O–Mn–O angle is approximately 56°, which is significantly lower than the value for the regular octahedral configuration (90°). This deformation generates some space to include one additional water molecule.

In order to confirm the above observations that are based on force field calculations, more precise ab initio calculations were performed with two water molecules (coordination 6) and three water molecules (coordination 7) using a model of the binding site with a limited number of atoms. The results for coordination number 7 exhibit much better agreement with the geometry taken from molecular dynamics than those obtained for the model with the coordination number of 6 (Fig. 5c and d). These results are in agreement with the hypothesis that steric requirements of UDP-Gal do not allow the Mn^{2+} ion the coordination number of seven in the simulation II. At the same time, we may conclude that the metal ion is very well located in the active site of the enzyme without the UDP-Gal, so it may easily access the active site prior to UDP-Gal.

2.4. Analysis of the tightly bound water molecules

As discussed above, analysis of the manganese environment indicated infinite (or at least very long) residence times for three coordinating water molecules. Analysis has been extended to all water molecules in contact with the protein in order to search for tightly bound water molecules that could play a role in the architecture of the binding site and might also become structural water molecules. A number of tightly bound water molecules were found to stay essentially in the same positions as those in the X-ray structure. In general, the 'internal' tightly bound waters stayed in their positions over the whole (or nearly whole) trajectory, whereas those captured on the surface of LgtC were often replaced by water molecules from the solution. It was confirmed that most of the 'internal' tightly bound waters stabilize the structure of protein, whether forming a bridge between two or three secondary structural motifs or keeping stable one single loop of the protein. With regard to the binding site, 15 water molecules have residence time longer than 1 ns in simulation III and 8 ns in simulation II (see supplementary materials, Table 1S). Representative picture taken from the end of both trajectories, displaying the protein backbone and substrate(s) (Fig. 7), show the typical arrangement of water molecules in the active site.

Simulation in the presence of nucleotide sugar—Four water residues from the X-ray structure stay in the active site during the entire MD trajectory, although with some small variations in hydrogen bonding pattern. Three of them are in direct contact with uridine–ribose part of UDP-Gal (Fig. 7a). Water 483 forms a bridge between OP2 of the pyrophosphate part of UDP-Gal and the backbone carbonyl group of His78. Water 324 is bound between the O2′ atom of ribose and the carboxyl group of Asp103. Water 382 forms a bridge between the carbonyl group of Leu102 and the amido group of Ala5, keeping together two amino acid strands of LgtC. The fourth stable water molecule (302) interacts with the previous one and also forms a bridge between helix Ser80-Leu87 and the neighboring strand of the protein (Tyr101). Four other water molecules arrived from the bulk solution at the beginning of MD simulation, approached the donor-binding site and stayed there for a reasonably long time (over 1000 ps), but their binding was not as stable as in the case of the X-ray water molecules (which were localized deeply inside the molecule). No water molecule can come close to the manganese ion, as the latter is steadily coordinated by six electronegative atoms from protein and donor substrate (see above).

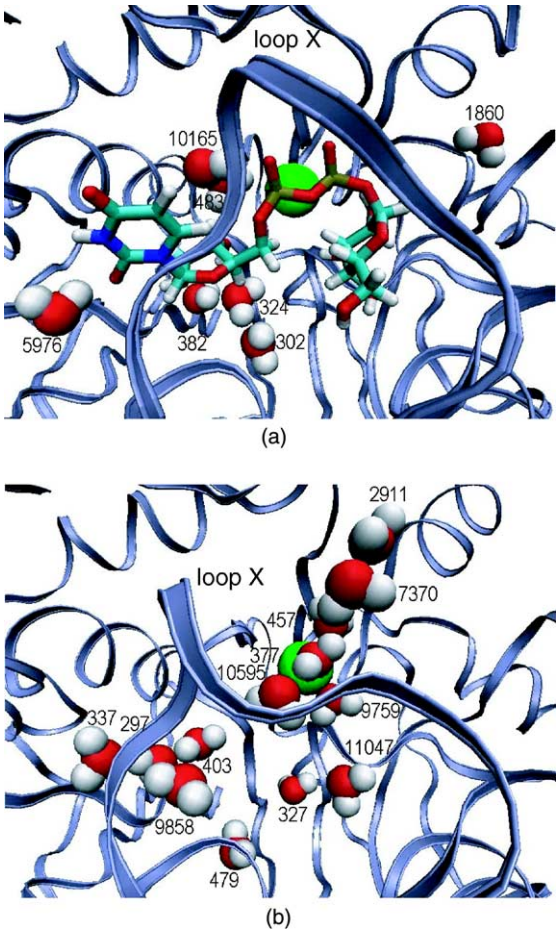


Figure 7. End-of-trajectory picture of several stable water molecules in simulations II (a) and III (b). The accessibility of the active site for solvent is higher in the absence of donor substrate than in its presence. Apart of this, the picture shows the change of conformation of loop X (loop Y was not included for better readability of the picture).

Since no acceptor is included in this simulation, it is of interest to analyze how water molecules occupy this empty space. The affinity of LgtC to the lactose acceptor is in the millimolar range and only three hydrogen bonds are observed in the crystal structure between the lactose substrate and the surrounding amino acids.[3] When examining the neighborhood of the three amino acids involved (Cys246, Asp130, and Gln189), two tightly bound water molecules could be localized in equivalent positions of two oxygen atoms of lactose, O3 of glucose, and O5 of galactose. Water 1860 came in 1658 ps from the bulk solution and formed hydrogen bonds with the hydroxyl group of Thr212 and the carboxyl group of Asp130. Water 6853 also came in 586 ps from the bulk solution and formed a hydrogen bond with Asp130 and Gln189 until 1498 ps (with a short interruption of 40 ps, forming a hydrogen bond with Cys246). No water remained for significant time at the theoretical location of the acceptor hydroxyl group (i.e., O4 of galactose). It therefore appears that the donor substrate is protected from hydrolysis not only by side chains of residues Ile76, Cys246, and Tyr186, but also by stable binding of water molecules 1860 and 6853 to residues mentioned above.

Simulation in the absence of nucleotide sugar—Similar remarks apply as for simulation II. The stable water molecules positioned around the active site of the protein are listed in the supplementary material. Many tightly bound waters stabilize the structure of the protein by keeping one or more structural motifs in a certain shape. For example, water 319 was found to stabilize loop X over nearly the entire trajectory, staying in a fixed position inside the loop, between carbonyl oxygens of Ile76, Ile79, and Ser80 and the side chain of Gln187. In contrast, several water molecules exchanged around the more flexible loop Y. Waters 377, 457, and 9759 coordinated the manganese ion as a consequence of absence of donor substrate. Several water molecules are located in the binding site for UDP-Gal where they fill the space for donor substrate (Fig. 7a). They exchange more frequently because of the absence of this substrate and due to the opening of loops X and Y.

3. Conclusions

Three molecular dynamics simulations were performed on fully solvated galactosyltransferase LgtC from *Neisseria meningitidis*. The key objective was to run a simulation with the manganese ion and the substrate donor UDP-Gal present and to compare it with the simulation where the UDP-Gal was removed. Essential dynamics analysis of the trajectories clearly revealed that some conformational movement occurs in loop X (residues 75–80) and substantial movement on loop Y (residues 246–251) close to the C-terminal of the protein. In the simulation with UDP-Gal this loop is tightly bound to the UDP-Gal. However, when UDP-Gal is not present in the system, both loops open almost by 10 Å and the active site is exposed to the solvent. Three water molecules enter it and coordinate strongly the manganese ion. We have identified residues and dihedrals that are responsible for the loops opening. This 'hinge region' is composed of the residues 246–247. Blocking this region could allow for fixing the loops, especially loop Y, and may be also for crystallization of the enzyme without UDP-Gal, which has so far been a problem.

Detailed analysis of the coordination sphere of manganese indicates that the coordination sphere of the ion is very stable when UDP-Gal is present. In this case, the coordination number is 6. When UDP-Gal is removed, the coordination number increases to 7 by coordination of three bulk water molecules. The coordination number of 7 is confirmed by quantum mechanics calculations. Also in this case, the coordination is very stable. Altogether, we may conclude that the manganese ion is

tightly bound in the active site of the enzyme, even if UDP-Gal is not present.

In the analysis of structural water molecules, we focused on those residing in the active site of the protein. We found out that the number of water molecules is higher (and their residence times are correspondingly shorter) in the absence of donor substrate than in its presence, suggesting a higher mobility of water molecules and better accessibility of the active site in the former case.

4. Methods

The simulations were performed using the AMBER-6.0 program package[19] using the module SANDER. The initial coordinates were taken from the X-ray structure of LgtC galactosyltranseferase with UDP-2-deoxy-2-fluorogalactose (UDP-Gal, donor analog) and 4-deoxy-lactose (acceptor analog), which is deposited in the Protein Databank[20] under the code 1GA8.[3] The original structure had to be modified. The four missing residues 218–221, which were not detected in X-ray experiment, were added in Sybyl 6.7[21] using protein loop search. In the simulation III, both donor and acceptor analog molecules were removed. In the simulations I and II, that differs only by the tautomeric state of His244, the acceptor analog molecule was removed and donor analog molecule was replaced with the native UDP-galactose. All of the simulated structures were built up by module XLEAP of AMBER using the AMBER force field parm94.[22] The parameters and charges for UDP-Gal were taken from our previous parameterization.[23] Hydrogens were also added by XLEAP. The system was solvated with TIP3P water molecules added as WATBOX216 cubes to form a rectangular water box. For details about number of water molecules and size of the box see Table 1. The minimum distance of the protein atoms from the walls of this rectangular box was set to 10 Å. In order to maintain electro-neutrality, 9 (trajectories I, II) or 7 (trajectory III) Na^+ counter ions were added by XLEAP.

A stepwise equilibration of the system was performed before the production phase of MD was started. After the energy minimization of hydrogen atoms, ions and noncrystallographic water molecules were minimized, followed by full minimization of ions and all water molecules. To finish the equilibration of the solvent, a 20 ps MD of ions and water molecules was performed at 298 K. The equilibration continued with minimization of the whole system with gradual decreasing of restraints, which were imposed on side chains of the protein residues. After the last step, where no restraints were put on the protein, a 70 ps molecular dynamics with slow heating from 10 to 298 K was carried out, followed by 30 ps dynamics. Both equilibration MD simulations were carried out under constant pressure of 1 atm.

The production phase of MD was performed at a constant pressure of 1 atm and a temperature of 298 K. A timestep of 1 fs was used for the simulations and the coordinates were saved every 1 ps. Bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm.[24] In all simulations, non-bonded interactions were treated with the Particle-Mesh Ewald (PME) method.[25] The nonbonded pairlist was updated every 10 MD steps using a cutoff of 9 Å.

The MD trajectories were analyzed with the modules CARNAL and PTRAJ of AMBER package. Essential dynamics analysis was performed with modules *trjconv*, *g_covar* and *g_anaeig* of GROMACS suit of programs.[26,27] Tightly bound water molecules analysis was performed with program RESTIME (R. Rittenhouse, 2000, unpublished program), and in-house programs (Z. Kriz, 2002).

Ab initio calculations were performed with Gaussian 98 molecular modeling package,[28] using the Hartree–Fock method with 6-31G* basis set. Two restricted models of manganese ion coordination surrounding were used. The first model represents coordination 7 and the second one coordination 6. First model consists of truncated Asp103, Asp105, and His244 residues, Mn^{2+} ion, and water molecules 377, 457, and 9759. The Asp103 and Asp105 were truncated to formiate to mimic δ-carboxyl groups. The His244 residue was truncated to imidazole ring. Initial coordinates were taken from the final coordinates of the trajectory III, coordinates of new hydrogen atoms were chosen in such a way that new bonds had a length of 1 Å. Models were calculated with multiplicity six and with initial wavefunction guess calculated with STO-3G basis set. Hydrogen atom coordinates were first optimized followed by full energy minimization of the whole structure. The second model was generated from the result of the first model by removing the water 457. This structure was also fully geometry optimized.

5. Miscellaneous

The drawings were prepared using the VMD program.[29]

Supporting information available

List of tightly bound water molecules in LgtC active site (Table 1S), and the history of 'pseudodihedral' angles as a representation of hinge (Cα245- Cα246- Cα247- Cα248) and internal movement (Cα249- Cα250- Cα251- Cα252) of the loop Y (Fig. 1S).

Acknowledgements

L.S. is supported by a grant from the French Ministère des Affaires Etrangères. The research has partially been supported (L.S., P.K.) by the Ministry of Education of the Czech Republic (grant LN00A0016). J.K.'s stays in Grenoble were financed by CNRS. We are also grateful to Supercomputing Centre in Brno and the CECIC in Grenoble for providing us with computational resources.

References

1. Gotschlich, E. *J. Exp. Med.* **1994**, *180*, 2181–2190.
2. Tzeng, Y. L.; Stephens, D. S. *Microbes Infect.* **2000**, *2*, 687–700.
3. Persson, K.; Ly, H. D.; Dieckelmann, M.; Wakarchuk, W. W.; Withers, S. G.; Strynadka, N. C. J. *Nature Struct. Biol.* **2001**, *8*, 166–175.
4. Ly, H. D.; Lougheed, B.; Wakarchuk, W. W.; Withers, S. G. *Biochemistry* **2002**, *41*, 5075–5085.
5. Gastinel, L. N.; Bignon, C.; Misra, A. K.; Hindsgaul, O.; Shaper, J. H.; Joziasse, D. H. *EMBO J.* **2001**, *20*, 638–649.
6. Boix, E.; Swaminathan, G. J.; Zhang, Y. N.; Natesh, R.; Brew, K.; Acharya, K. R. *J. Biol. Chem.* **2001**, *276*, 48608–48614.
7. Boix, E.; Zhang, Y. N.; Swaminathan, G. J.; Brew, K.; Acharya, K. R. *J. Biol. Chem.* **2002**, *277*, 28310–28318.
8. Patenaude, S. I.; Seto, N. O.; Borisova, S. N.; Szpacenko, A.; Marcus, S. L.; Palcic, M. M.; Evans, S. V. *Nature Struct. Biol.* **2002**, *9*, 685–690.
9. Nguyen, H. P.; Seto, N. O.; Cai, Y.; Leinala, E. K.; Borisova, S. N.; Palcic, M. M.; Evans, S. V. *J. Biol. Chem.*, **2003**, *278*, 49191–49195.
10. Heissigerova, H.; Breton, C.; Moravcova, J.; Imberty, A. *Glycobiology* **2003**, *13*, 377–386.
11. Andre, I.; Tvaroska, I.; Carver, J. P. *Carbohydr. Res.* **2003**, *338*, 865–877.
12. Gunasekaran, K.; Ma, B.; Ramakrishnan, B.; Qasba, P. K.; Nussinov, R. *Biochemistry* **2003**, *42*, 3674–3687.
13. Resat, H.; Mezei, M. *Biophys. J.* **1996**, *71*, 1179–1190.
14. Amadei, A.; Linssen, A. B.; Berendsen, H. J. *Proteins* **1993**, *17*, 412–425.
15. Van Aalten, D. M. F.; De Groot, B. L.; Findlay, J. B. C.; Berendsen, H. J. C.; Amadei, A. *J. Comput. Chem.* **1997**, *18*, 169–181.
16. Wakarchuk, W. W.; Cunningham, A.; Watson, D. C.; Young, N. M. *Protein Eng.* **1998**, *11*, 295–302.
17. Breton, C.; Bettler, E.; Joziasse, D. H.; Geremia, R. A.; Imberty, A. *J. Biochem. (Tokyo)* **1998**, *123*, 1000–1009.
18. Busch, C.; Hofmann, F.; Selzer, J.; Munro, S.; Jeckel, D.; Aktories, K. *J. Biol. Chem.* **1998**, *273*, 19566–19572.
19. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P.A. University of California: San Francisco, 1999.
20. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucl. Acids Res.* **2000**, *28*, 235–242.
21. SYBYL; 6.4 ed.; 1699 S. Hanley Road, Suite 303, St Louis, MO 63144, USA.
22. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
23. Petrova, P.; Koca, J.; Imberty, A. *J. Am. Chem. Soc.* **1999**, *121*, 5535–5547.
24. Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. *J. Comp. Chem.* **1977**, *23*.
25. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
26. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comp. Phys. Commun.* **1995**, 43–56.
27. Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Mod.* **2001**, 306–317.
28. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C.Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J.A. 98, Revision A. 9 ed.; Gaussian, Inc., 1998.
29. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, 33–38.