

Enhanced conformational sampling of carbohydrates by Hamiltonian replica-exchange simulation

Sushil Kumar Mishra^{2,3}, Mahmut Kara⁴,
Martin Zacharias⁴, and Jaroslav Koča^{1,2,3}

²Central European Institute of Technology, Masaryk University, Kamenice 5, 61137 Brno, Czech Republic; ³National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kotlářská 2, 61137 Brno, Czech Republic; and ⁴Physik-Department T38, Technische Universität München, James-Frank-Str. 1, Garching D-85748, Germany

Received on June 20, 2013; revised on September 29, 2013; accepted on October 14, 2013

Knowledge of the structure and conformational flexibility of carbohydrates in an aqueous solvent is important to improving our understanding of how carbohydrates function in biological systems. In this study, we extend a variant of the Hamiltonian replica-exchange molecular dynamics (MD) simulation to improve the conformational sampling of saccharides in an explicit solvent. During the simulations, a biasing potential along the glycosidic-dihedral linkage between the saccharide monomer units in an oligomer is applied at various levels along the replica runs to enable effective transitions between various conformations. One reference replica runs under the control of the original force field. The method was tested on disaccharide structures and further validated on biologically relevant blood group B, Lewis X and Lewis A trisaccharides. The biasing potential-based replica-exchange molecular dynamics (BP-REMD) method provided a significantly improved sampling of relevant conformational states compared with standard continuous MD simulations, with modest computational costs. Thus, the proposed BP-REMD approach adds a new dimension to existing carbohydrate conformational sampling approaches by enhancing conformational sampling in the presence of solvent molecules explicitly at relatively low computational cost.

Keywords: adaptive biasing force simulations / biasing potential replica-exchange simulation / conformational sampling / molecular dynamics simulation / saccharides

Introduction

Carbohydrates that are covalently linked to proteins (glycoproteins) and lipids (glycolipids) are the most prominent cell

surface-exposed structures and are implicated in many biological processes in nearly all living systems: molecular recognition (Feizi and Mulloy 2003), viral entry (Feizi et al. 2008), cell-cell communication (Geijtenbeek et al. 2000), bacterial-host infection (Karlsson 1999), immunological protection (Morelli et al. 2011), fertilization, embryogenesis (Tiemeyer and Goodman 1996; Vacquier and Moy 1997), neural development (Gama 2009), cell proliferation and organization into specific tissues being some of them. However, the structural complexity of carbohydrates in living systems has hindered the expected development in the fields of glycobiology. Determination of the three-dimensional structure of carbohydrates and understanding their molecular recognition by proteins are key challenges of structural glycobiology. Complexity arises mainly because of certain unusual characteristics of carbohydrates: different types and numbers, the anomeric effect and their ability to form glycosidic linkages at various positions. When compared with proteins and nucleic acids, carbohydrates are a far more diverse group of molecules, due to the variability in their constituent saccharide monomers. The complexity of large saccharides increases exponentially with the variation found in the carbon backbone length of the monomer, anomericity and side group orientation. For instance, if the combinatorial potential of the individual units is fully exploited, two different amino acids could form only two dipeptides, but two different hexose sugars could produce up to 11 different disaccharides.

Monosaccharides can oligomerize via various kinds of linkages to form diverse, branched oligo- and polysaccharide structures. Two monosaccharide units may also have different ring conformations. Each ring conformation and glycosidic linkage between saccharides usually has a number of possible low-energy conformations, and taken together this makes it difficult to determine the structure experimentally. Recent advances in X-ray crystallography and NMR spectroscopy have provided increased understanding of the structural aspects of complex saccharides, although these experimental approaches are limited by the structural flexibility and heterogeneity of saccharides (Lutke 2004; Lütke and Lieth 2004).

This knowledge has enabled the problem of determining the structure of saccharides to be divided into two steps. The first step is concerned with determining the ring conformation of individual monosaccharide units and the second is to find the preferred orientations of monosaccharide units along the connecting glycosidic bond. Biomolecular interactions of carbohydrates with proteins depend on the sugar ring conformation. The affinity of this interaction can only be determined

¹To whom correspondence should be addressed: Tel: +420-54949-4947; Fax: +420-54949-2556; e-mail: jaroslav.koca@ceitec.muni.cz

through accurate representation of the ring conformations and their transition from unbound to the bound state (Gandhi and Mancera 2008, 2009, 2011). Classical molecular dynamics (MD) simulations with commonly available force fields may not accurately model transitions of the ring conformations (Gandhi and Mancera 2010). The variety of possible configurations of individual monosaccharides: D- or L-isomers, α - or β -anomers and the number of possible ring conformations are challenging topics in glycobiology. However, if a preferred conformation of a sugar ring is known, the structure of an oligosaccharide is determined by finding the relevant dihedral angle Φ and Ψ along the glycosidic linkage between the known monosaccharide structures (Supplementary data, Figure S1). Focus of the present study is to identify the preferred arrangement of monosaccharide units in oligosaccharides by sampling the Φ and Ψ dihedral angles of each glycosidic linkage assuming that each monosaccharide adopts a known low-energy conformation.

Computational studies of oligosaccharides have evolved in several directions, including quantum and molecular mechanical studies of their conformation and energetics. To date, various computational approaches to perform a conformation analysis of oligosaccharides in vacuum, such as adiabatic potential energy surface maps (Imberty, Gerber, et al. 1990; Imberty, Tran, et al. 1990), the CICADA approach (Koča 1994), Monte Carlo simulations (Peters et al. 1993) and genetic algorithms (Nahmany et al. 2005; Xia et al. 2007) have been used. Due to the particular limits related to these computational approaches, such methods are not suitable for including explicit solvent. However, interactions with the aqueous environment are crucial to determining carbohydrate conformations; thus, solvent effects need to be included to understand the conformational behavior (Kirschner et al. 2008).

The main methodologies that can be applied with implicit/explicit solvent models are classical MD (cMD) simulation (Hardy and Sarko 1993; Naidoo and Brady 1997) and replica-exchange simulations (Re et al. 2012). All the methods that perform conformational searches in a vacuum or by using an implicit solvent model do not include the effect of hydrogen bonding between the sugar and water molecules explicitly. Such methods are also limited by the number of glycosidic bonds to be scanned. cMD in the presence of explicit solvent molecules can be used with current computers on the time scale of 0.1–1 μ s (Almond and Sheehan 2003). But, cMD is quite limited when used for sampling, because the structures may get trapped in some free-energy minima and are unlikely to cross barriers higher than a few $k_B T$. Thus, only a subset of the relevant conformations can be typically observed in classical explicit solvent MD simulations. This difficulty limits the application of cMD simulations to small fragments such as di- and trisaccharides. Therefore, MD-based advanced sampling approaches, such as replica-exchange or adaptive biasing force methods, are of great interest.

One of the simplest and widely used MD-based computational methods used to enhance the conformational sampling of proteins and nucleotides is the replica-exchange approach, where several copies of a system are simulated using MD at different temperatures (T-REMD) and neighboring replicas are allowed to exchange the conformation with a pre-specified

transition probability (Sugita and Okamoto 1999). In order to achieve efficient exchanges, the number of replicas to cover a given temperature range grows with the square root of the total number of atoms in the system. The application to systems that include many explicit solvent molecules is therefore particularly limited. As an alternative to varying the temperature, the force field or Hamiltonian for a certain part of the system can also be varied along the replicas. This approach offers an advantage that exchanges between replicas are independent of the part of the Hamiltonian that does not differ between replicas. Such a replica-exchange simulation requires much fewer replicas compared with T-REMD. One variant of such approaches is the biasing potential-based replica-exchange molecular dynamics (BP-REMD) where a biasing potential along the backbone dihedral is applied to allow effective transitions between various conformations (Kannan and Zacharias 2007). Such a scheme has been effectively used to allow the backbone transition in peptides and DNA molecules to study the folding of peptides, proteins and nucleic acids (Curuksu and Zacharias 2009; Kannan and Zacharias 2009, 2010). To the best of our knowledge, this is the first study to apply the BP-REMD method to promote the coupled transition along the dihedral angle (glycoside linkage) in saccharides. Elucidation of the three-dimensional (3D) structures and the dynamic properties of oligosaccharides in water is a prerequisite to understanding the process of protein–carbohydrate recognition and for the carbohydrate-based rational design.

This study is dedicated to understanding the dynamic behavior of the above-mentioned di- and trisaccharides in solution (Figure 1). Before we used this approach to study the dynamic behavior of several biologically relevant saccharides, we evaluated this approach on a number of disaccharides which have 1 \rightarrow 2, 1 \rightarrow 3, 1 \rightarrow 4 and 1 \rightarrow 6 glycosidic linkages. The other studied saccharides were antigenic determinant blood group B (BGB), Lewis A (Le^a) and Lewis X (Le^x) trisaccharides. BGB is an important antigen found in erythrocyte membranes and it plays a crucial role in blood transfusion and tissue transplantation. There has been contradictory evidence from both NMR and molecular modeling as to whether it is a rigid structure or stays in equilibrium between two possible conformations in solution (Imberty et al. 1995; Azurmendi and Bush 2002).

The carbohydrate-dependent blood-group antigen Le^a trisaccharide is found on the surface of erythrocytes and is capable of binding covalently to proteins and/or lipids. Lewis antigens have been found to be expressed at a high level in many types of cancer (Sakamoto et al. 1986; Murata et al. 1992; Lloyd 2000). Le^a trisaccharide is also used for blood group typing and antibody binding assays. Le^x is known to be a tumor-associated antigen; it is involved in the fucosylation of GlcNAc residues in poly-lactosamine chains, metastasis and tumor progression in carcinomas (Sakamoto et al. 1986; Nishihara et al. 1999) and it might be a promising target for cancer treatment, including antibody-based immunotherapy. Knowing the structural preferences of Le^a trisaccharides could play a key role in designing drugs which might inhibit the activity and binding of biological targets. Thus, techniques to determine the 3D structure of such biologically interesting saccharides in water are of great interest and can be further used in carbohydrate-based drug design and/or vaccine development.

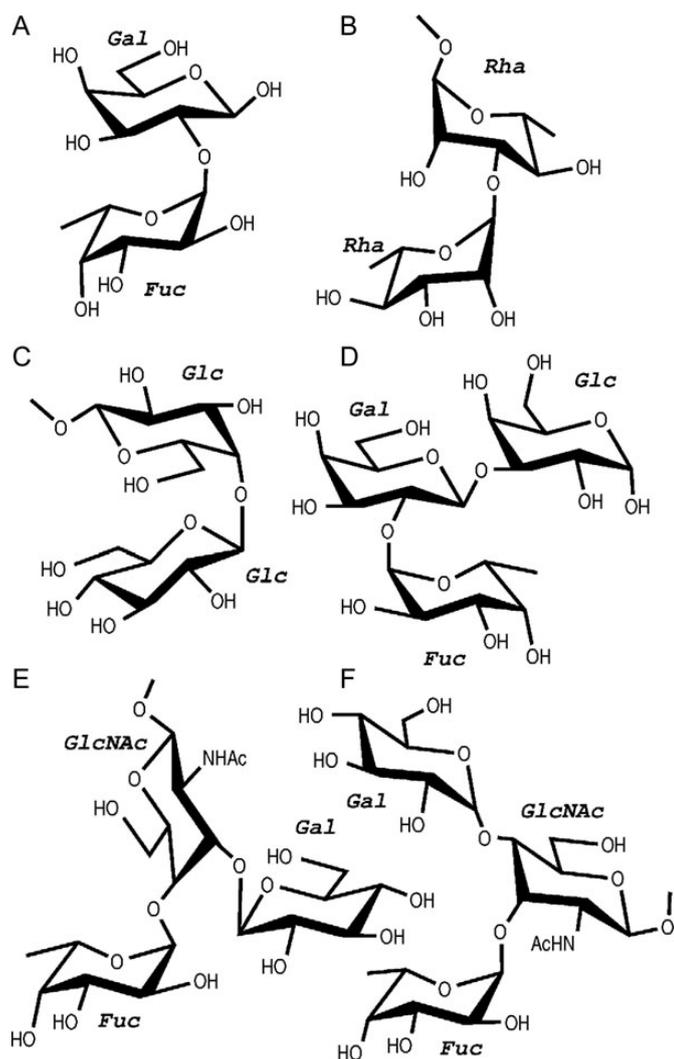


Fig. 1. Schematic representations of di- and trisaccharides used for conformational sampling. Labels: α -L-Fuc(1 \rightarrow 2) β -D-Gal (A), α -L-Rha(1 \rightarrow 3) α -L-MeRha (B), β -D-Glc(1 \rightarrow 4) β -D-Glc (C), BGB trisaccharide (α -D-Gal(1 \rightarrow 3)[α -L-Fuc(1 \rightarrow 2)] β -D-MeGal) (D), Le^a trisaccharides (α -L-Fuc(1 \rightarrow 4)[β -D-Gal(1 \rightarrow 3)] β -D-MeGlcNAc) (E), Le^x trisaccharides (α -L-Fuc(1 \rightarrow 3)[β -D-Gal(1 \rightarrow 4)] β -D-MeGlcNAc) (F).

In this study, we propose a BP-REMD protocol to enhance the conformational sampling of oligosaccharides and evaluate its performance on several di- and trisaccharides in the presence of an explicit solvent. Thus, in this article, we are mainly focused on two goals. First, we present a BP-REMD simulation procedure that enables efficient sampling of the Φ and Ψ torsion angles along the glycosidic linkage. Second, we also validate the GLYCAM06 force field by using it for a structure prediction of di- and trisaccharides. Moreover, we then present the structural and dynamic behavior of the biologically relevant trisaccharides BGB, Le^a and Le^x .

Results and discussion

Comparison of sampling efficiency on disaccharides

Disaccharides with the glycosidic linkage 1 \rightarrow 2, 1 \rightarrow 3, 1 \rightarrow 4 and 1 \rightarrow 6 were used for calibrating and testing the BP-REMD

sampling scheme. Three types of MD simulations were performed: in the first simulation, systems were subjected to a 30 ns long cMD simulation (for some cases, the simulations were extended to 1 μ s to check convergence, see below). In the second type, a replica-exchange simulation was set at various levels of biasing potential in the force field for each replica (see *Methods*). This resulted in the destabilization of favorable dihedral angle combinations and the overcoming of local free-energy barriers, improving the exploration of all the major possible free-energy minima. The BP-REMD simulations were performed for 4 ns in each replica (with seven replicas resulting in a total simulation time of 7×4 ns which is approximately equivalent to the computational cost of the 30 ns cMD runs). However, only the reference replica under the control of the original force field was used for the analysis. In order to assess the effectiveness of BP-REMD, a third variant of MD simulation, ABF, was performed along the Φ and Ψ dihedral angles. In ABF simulation, an average biasing force was accumulated in bins of 4° , which was then used to estimate the free-energy profile for the Φ and Ψ dihedral angles in the range of -180 to 180° . Such an approach is, however, restricted by the dimension of the scan; it was not feasible to use such an approach with trisaccharides.

The MD trajectories of α -L-Fuc(1 \rightarrow 2) β -D-Gal in explicit solvent at $T = 300^\circ\text{K}$ show structures exhibiting limited flexibility in moving between the various Φ and Ψ torsions in the local minima M1. However, the range of Φ and Ψ torsion sampled in BP-REMD simulations is quite wide and is in good agreement with the ABF simulations. This shows that compared with cMD, BP-REMD substantially enhances conformational sampling and more possible conformations of saccharides are seen in water. In the ABF simulation, we also observed that the global minimum for M1 was quite wide and BP-REMD detected the global minimum correctly, which was slightly shifted from the one seen in cMD simulations.

The cMD simulation indicated only the M1 region in the Ramachandran plot had been sampled, whereas BP-REMD simulation clearly enhanced the sampling and two other possible conformations of disaccharides in water were assessed (Figure 2). Moreover, the driving force for the orientation of the sugar units seems to result from solvent interactions, because we did not observe any strong intramolecular hydrogen bond between these two sugar residues in MD simulations (Figure 2D and E).

With another disaccharide D2, which has a 1 \rightarrow 3 glycosidic linkage, there were mainly three minima. In cMD, it was usually initially trapped in global M1 minima and the M2 minimum was not sampled. When the starting conformation of the cMD simulation was at M2, it initially samples the local M2 minima and then the global M1 minimum was also sampled in the final stage of this simulation (Supplementary data, Figure S2). In both cases, the M3 conformations were not sampled in cMD simulations. In contrast in BP-REMD simulations, both M1 and M2 were properly sampled, and then the method also enabled sampling of the M2 and M3 regions in much shorter overall simulation times than the cMD run (Figure 3). The structure of D2 in M1 is able to form electrostatic contacts between the O2 and O4 hydroxyls of Me-Rha and the oxygen of the glycosidic bond, whereas in M2, O2 interacts with O5 of the L-Rha (Figure 3D and E). A third disaccharide β -D-Glc(1 \rightarrow 4) β -D-Glc exhibited a

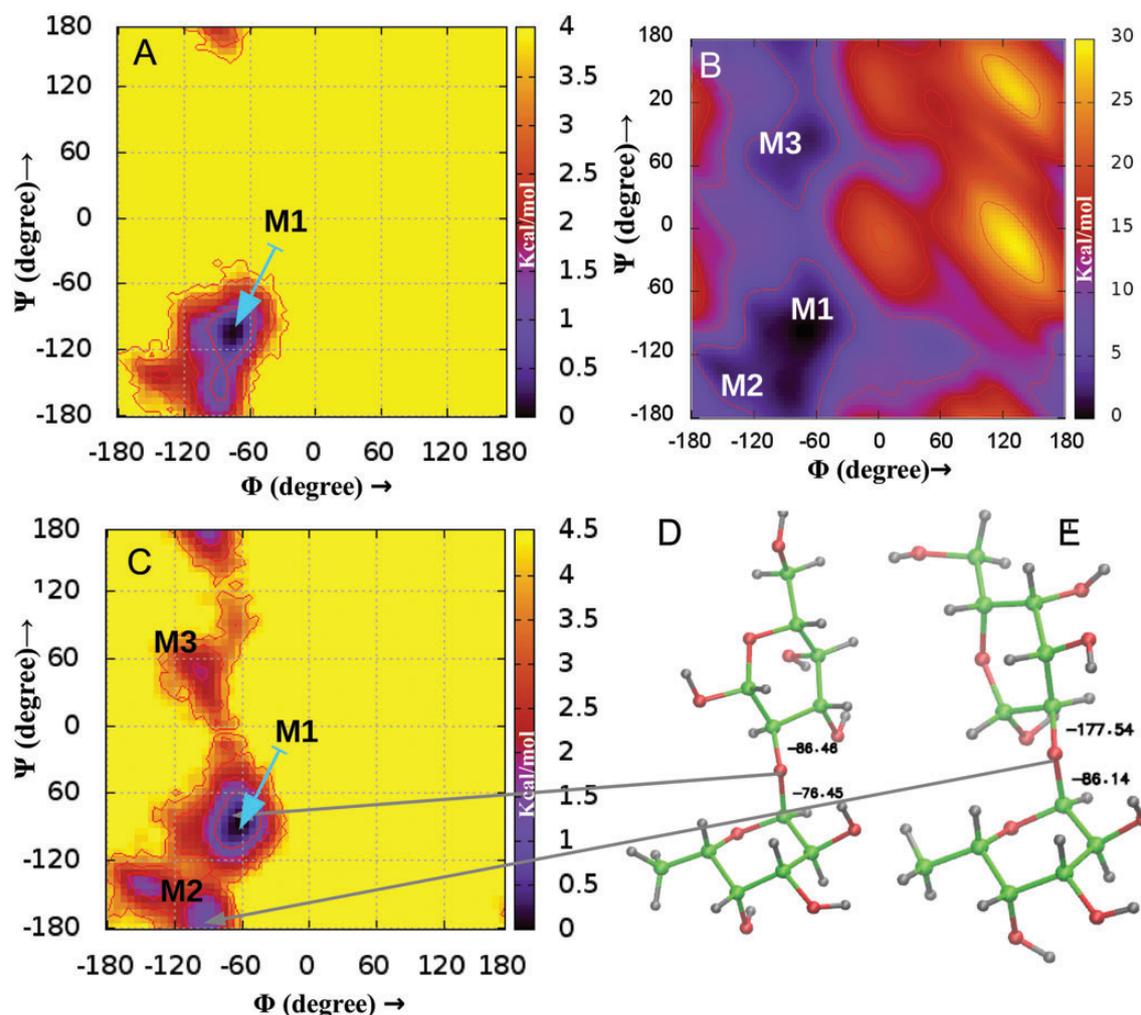


Fig. 2. α -L-Fuc(1 \rightarrow 2) β -D-Gal Φ and Ψ dihedral angle distributions in explicit solvent. The maps represent the free energy calculated from the probability distribution of Φ - Ψ dihedral angles obtained from: (A) cMD simulation for 30 ns at 300°K; (B) free energy obtained from numerical integration of gradients obtained from ABF simulation covering full range $[-180^\circ; +180^\circ]$ in 90 bins for 50 ns; (C) sampling in reference replica of the BP-REM simulation for 4 ns per replica at 300°K; (D) snapshot of structure of saccharide taken from region M1 of BP-REM and (E) snapshot of structure of saccharide taken from region M2 of BP-REM.

quite similar free-energy profile for the 1 \rightarrow 4 glycosidic linkage, except for the fact that M2 and M3 appeared to be deeper compared with D2 (Figure 4). In M1, hydroxyl O3 of D-Glc forms a polar interaction with O5, whereas in M2 this interacts with the O2 hydroxyl of D-Glc at the non-reducing end (Figure 4D and E). In both the cases, 30 ns cMD simulation was insufficient to sample the M3 region but BP-REM was efficient enough to enhance the sampling of M1 and M2 and then also promoted sampling of the M3 region (Figure 4A and C).

We applied the simulations to two disaccharides known to have multiple conformations in solution, namely isomaltose, α -D-Neu5Ac(2 \rightarrow 6) β -D-Gal (D4) and α -D-Glc(1 \rightarrow 6) β -D-Glc (D5) having (1 \rightarrow 6) and (2 \rightarrow 6) glycosidic linkages. These disaccharides unambiguously populate multiple conformations due to the additional torsional angle ω (denoted by atoms O1—C6—C5—O5) involved. Both systems are relevant both biologically and for the purpose of theoretical conformational studies. The glycosidic dihedral angle sampling for D4 is crucial to understand the binding of human influenza A virus to

cell surface receptors (Choi et al. 2009; Javaroni et al. 2009). Due to these facts, we extended cMD simulations of these disaccharides up to 1 μ s and compared results with the BP-REM simulation.

The α (2 \rightarrow 6) linkage of α -D-Sial (2 \rightarrow 6) β -D-Gal shows an equilibrium of two conformers along the Φ angles at -60° and $+60^\circ$ and of the three staggered rotamers of the ω angle (Supplementary data, Figure S3). One of the minima, namely M3, was rarely visited during 30 ns cMD simulation (Supplementary data, Figure S3A and B), but efficiently sampled by BP-REM (Supplementary data, Figure S3C and D), and also by the 1 μ s cMD simulation (Supplementary data, Figure S3E and F). BP-REM simulation reasonably reflected previously observed conformations from calculations and experiments (Poppe 1992; Choi et al. 2009).

Last disaccharide, D5 is commonly known as β -isomaltose and found in starch, glycogen branching. The α (1 \rightarrow 6) glycosidic linkage of isomaltose is found in starch and glycogen branching. In this case, the 30 ns cMD (Supplementary data,

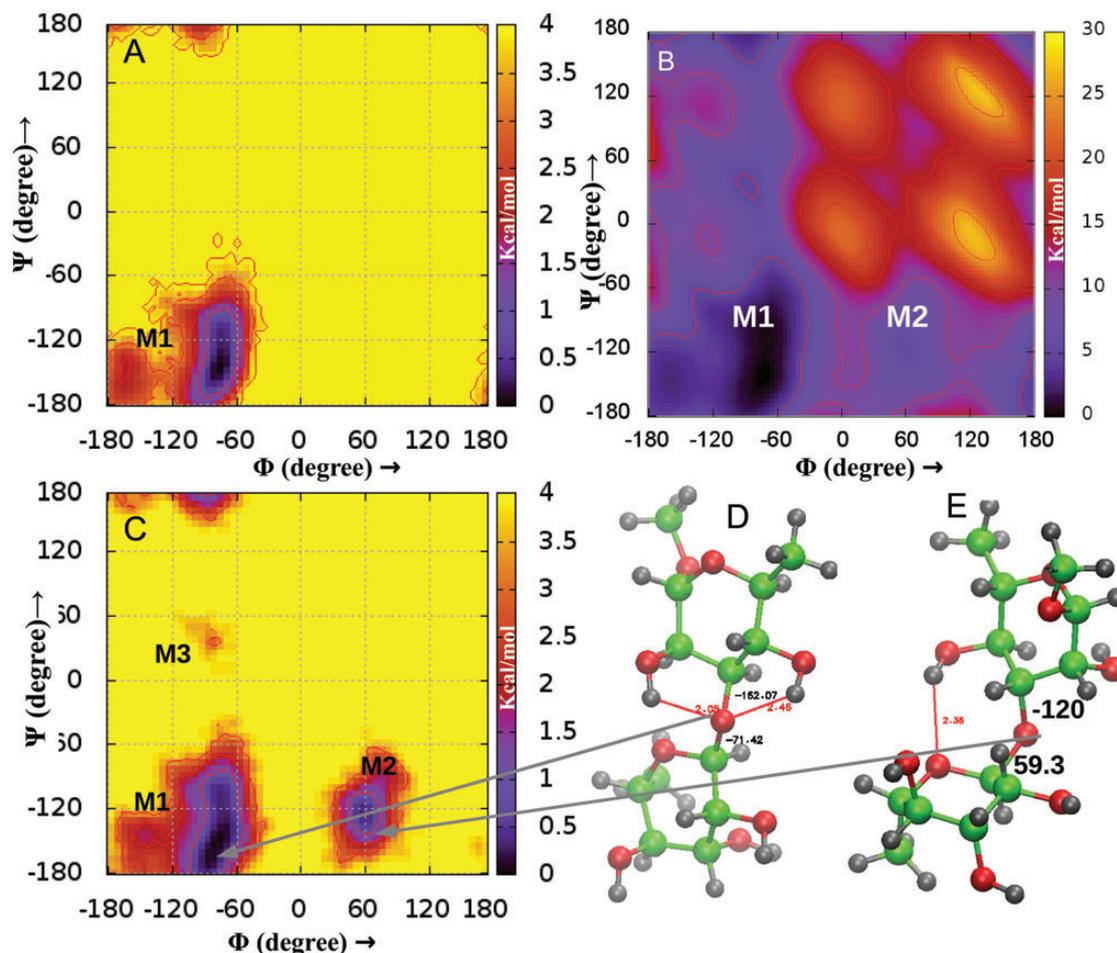


Fig. 3. α -L-Rha(1 \rightarrow 3) α -L-MeRha Φ - and Ψ dihedral angles distributions in explicit solvent. The maps represent the free energy calculated from the probability distribution of the Φ - Ψ dihedral angles obtained from: (A) cMD simulation for 30 ns at 300°K; (B) free energy obtained from the numerical integration of gradients obtained from ABF simulation covering full range [-180°; +180°] in 90 bins for 50 ns; (C) sampling in reference replica of BP-REMD simulation for 4 ns per replica at 300°K; (D) snapshot of structure of saccharide taken from region M1 of BP-REMD and (E) snapshot of structure of saccharide taken from region M2 of BP-REMD. Red dotted lines represent a possible hydrogen bond between those atoms.

Figure S4A and B) and the BP-REMD (Figure S4C and D) capture a realistic part of the free-energy surface. However, BP-REMD (Supplementary data, Figure S4C and D) samples the M5 region more efficiently, similar to what we see in the 1 μ s cMD simulation (Supplementary data, Figure S4E and F). In D5, the variation of the ϕ angle is smaller than ψ and ω angles. Compared with other glycosidic linkages, it can be seen that (1 \rightarrow 6) linkage are conformationally versatile and exists in many possible stable conformations. The existence of two or more alternative conformations differing in the value of the ω dihedral angle for the (1 \rightarrow 6)-linked disaccharides agrees with the results from previous computational (Kony et al. 2004; Pereira et al. 2006; Choi et al. 2009; Javaroni et al. 2009; Perić-Hassler et al. 2010) and experimental measurements (Ohruí et al. 1985; Poppe 1993; Roën, Mayato, et al. 2008; Roën, Padrón, et al. 2008). Thus, the BP-REMD approach works equally well for (1 \rightarrow 6) and (2 \rightarrow 6) glycosidic linkages (Supplementary data, Figure S3 and S4).

It is known that transitions in the linkage between sugar units may sensitively depend on the sugar ring conformation. The

occurrence of alternative ring conformations besides of the starting chair form was carefully checked in all simulations (Supplementary data, Table S1 and Figures S6 and S7). Although few transient transitions to alternative states occurred, the sampled conformations adopted one (chair) sugar ring conformation (Supplementary data, Figures S6 and S7). In summary, BP-REMD predictions were generally in good agreement with the preferred regions seen in the more rigorous and computationally expensive ABF simulation and in the long 1 μ s cMD simulations. Therefore, we subsequently used this BP-REMD protocol for the study of biologically relevant trisaccharides.

BGB trisaccharide

As the configuration of L-fuc and D-Gal along their glycosidic bond with D-MeGal are of importance in molecular recognition, it is interesting to know to what extent BGB can be flexible in an aqueous environment. Earlier studies of BGB using NMR and molecular modeling concluded that BGB is essentially rigid (Cagas and Bush 1990), but later a more

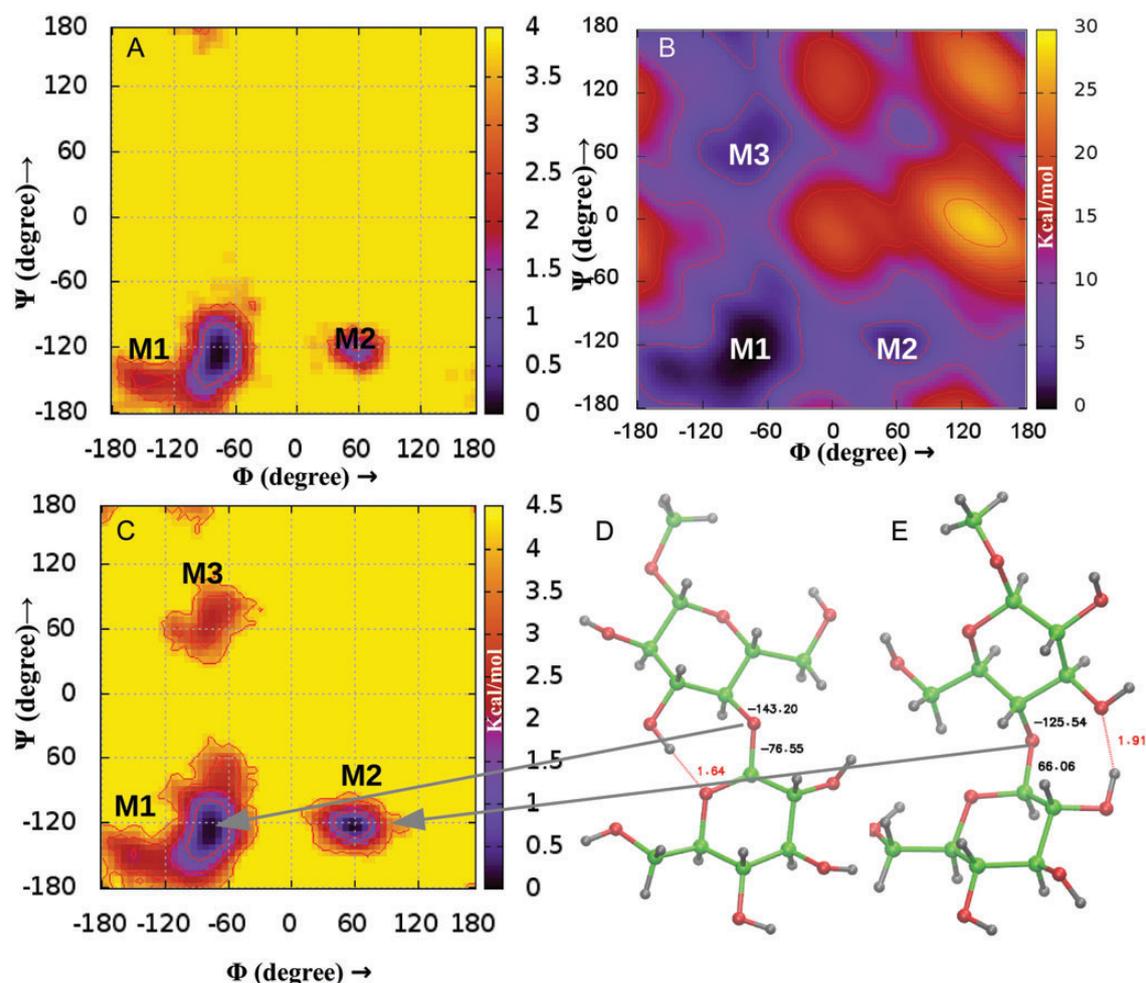


Fig. 4. β -D-Glc(1 \rightarrow 4) β -D-Glc Φ and Ψ dihedral angle distributions in explicit solvent. The maps represent the free energy calculated from the probability distribution of Φ - Ψ dihedral angles obtained from: (A) cMD simulation for 30 ns at 300K; (B) free energy obtained from numerical integration of gradients obtained from ABF simulation covering full range $[-180^\circ; +180^\circ]$ in 90 bins for 50 ns; (C) sampling in reference replica of BP-REMD simulation for 4 ns per replica at 300K; (D) snapshot of structure of saccharide taken from region M1 of BP-REMD and (E) snapshot of structure of saccharide taken from region M2 of BP-REMD. Red dotted lines represent a possible hydrogen bond between those atoms.

sophisticated molecular modeling approach suggested the possibility of two main conformations (Imberty et al. 1995). We performed cMD and BP-REMD simulations to investigate the possible conformations of the BGB trisaccharide in aqueous solution using the GLYCAM06 force field. We also compared the sampling performance of a BP-REMD simulation on the evolution of Φ and Ψ for each linkage with cMD simulations. In this test case, a 30 ns cMD simulation was able to sample most of the main conformers that were seen in the BP-REMD simulations. Nevertheless, the sampling of the free-energy minima in BP-REMD was enhanced and an additional conformer (region M2 in Figure 5D) was observed.

It was found that the minimum energy conformer is mainly characterized by a van der Waals interaction between the H3, H5 of α -D-Gal and H1 of α -L-Fuc units (Figure 5E). Otter et al. (1999) had previously shown that the hydroxyl groups of D-Gal and α -L-Fuc in this conformer were too far apart for forming a significant polar interaction. Moreover, we observed a possible polar interaction between the hydroxyl O2 of L-Fuc

and O6 of D-Gal. A histogram of the possible hydrogen bond, shown in Figure 6A, was generated to ascertain the possibility of a hydrogen bond between O2 and O6 hydroxyls, although no such hydrogen bond or water bridge between L-Fuc and D-Gal has been observed in the crystal structure (Otter et al. 1999). This can be explained by the fact that the saccharide adopts many more conformations during simulation and not just the single conformation which is seen in the crystal structure. Additionally, another possible intramolecular hydrogen bond was observed between the O2 and O4 hydroxyls of α -D-Gal and β -D-MeGal, respectively. The histogram of the distance of this hydrogen bond shows bond length shorter than 4Å for either of these two possible hydrogen bonds (Figure 6B). The BP-REMD simulation of BGB resulted in some conformations which were not seen in cMD (region M2 in Figure 5D). These forms do not allow for significant electrostatic interactions between L-Fuc and D-Gal but a hydrogen bond between the O2 and O4 hydroxyls of α -D-Gal and β -D-MeGal was still possible. Similar to the disaccharides, the sampled states were

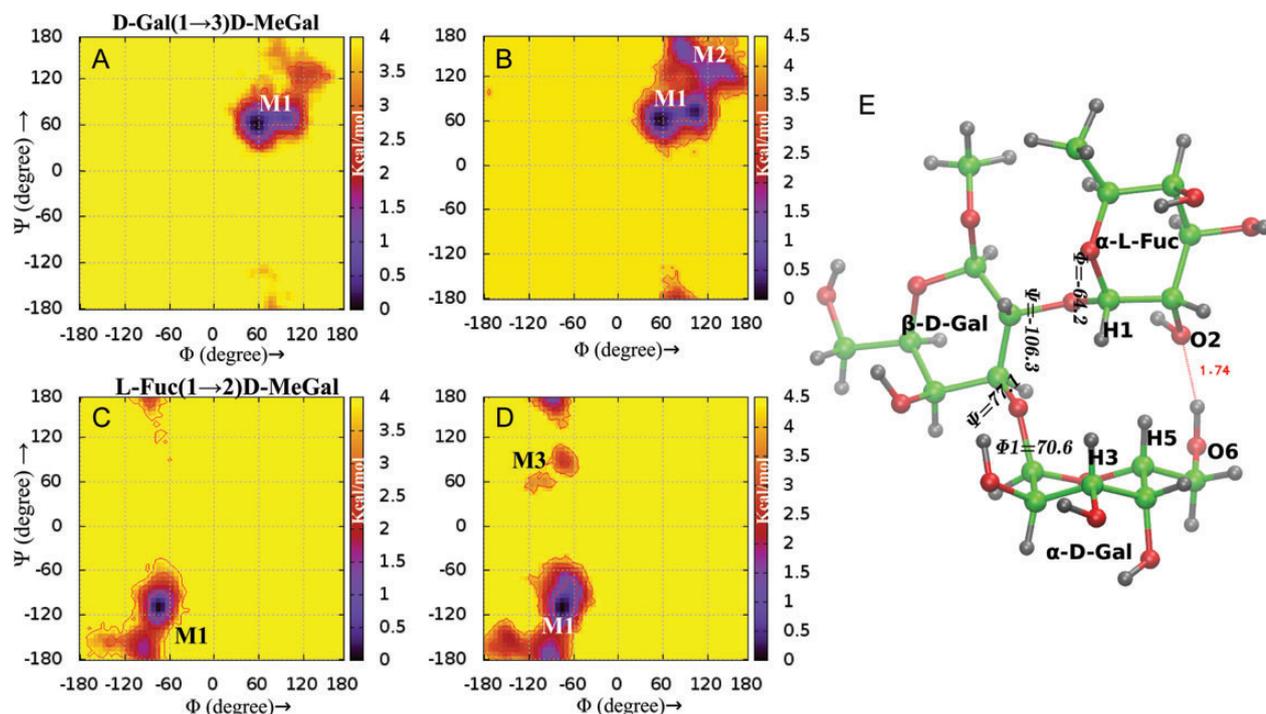


Fig. 5. Free-energy maps obtained from probability distribution of Φ and Ψ dihedral angle distributions of linkage α -D-Gal(1 \rightarrow 3) β -D-MeGal. (A) cMD simulation for 30 ns at 300°K; (B) sampling in reference replica of BP-REM simulation for 4 ns per replica at 300°K; (C) and (D) are respective maps for linkage α -L-Fuc(1 \rightarrow 2) β -D-MeGal and (E) snapshot of conformer of BGB belongs to region M1 of both linkages in BP-REM. Red dotted lines represent a possible hydrogen bond between those atoms.

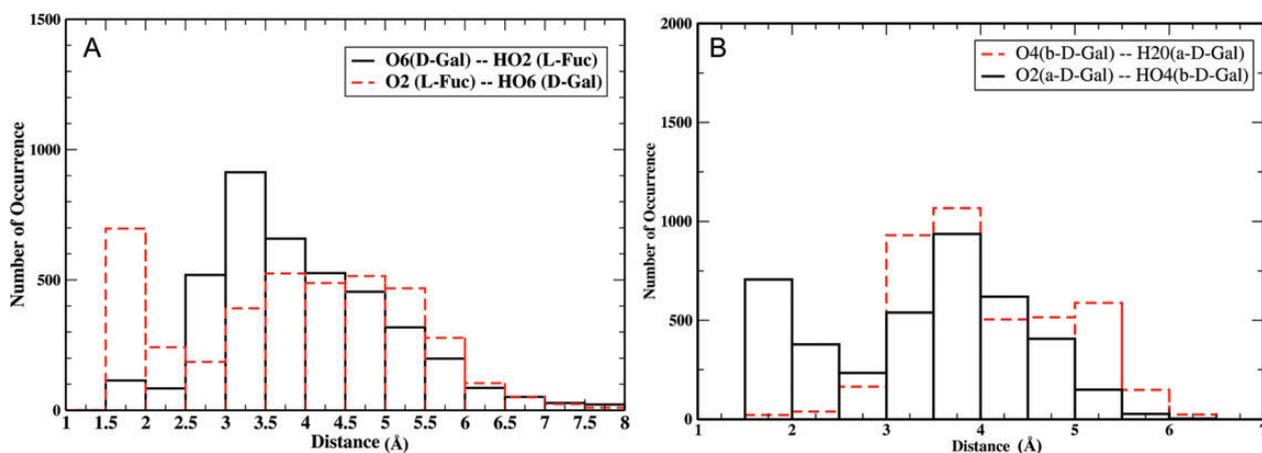


Fig. 6. Histogram of intramolecular hydrogen bond distances in BGB measured from BP-REM trajectory of first replica between hydroxyls: (A) O2 (α -L-Fuc) and O6 (α -D-Gal) and (B) O2 (α -D-Gal) and O4 (β -D-MeGal).

dominated by one type of monomer ring conformation (Supplementary data, Figure S8 and Table S1). Overall, BP-REM simulations in explicit solvent showed a significant improvement compared with previous simulations in implicit solvent, suggesting that BGB is not completely rigid in structure. A good agreement was found between the sampling of the Φ and Ψ torsion angles in solution by NMR (Azurmendi and Bush 2002) and the present BP-REM.

Le^x and Le^a trisaccharides

Initial conformational studies of Lewis oligosaccharides were first reported in the early 1980s where the relatively rigid behavior of these saccharides was shown through NMR and molecular modeling (Thøgersen et al. 1982). There were many later efforts to study the conformational behavior of Lewis oligosaccharides and the results indicated the existence of only a small group of closely related conformations (Biswas and Rao

1982; Yuriev et al. 2005). An interesting observation relating the dynamically behavior of Le^a and two low-energy conformations of Le^x were found during in vacuo simulation studies by Bush et al. (Mukhopadhyay and Bush 1991; Miller et al. 1992). Some other later modeling studies also supported the idea that the glycosidic linkage of Le^x and Le^a is not as rigid as indicated by earlier NMR studies (Yuriev et al. 2005).

To provide a general understanding of the conformation of and Le^x and Le^a epitopes, we analyzed the relative orientation of the L-Fuc and D-Gal rings from GlcNAc in the form of Φ and Ψ dihedral angles. The torsion angles Φ and Ψ from the cMD and BP-REMD simulations of Le^a and Le^x were monitored to fully determine its conformation and are presented in Figure 7. An inspection of these plots showed a good concordance among low-energy structures and low values of the metric function obtained from the orientation tensor measured by NMR (Azurmendi et al. 2002). The free-energy maps obtained from BP-REMD had the same aspect as the one previously measured: the main low-energy conformer of Le^x consisted of minima ($\Phi = -68.5 \pm 8$; $\Psi = -110.6 \pm 7.5$ between α -L-Fuc (1 \rightarrow 4)Me- β -D-GlcNAc, and $\Phi = -70.3 \pm 8.8$; $\Psi = 141.3 \pm 7.1$ between β -D-Gal(1 \rightarrow 3)Me- β -D-GlcNAc linkage) labeled M1 in Figure 7B and D (Yuriev et al. 2005). The standard deviation of the Φ and Ψ dihedral angles in both glycosidic linkages shows that Le^x is quite rigid and only one narrow free-energy minimum for both the glycosidic linkages was seen in the cMD simulation (Table I).

To obtain a more general characterization of the conformation of the Lewis epitopes, we analyzed the dynamics of the

molecule and captured the relative orientation of the L-Fuc and D-Gal rings. At the global minimum M1, L-Fuc was positioned parallel to the top of D-Gal and almost always tightly stacked in a rigid conformation. In this conformation, mainly non-polar interactions were possible between these units (Figure 8A). Moreover, a histogram of the intramolecular hydrogen bond showed that hydroxyl O2 of L-Fuc was able to establish a weak hydrogen bond between hydroxyl O2 of L-Fuc and atoms O2N and O3 of the GlcNAc, similar to what was observed in a recent spectroscopic study by Su et al. (2009; Figure 9A). The rest of the hydroxyl groups in Le^x mainly interact with solvent molecules. The BP-REMD simulation of Le^x provided an enhanced sampling compared with cMD and another possible conformation of the Le^x was observed (M2), where both the L-Fuc and D-Gal changed its orientation and displayed an inter-ring linkage between the HO6 (D-Gal) and O6 (GlcNAc) atoms (Figure 8B). The free energy of conformer M2 is ~ 3 kcal/mol larger than that of conformer M1, which agrees with the trend obtained by ab initio calculation (Csonka et al. 2000).

The average dihedral angle values were in good agreement with previous studies, where two or more low-energy conformers of Le^x were found (Miller et al. 1992; Yuriev et al. 2005). A clear distinction was seen between the intrinsic conformational preference of Le^x in vacuum and in the presence of other environments (water or protein; Su et al. 2009). Because it provides an advantage over existing computational approaches by including an explicit solvent, BP-REMD could thus be one of the best suited approaches for the conformational sampling of such saccharides.

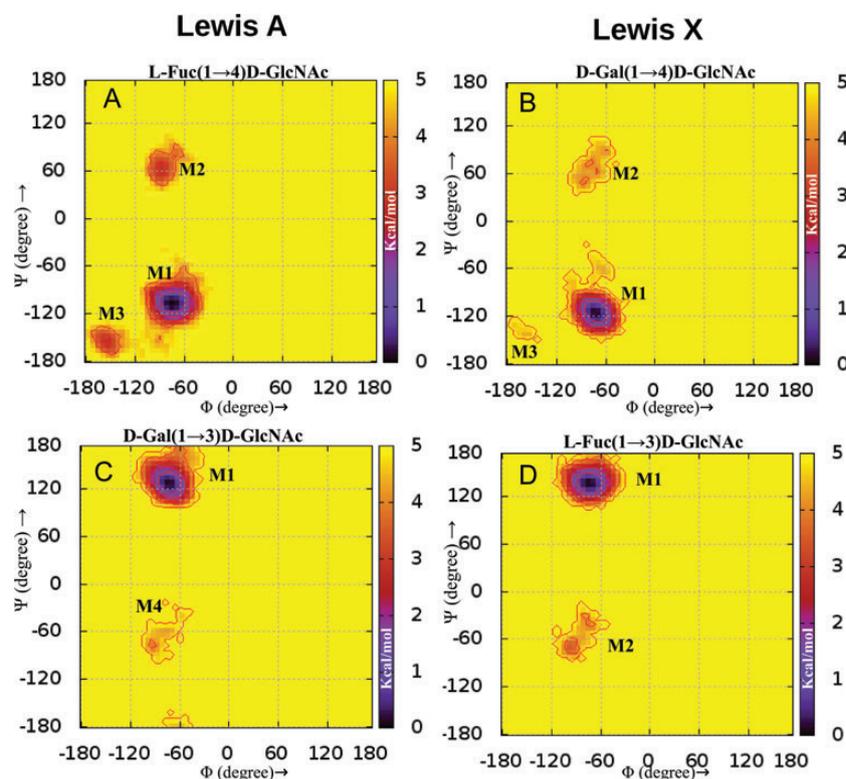


Fig. 7. Free-energy maps obtained from probability distribution of Φ and Ψ dihedral angles in reference replica of BP-REMD simulation: (A) sampling along (1 \rightarrow 4) of Le^a ; (B) sampling along (1 \rightarrow 4) of Le^x ; (C) sampling along (1 \rightarrow 3) of Le^a and (D) sampling along (1 \rightarrow 3) of Le^x .

Table I. Average values and standard deviation of Φ and Ψ glycosidic dihedral angles from BP-REMD simulations compared with cMD simulations and previous studies

Epitope	Ψ		Φ		Method	Reference	
	α -D-Gal(1 \rightarrow 3) β -D-MeGal	Ψ	α -L-Fuc(1 \rightarrow 2) β -D-MeGal	Ψ			
BGB	61.1 \pm 9.2	65.6 \pm 19.4	-72.6 \pm 15.6	-102.8 \pm 40.1	BP-REMD (M1)		
	102.8 \pm 10.5	74.3 \pm 23.4	-71.5 \pm 13.7	-92.3 \pm 44.9	BP-REMD (M2)		
	103.4 \pm 13.3	90.4 \pm 45.0	-76.5 \pm 15.7	86.5 \pm 14.6	BP-REMD (M3)*		
	57	64.6	-68.1	-91.8	X-ray	Otter et al. (1999)	
	57	58	-70	-94	NMR	Otter et al. (1999)	
	55	70	-65	-100	HSEA (1 ^a)	Thøgersen et al. (1982)	
	58	68	-69	-97	NMR (Fam. 1)	Azurmendi and Bush (2002)	
	68	81	-81	-156	NMR (Fam. 2)	Azurmendi and Bush (2002)	
	53	60	-73	-136	NMR (TRAMITE)	Azurmendi and Bush (2002)	
	59	70	-66	-94	NMR (Oriented)	Azurmendi and Bush (2002)	
	Le ^x	α -L-Fuc(1 \rightarrow 3) β -D-GlcNAc		β -D-Gal(1 \rightarrow 4)] β -D-GlcNAc		BP-REMD (M1)	
		-70.3 \pm 8.8	141.3 \pm 7.1	-68.5 \pm 8.0	-110.6 \pm 7.5	BP-REMD (M2)*	
		-84.8 \pm 11.9	-54 \pm 16.2	-74.4 \pm 11.4	71 \pm 15.6	BP-REMD (M3)*	
-59.1 \pm 10.6		150.2 \pm 6.4	-156.9 \pm 8.4	-134.6 \pm 6.9	cMD (M1)		
-70.3 \pm 8.7		141.2 \pm 7.1	-68.7 \pm 7.9	-110 \pm 7.3	NMR (HESA)	Thøgersen et al. (1982)	
-65		145	-65	-110	NMR (NOSEY)	Miller et al. (1992)	
-45		145	-55	-120	MD (CHARMm) conf. A	Miller et al. (1992)	
-45		144	-57	-118	MD (CHARMm) conf B.	Miller et al. (1992)	
-83		143	-65	-108	MD Conf A (CHARMm)	Miller et al. (1992)	
-59		152	-61	-112	MD Conf B (CHARMm)	Miller et al. (1992)	
-64		152	-64	-112	CS (MM3)	Imberty et al. (1995)	
-81.3		151	-74.9	-104	X-Ray	Pérez et al. (1996)	
-76.4		138.3	-70.4	-106	NMR (RDC)	Azurmendi et al. (2002)	
-73.1		149.5	-66.3	-105			
Le ^a		α -L-Fuc(1 \rightarrow 4) β -D-GlcNAc		β -D-Gal(1 \rightarrow 3)] β -D-GlcNAc		BP-REMD (M1)	
	-70.5 \pm 8.9	-99.8 \pm 19.4	-69.7 \pm 7.8	131.8 \pm 13.1	BP-REMD (M2)*		
	-82 \pm 9.4	69.3 \pm 11.3	-61.8 \pm 12.1	134.6 \pm 42.7	BP-REMD (M3)*		
	-150.4 \pm 8.4	-149.6 \pm 8.1	-73.6 \pm 10.9	134.5 \pm 8.9	BP-REMD (M4)*		
	-82.7 \pm 10.4	-88.4 \pm 76.8	-77.2 \pm 13.9	-81.3 \pm 42.5	cMD (M1)		
	-70.6 \pm 8.8	-101.8 \pm 7.1	-69.4 \pm 7.6	131.3 \pm 7.3	NMR (NOSEY)	Thøgersen et al. (1982)	
	-70	-100	-70/-80	140	MD (CHARMm)	Mukhopadhyay and Bush (1991)	
	-71	-97.7	-72	137	MD, SA (AMBER)	Kogelberg and Rutherford (1994)	
	-77	-99	-73	134	CS (MM3)	Imberty et al. (1995)	
	-79.4	-97.7	-75.4	144.1	NMR (RDC)	Azurmendi et al. (2002)	
	-76.9	-98.2	-70	141.2			

It can be seen that BP-REMD enables transitions from the global minimum, which were not usually seen in previous studies and thus enhances the conformational sampling of the saccharide. The average Φ and Ψ values of minimum energy conformations from BP-REMD simulations are in good agreement with the reported values from calculation and experiment. Standard deviation of Φ and Ψ dihedral angles sampled during cMD and BP-REMD simulation shows that Le^a and Le^x form a quite rigid structure in their stacked conformation. The glycosidic dihedral angles are defined as $\Phi = O5-C1-Ox-Cx$ and $\Psi = O5-C1-Ox-Cx+1$. Data from the literature reported in terms of the dihedral angles $\Phi = HC1-C1-O1-Cn'$ and $\Psi = C1-O1-Cn'-HCn'$ and/or $\Phi = O5-C1-Ox-Cx$ and $\Psi = Cx-1-Cx-Ox-C1$ were recalculated to Φ and Ψ values (used in this study) to enable a uniform comparison. CS, conformational search; SA, simulated annealing; RDC, residual dipolar coupling. All the values of Φ and Ψ are in degrees.

*Unique conformations that were not sampled during 30 ns cMD but were sampled by BP-REMD simulations.

Le^x and Le^a are isomers in which only the glycosidic linkage of the L-Fuc and D-Gal rings to GlcNAc are swapped. Similarly, the cMD simulation of Le^a was completely trapped in its global minimum ($\Phi = -70.671 \pm 8.8$; $\Psi = -101.8 \pm 7.1$ between α -L-Fuc(1 \rightarrow 4)Me- β -D-GlcNAc and $\Phi = -69.4 \pm 7.6$; $\Psi = 131.3 \pm 7.3$ between β -D-Gal(1 \rightarrow 3)Me- β -D-GlcNAc), but BP-REMD was able to sample a few other possible conformations (Figure 7 and Table I). At its global minimum, the Le^a gave a similar stacked orientation of L-Fuc and D-Gal as was seen for Le^x, except the orientation of the GlcNAc ring was flipped by 180° along its principal axis, so that the N-acetyl group of Le^a extended into the same region as the CH₂OH of the Le^x (Figure 10A). This was in agreement with the previous studies (Azurmendi et al. 2002; Yuriev et al. 2005). In this conformer, the hydroxyl O2 of D-Gal, instead of the L-Fuc in Le^x, formed polar interactions with the O2N and O3 oxygen of the

GlcNAc residues (Figure 9B). Thus, it was assumed that such a structure of Le^a and Le^x, in which most of the hydroxyl groups are free to form a network of hydrogen bonds with the polar side chains of the protein, enabled them to exhibit specific and selective binding with lectins and/or proteins.

This contrasts sharply with the other three conformations observed only in the BP-REMD simulation of Le^a, where either L-Fuc or D-Gal or both changed their orientation and established a ring-ring hydrogen bond between them (Figure 10B-D). A significant preference was seen for one conformation of Le^a (>95%) and Le^x (>99%), supporting the finding that Le^x was more rigid than Le^a (Imberty et al. 1995). In the conformation shown in Figure 10B, L-Fuc flips by 180° and the O2 hydroxyl forms a polar interaction with the O5 of the D-Gal. In another confirmation, the L-Fuc is accommodated in such a way that the O4 hydroxyl of D-Gal forms a polar interaction with the O5 of

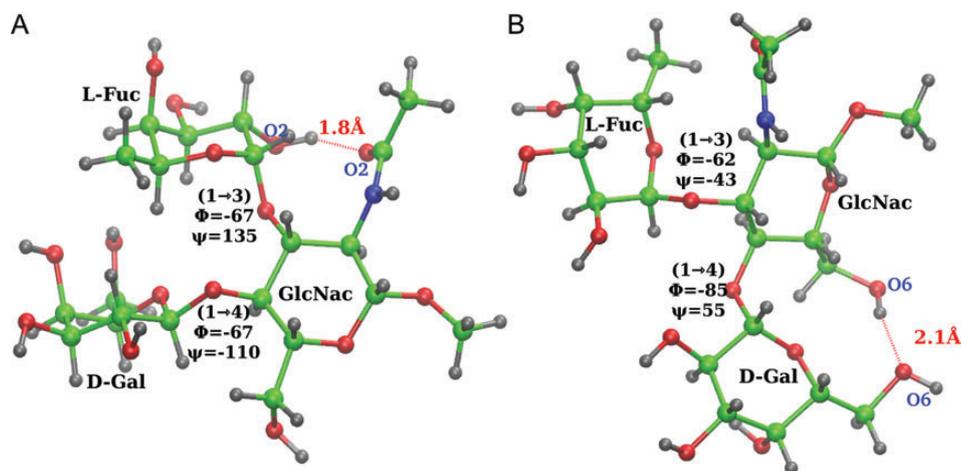


Fig. 8. Graphical representation of main low-energy conformations of Le^x trisaccharide: (A) tightly stacked and rigid conformation which was mainly seen in simulation and (B) one of conformers occasionally seen, which do not exhibit stacking like minimum low-energy conformation.

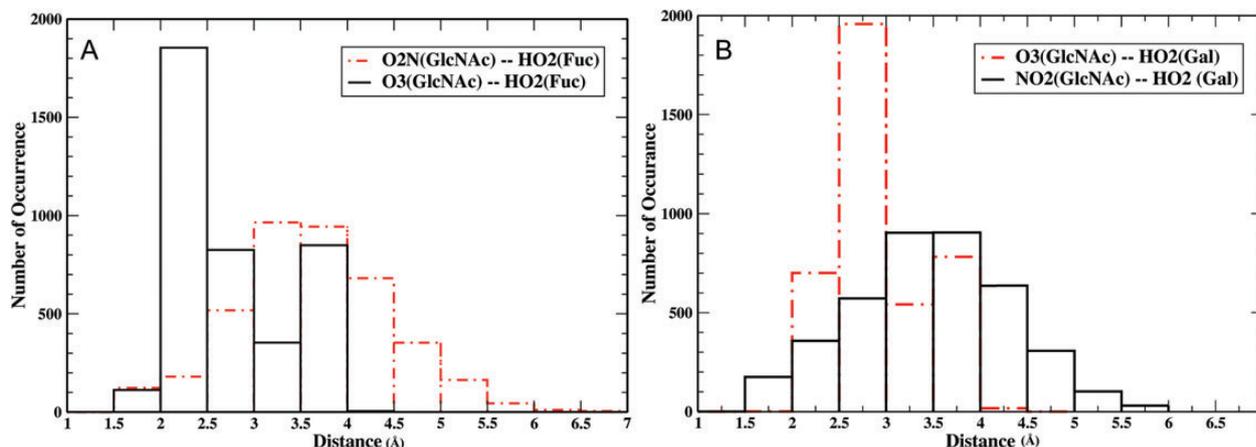


Fig. 9. Histogram of intramolecular hydrogen bond distances measured from BP-REMD trajectory of first replica: (A) hydrogen bond between hydroxyl O2 (L-Fuc) and O2N/O3 (GlcNAc) in Le^x and (B) hydrogen bond between hydroxyl O2 (D-Gal) and O2N/O3 (GlcNAc) in Le^a .

the L-Fuc (Figure 10C). In the fourth observed conformation, D-Gal also moved along the 1→3 linkage and hydroxyls O2 and O6 of D-Gal exhibited a polar interaction with the O5 atom of L-Fuc and O3N of GlcNAc, respectively (Figure 10D).

It was realized that Lewis determinants generally maintain a similar conformation in their free and bound state and hydroxyl O3 and O4 of D-Gal were also important in its binding to lectin (Yuriev et al. 2005). In our simulations, the O3 and O4 of D-Gal did not participate in any intramolecular interaction and were always facing toward the surface; thus, it is reasonable to assume that these are energetically less favorable conformations that may be much less populated in solution but could also be selected for in molecular recognition. We extended cMD simulations of Le^x and Le^a up to 1 μ s, and it was seen that even this time scale is insufficient to sample some conformations which can be seen in the BP-REMD simulations and correlate well with experimental studies (Supplementary data, Figure S5). Similar to the other cases, the sugar pucker conformations showed only rare transitions to alternative states (Supplementary data, Figures S9 and S10).

Conclusions

A biasing potential-based Hamiltonian replica-exchange MD simulation in explicit TIP3P water was performed that specifically promotes dihedral transition along the glycosidic linkage but still preserves canonical sampling of possible conformational states. Application of the method to di- and trisaccharides was primarily assessed by comparing the Φ and Ψ dihedral angles sampled in cMD and BP-REMD simulations with the Φ and Ψ dihedral angles sampled using ABF simulations or NMR experiments. Since the computationally demanding ABF approach is limited by the dimension of the scan and is not suitable for trisaccharides, Φ and Ψ dihedral angles of trisaccharides sampled by BP-REMD and cMD were thus only compared with available experimental measurements. Good agreement between the BP-REMD and ABF simulations was obtained, suggesting that the current set up of BP-REMD can be extended and applied to larger oligosaccharides. We further investigated the conformational behavior of the BGB, Le^x and

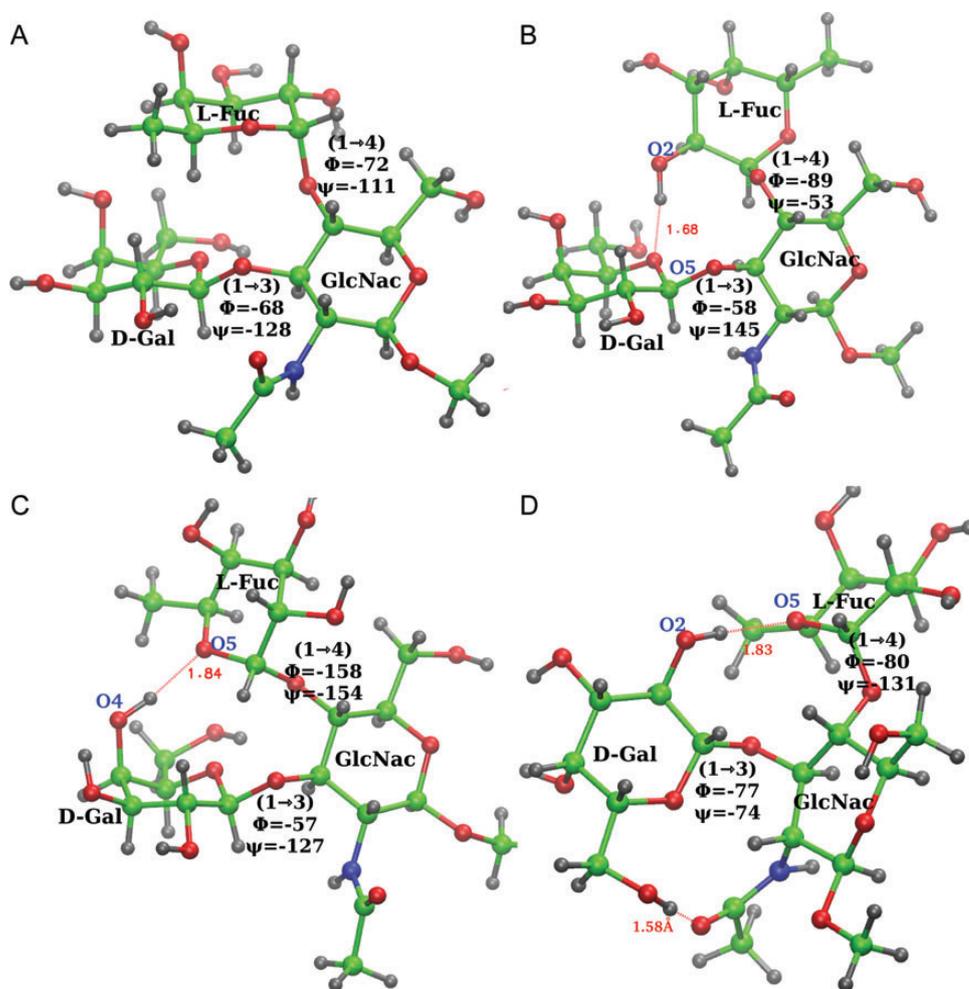


Fig. 10. Graphical representation of main low-energy conformations of Le^a trisaccharide: (A) tightly stacked and rigid conformation which was mainly seen in simulations and (B–D) conformations seen occasionally, which exhibit inter-ring linkage.

Le^a trisaccharides by the GLYCAM06 force field, using BP-REMD simulations and could limit the required computational resources. Succinctly, the protocol involved the following steps: (i) perform a high-temperature, implicit-solvent MD simulation to compute the approximate positions of local free-energy minima; (ii) use this information to set up a Hamiltonian replica-exchange simulation which allows transitions at those regions and (iii) create 2D Ramachandran-type free-energy profiles. It should be emphasized that the exact placement of the dihedral biasing potentials to promote transitions is not critical for the canonical sampling in the replica that runs under the control of the original force field (no biasing) and which was used for analysis of the sampling data. However, it might also be possible to extend the approach by checking the stable points of the Φ and Ψ dihedral angles during the explicit solvent BP-REMD and adjust and optimize the biasing potential accordingly. This will be explored in future simulation studies.

Simulation results on the trisaccharides using BP-REMD simulations were in good agreement with available X-ray, NMR and other spectroscopic measurements, suggesting that

there are no major flaws in the GLYCAM06 force field. The conformational analysis provided insights into the structural and conformational behavior of trisaccharides. BGB existed in a number of possible conformational states which demonstrated their conformational flexibility. Whereas the Le^x and Le^a trisaccharides were found to be quite rigid, Le^a was seen to be slightly more flexible than Le^x , but both saccharides existed mainly in one conformation, where L-Fuc and D-Gal were stacked on the top of each other. There were no inter-ring hydrogen bonds seen in this stacked conformer. This contrasted sharply with the other conformers, which display a weak hydrogen bond between the GlcNac and L-Fuc (NH—OH₂ or O₃—HO₂) residues. Therefore, non-polar interaction seemed to be responsible for the conformational preference of Le^x and Le^a . However, certain polar interactions were seen, albeit such interactions were not always strong and stable, possibly due to other conformations of Le^x . A clear distinction between the intrinsic conformational preference of Le^x in “in vacuo” and in a solvated, hydrated crystalline and protein-bound environment was recently found by Su et al. (2009), where the conformation of the Le^x antigen was determined by IR-UV ion dip

spectroscopy.⁵³ Theoretical predictions using BP-REMD are well suited to investigate the intrinsic conformational preference of BGB, Le^x and Le^a trisaccharides compared with existing approaches, which either exclude solvent atoms explicitly or are hampered by poor conformational sampling or high computational cost. In the future, this approach can be used in carbohydrate-based drug design and/or vaccine development processes, where information on the intrinsic conformational preference of large saccharides is required.

In summary, the proposed BP-REMD protocol adds a new dimension to existing approaches to the conformational sampling of saccharides by enabling: (i) an effective conformational sampling in the presence of explicit solvent, (ii) fast calculations by using fewer replicas than temperature-based replica-exchange simulations and (iii) the estimation of free-energy profiles along the glycosidic linkage. The computational cost of simulations is expected to be proportional to the square root of the number of dihedral linkages, instead of the total number of atoms in T-REMD. Theoretically, it should have no issues with extending the conformational sampling of saccharides to a protein-bound environment, but that will be attempted in our future work. It should also be emphasized that the BP-REMD approach can also be used to improve the sampling of possible ring conformations of each monosaccharide unit by including appropriate biasing potentials on the dihedral angles controlling the ring conformation. However, current force fields may not be accurate enough to realistically represent the ring conformational preferences (Gandhi and Mancera 2010) and we will therefore explore this possibility in future studies.

Methods

Test systems and MD simulation

Figure 1 shows a schematic representation of the saccharides with 1→2, 1→3 and 1→4 glycosidic linkages that are used as test systems in this study. The conventions proposed by the Commission on Nomenclature IUPAC (1989) and IUPAC-IUB (1997) are used throughout this paper. For (1→x) glycosidic linkages, the torsion angles that describe the orientation of two consecutive monosaccharide units are defined by: $\Phi = O5'-C1'-Ox-Cx$ and $\Psi = C1'-Ox-Cx-Cx+1$. The primed atoms correspond to the non-reducing residue. The initial structures of all the saccharides were generated using Glycam Biomolecule Builder (Woods 2005). Pyranose rings were set to their chair conformations. All cMD simulations were performed using the package AMBER11 (Case 2010a) with the GLYCAM06 (Kirschner et al. 2008) force field for carbohydrates.

Implicit solvent MD simulation. Initially, a 4-ns MD simulation of each saccharide at 600°K was performed using an advanced generalized born implicit solvent model (IGB = 5 in AMBER). Values of the Φ and Ψ dihedral angles sampled in this short simulation were extracted using the *ptraj* package of AmberTools, version 1.5, and used to set up biasing regions in the BP-REMD simulations. In order to check if the high-temperature simulations result in oversampling of irrelevant conformations especially alternative sugar pucker states that may also affect transitions of the glycosidic linkages, the sampled sugar puckers were compared with simulations in

explicit solvent (Supplementary data, Table S1). Small shifts of the mean sugar pucker parameters were observed which were, however, within the standard deviations of the sampled states (Supplementary data, Table S1), indicating broader sampling (as desired) but no significant deviation from the explicit solvent simulations.

cMD simulation. Saccharides were solvated in a box of 12 Å TIP3P water (Jorgensen et al. 1983) molecules. An ionic strength of 0.15 M was created by adding Na⁺ and Cl⁻ ions. The MD simulations were carried out using the package AMBER11 (Case 2010a) with the GLYCAM06 force field (Kirschner et al. 2008) and constant pressure periodic boundary conditions. The water and ions were first equilibrated so that the system was initially subjected to harmonic restraint (25 kcal/mol) and solvent molecules and ions were minimized. The system was then heated up to 300°K in steps of 100°K, followed by a gradual removal of the restraints and then finally a 0.5-ns equilibration without any restraints at 300°K. An integration time step of 2 fs was used together with the SHAKE algorithm (Ryckaert et al. 1977) to constrain the bonds connecting hydrogen atoms. The temperature was regulated using a Langevin thermostat (Nosé 1984) with a coupling constant of 2 ps⁻¹. The pressure was maintained at 1 bar using the Berendsen weak-coupling algorithm (Berendsen et al. 1984) with a pressure relaxation time of 1.2 ps. Electrostatic interactions were evaluated using the smooth Particle Mesh Ewald method (Essmann et al. 1995), with a non-bonded interaction cutoff of 10 Å. Coordinates were saved after each 1 ps. The *ptraj* package of AmberTools version 1.5 was used to calculate the Φ and Ψ dihedral angles from the MD trajectories.

Backbone BP-REMD for saccharides

A Hamiltonian replica-exchange simulation of an equilibrated system was performed to reduce the free-energy barriers and promote glycosidic dihedral transitions in the saccharides. Total seven replicas were run, including one reference replica without any biasing potential. Each replica was subjected to 4 ns of MD simulation under periodic boundary conditions. A 2-fs integration time step was used and the non-bonded cutoff was set to 9 Å. All of these simulations were performed in AMBER9 (Case 2010b) with the GLYCAM06 (Kirschner et al. 2008) force field for carbohydrates.

The BP-REMD method employs a 2D biasing potential for each Φ and Ψ dihedral angle pair that is involved in the linkage between two saccharide units following the approach described in previous article (Kara and Zacharias 2013). The form of the biasing potential is similar to a 2D Gaussian function that destabilizes a favorable regime in the coupled Φ and Ψ dihedral space. It is characterized by the center of the biasing potential and the width. The width of the biasing potential was 60° and the center of favorable regions was derived from the short implicit solvent simulations (see above paragraph). Typically, the biasing along the replica runs involved biasing three favorable regimes of the Φ and Ψ dihedral space. We follow the protocol and use the number of replicas that were used in simulations of nucleic acids, which were also optimal for the test cases in this study (Kara and Zacharias 2013). For replicas 1–7, maximum biasing potentials (E_{\max}) of 0, 1, 2, 3, 4, 5 and 6 kcal/mol were

applied along the Φ and Ψ angles of the glycosidic linkage, respectively. The biasing potentials in the replica runs destabilize favorable Φ and Ψ angle regions, promote transitions between conformational substates and also improve the sampling in the reference replica via exchanges with the reference replica. Each replica evolved independently and after each 1 ps time interval exchanges between neighboring replicas were attempted according to the following Metropolis criterion (Okamoto 2004):

$$w(x_i \rightarrow x_j) = 1 \text{ for } \Delta \leq 0 \quad (1)$$

$$w(x_i \rightarrow x_j) = e^{-\Delta} \text{ for } \Delta > 0 \quad (2)$$

where

$$\Delta = \beta[(E^j(r_j) - E^j(r_i)) - (E^i(r_j) - E^i(r_i))] \quad (3)$$

where E^j and E^i are the energy of neighboring configurations using the force field for replicas j and i , respectively. $\beta = 1/k_B T$ is the inverse thermal energy.

ABF calculation. For comparison purposes, adaptive bias MD simulations, with an adaptive biasing force (Darve and Pohorille 2001), were performed to explore the free-energy landscape of saccharides in water. Since ABF scans are limited by the dimension of scan, they are only feasible for the conformational sampling of disaccharides. ABF calculations were performed using the in-house software PMFLib (Kulhanek 2010), which uses the *pmemd* program of AMBER. The collective variables chosen for the scan were the Φ and Ψ dihedral angles along the glycosidic linkage of the disaccharides. The conditions of these simulations, unless specified otherwise, were almost the same as those obtained from cMD simulations.

Conformational sampling was performed in a vacuum as well as in the presence of water. Φ and ψ were handled as coupled variables covering the full range $[-180^\circ; +180^\circ]$ in 90 bins. In each case, 50 ns of simulation was sufficient to sample the free-energy space properly. Reconstruction of the complete free-energy landscape was achieved by the numerical integration of the 2D gradients.

Conformational clustering

To characterize the conformational states of carbohydrates, we generated Ramachandran plots of the glycosidic Φ and Ψ angles that display the conformational space of each linkage. Such plots were compared with conformations measured by NMR experiments. The package *ptraj* was used to calculate the Φ and Ψ dihedral angles from the MD trajectories for the systems. A 2D plot of the Φ and Ψ dihedral angles, called the conformational profile, shows the conformational space of the linkage. This 2D conformational profile was then used to calculate the probability (P) of there being conformations in the bins of 6° , and then the free-energy profile of the conformational space was obtained by the following equation (in the reference replica):

$$\Delta G = -RT \log(P) \quad (4)$$

Free-energy maps were set to $\Delta G = 0$ kcal/mol at the global minimum. The value of ΔG at grid points that were never visited during cMD simulation (formally infinite) was set to the maximal value ΔG_{\max} .

Supplementary Data

Supplementary data for this article is available online at <http://glycob.oxfordjournals.org/>.

Funding

This work was supported by the Ministry of Education of the Czech Republic (LH13055), the Grant Agency of the Czech Republic (13-25401S) and European Social Fund (CZ.1.07/2.3.00/30.0009). This work was further supported by the project “CEITEC—Central European Institute of Technology” (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund. MZ acknowledges support by the Deutsche Forschungsgemeinschaft (DFG) grant (Za-153/17-1). Access to the MetaCentrum computing facilities provided under the programme “Large Infrastructure Projects for Research, Development, and Innovation” (LM2010005) and CERIT-SC computing facilities provided by the CERIT Scientific Cloud programme center, part of the Operational Program Research and Development for Innovation (reg. no. CZ.1.05/3.2.00/08.0144) are appreciated.

Conflict of interest

None declared.

Abbreviations

ABF, adaptive biasing force; BGB, blood group B; BP-REMD, biasing potential-based replica-exchange molecular dynamics; cMD, classical MD; 3D, three-dimensional; D4, α -D-Neu5Ac (2 \rightarrow 6) β -D-Gal; D5, α -D-Glc(1 \rightarrow 6) β -D-Glc; Le^a, Lewis A; Le^x, Lewis X; MD, molecular dynamics.

References

- Almond A, Sheehan JK. 2003. Predicting the molecular shape of polysaccharides from dynamic interactions with water. *Glycobiology*. 13:255–264.
- Azumendi HF, Bush CA. 2002. Conformational studies of blood group A and blood group B oligosaccharides using NMR residual dipolar couplings. *Carbohydr Res*. 337:905–915.
- Azumendi HF, Martin-Pastor M, Bush CA. 2002. Conformational studies of Lewis X and Lewis A trisaccharides using NMR residual dipolar couplings. *Biopolymers*. 63:89–98.
- Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. 1984. Molecular dynamics with coupling to an external bath. *J Chem Phys*. 81:3684–3690.
- Biswas M, Rao VSR. 1982. Conformational studies on the ABH and Lewis blood group oligosaccharides. *Carbohydr Polymers*. 2:205–222.
- Cagas P, Bush CA. 1990. Determination of the conformation of Lewis blood group oligosaccharides by simulation of two-dimensional nuclear overhauser data. *Biopolymers*. 30:1123–1138.
- Case DA, Darden TA, Cheatham III TE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, et al. 2010a. *AMBER 11*. University of California.

- Case DA, Darden TA, Cheatham III TE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, et al. 2010b. *AMBER 09*. University of California.
- Choi Y, Kim H, Cho KW, Paik SR, Kim H-W, Jeong K, Jung S. 2009. Systematic probing of an atomic charge set of sialic acid disaccharides for the rational molecular modeling of avian influenza virus based on molecular dynamics simulations. *Carbohydr Res*. 344:541–544.
- Csonka GI, Sosa CP, Csizmadia IG. 2000. Ab initio study of lowest-energy conformers of Lewis X (Lex) trisaccharide. *J Phys Chem A*. 104:3381–3390.
- Curuku J, Zacharias M. 2009. Enhanced conformational sampling of nucleic acids by a new Hamiltonian replica exchange molecular dynamics approach. *J Chem Phys*. 130:104110.
- Darve E, Pohorille A. 2001. Calculating free energies using average force. *J Chem Phys*. 115:9169–9183.
- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. 1995. A smooth particle mesh Ewald method. *J Chem Phys*. 103:8577–8593.
- Feizi T, Ernst B, Hart GW, Sinaý P. 2008. *Glycobiology of AIDS Carbohydrates in Chemistry and Biology*. Weinheim, Germany: Wiley. p. 851–866.
- Feizi T, Mulloy B. 2003. Carbohydrates and glycoconjugates. *Glycomics: The new era of carbohydrate biology*. *Curr Opin Struct Biol*. 13:602–604.
- Gama CI. 2009. Understanding the chemical basis of neuronal development and communication: I. The role of fucose alpha (1-2) galactose carbohydrates in neuronal growth II. Structure-function analysis of chondroitin sulfate in the brain. PhD Dissertation, California Institute of Technology.
- Gandhi NS, Mancera RL. 2008. The structure of glycosaminoglycans and their interactions with proteins. *Chem Biol Drug Des*. 72:455–482.
- Gandhi NS, Mancera RL. 2009. Free energy calculations of glycosaminoglycan-protein interactions. *Glycobiology*. 19:1103–1115.
- Gandhi NS, Mancera RL. 2010. Can current force fields reproduce ring puckering in 2-O-sulfo-alpha-L-iduronic acid? A molecular dynamics simulation study. *Carbohydr Res*. 345:689–695.
- Gandhi NS, Mancera RL. 2011. Molecular dynamics simulations of CXCL-8 and its interactions with a receptor peptide, heparin fragments, and sulfated linked cyclitols. *J Chem Inf Model*. 51:335–358.
- Geijtenbeek TB, Torensma R, van Vliet SJ, van Duijnhoven GC, Adema GJ, van Kooyk Y, Figdor CG. 2000. Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell*. 100:575–585.
- Hardy BJ, Sarko A. 1993. Conformational analysis and molecular dynamics simulation of cellobiose and larger cellooligomers. *J Comput Chem*. 14:831–847.
- Imberty A, Gerber S, Tran V, Pérez S. 1990. Data bank of three-dimensional structures of disaccharides, a tool to build 3-D structures of oligosaccharides. *Glycoconj J*. 7:27–54.
- Imberty A, Mikros E, Koca J, Mollicone R, Oriol R, Pérez S. 1995. Computer simulation of histo-blood group oligosaccharides: Energy maps of all constituting disaccharides and potential energy surfaces of 14 ABH and Lewis carbohydrate antigens. *Glycoconj J*. 12:331–349.
- Imberty A, Tran V, Pérez S. 1990. Relaxed potential energy surfaces of N-linked oligosaccharides: The mannose- α (1 \rightarrow 3)-mannose case. *J Comput Chem*. 11:205–216.
- Javoroni F, Ferreira ABB, da Silva CO. 2009. The (alpha-1,6) glycosidic bond of isomaltose: A tricky system for theoretical conformational studies. *Carbohydr Res*. 344:1235–1247.
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 79:926–935.
- Kannan S, Zacharias M. 2007. Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential. *Proteins*. 66:697–706.
- Kannan S, Zacharias M. 2009. Folding simulations of Trp-cage mini protein in explicit solvent using biasing potential replica-exchange molecular dynamics simulations. *Proteins*. 76:448–460.
- Kannan S, Zacharias M. 2010. Application of biasing-potential replica-exchange simulations for loop modeling and refinement of proteins in explicit solvent. *Proteins*. 78:2809–2819.
- Kara M, Zacharias M. 2013. Influence of 8-oxoguanosine on the fine structure of DNA studied with biasing-potential replica exchange simulations. *Biophys J*. 104:1089–1097.
- Karlsson KA. 1999. Bacterium-host protein-carbohydrate interactions and pathogenicity. *Biochem Soc Trans*. 27:471–474.
- Kirschner KN, Yongye AB, Tschappel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ. 2008. GLYCAM06: A generalizable biomolecular force field. *Carbohydrates*. *J Comput Chem*. 29:622–655.
- Koča J. 1994. Computer program CICADA—travelling along conformational potential energy hypersurface. *J Mol Struct THEOCHEM*. 308:13–24.
- Kogelberg H, Rutherford TJ. 1994. Studies on the three-dimensional behaviour of the selectin ligands Lewis(a) and sulphated Lewis(a) using NMR spectroscopy and molecular dynamics simulations. *Glycobiology*. 4:49–57.
- Kony D, Damm W, Stoll S, Hünenberger PH. 2004. Explicit-solvent molecular dynamics simulations of the β (1 \rightarrow 3)- and β (1 \rightarrow 6)-linked disaccharides β -laminarabiose and β -gentiobiose in water. *J Phys Chem B*. 108:5815–5826.
- Kulhanek P. 2010. *PMFLib*. NCBR. Brno, Czech Republic: Masaryk University.
- Lloyd KO. 2000. The chemistry and immunochemistry of blood group A, B, H, and Lewis antigens: Past, present and future. *Glycoconj J*. 17:531–541.
- Luttkete T. 2004. Carbohydrate Structure Suite (CSS): Analysis of carbohydrate 3D structures derived from the PDB. *Nucleic Acids Res*. 33:D242–D246.
- Lütke T, von der Lieth C-W. 2004. pdb-care (PDB Carbohydrate RESidue check): A program to support annotation of complex carbohydrate structures in PDB files. *BMC Bioinformatics*. 5:69.
- Miller KE, Mukhopadhyay C, Cagas P, Bush CA. 1992. Solution structure of the Lewis x oligosaccharide determined by NMR spectroscopy and molecular dynamics simulations. *Biochemistry*. 31:6703–6709.
- Morelli L, Poletti L, Lay L. 2011. Carbohydrates and immunology: Synthetic oligosaccharide antigens for vaccine formulation. *Eur J Org Chem*. 2011:5723–5777.
- Mukhopadhyay C, Bush CA. 1991. Molecular dynamics simulation of Lewis blood groups and related oligosaccharides. *Biopolymers*. 31:1737–1746.
- Murata K, Egami H, Shibata Y, Sakamoto K, Misumi A, Ogawa M. 1992. Expression of blood group-related antigens, ABH, Lewis(a), Lewis(b), Lewis(x), Lewis(y), CA19-9, and CSLEX1 in early cancer, intestinal metaplasia, and uninvolved mucosa of the stomach. *Am J Clin Pathol*. 98:67–75.
- Nahmany A, Strino F, Rosen J, Kemp GJL, Nyholm P-G. 2005. The use of a genetic algorithm search for molecular mechanics (MM3)-based conformational analysis of oligosaccharides. *Carbohydr Res*. 340:1059–1064.
- Naidoo KJ, Brady JW. 1997. The application of simulated annealing to the conformational analysis of disaccharides. *Chem Phys*. 224:263–273.
- Nishihara S, Iwasaki H, Kaneko M, Tawada A, Ito M, Narimatsu H. 1999. α 1,3-Fucosyltransferase 9 (FUT9; Fuc-TIX) preferentially fucosylates the distal GlcNAc residue of polylactosamine chain while the other four α 1,3FUT members preferentially fucosylate the inner GlcNAc residue. *FEBS Lett*. 462:289–294.
- Nosé S. 1984. A unified formulation of the constant temperature molecular dynamics methods. *J Chem Phys*. 81:511–519.
- Okamoto Y. 2004. Generalized-ensemble algorithms: Enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J Mol Graph Model*. 22:425–439.
- Ohrui H, Nishida Y, Watanabe M, Hori H, Meguro H. 1985. ¹H-NMR studies on (6R)- and (6S)-deuterated (1-6)-linked disaccharides: Assignment of the preferred rotamers about C5-C6 bond of (1-6)-disaccharides in solution. *Tetrahedron Lett*. 26:3251–3254.
- Otter A, Lemieux RU, Ball RG, Venot AP, Hindsgaul O, Bundle DR. 1999. Crystal state and solution conformation of the B blood group trisaccharide α -L-Fucp-(1 \rightarrow 2)-[α -D-Galp]-[1 \rightarrow 3]- β -D-Galp-OCH₃. *Eur J Biochem*. 259:295–303.
- Pereira CS, Kony D, Baron R, Müller M, van Gunsteren WF, Hünenberger PH. 2006. Conformational and dynamical properties of disaccharides in water: A molecular dynamics study. *Biophys J*. 90:4337–4344.
- Perić-Hassler L, Hansen HS, Baron R, Hünenberger PH. 2010. Conformational properties of glucose-based disaccharides investigated using molecular dynamics simulations with local elevation umbrella sampling. *Carbohydr Res*. 345:1781–1801.
- Pérez S, Mouhous-Riou N, Nifant'ev NE, Tsvetkov YE, Bacht B, Imberty A. 1996. Crystal and molecular structure of a histo-blood group antigen involved in cell adhesion: the Lewis x trisaccharide. *Glycobiology*. 6:537–542.
- Peters T, Meyer B, Stuike-Prill R, Somorjai R, Brisson JR. 1993. A Monte Carlo method for conformational analysis of saccharides. *Carbohydr Res*. 238:49–73.
- Poppe L. 1993. Modeling carbohydrate conformations from NMR data: Maximum entropy rotameric distribution about the C5-C6 bond in gentiobiose. *J Am Chem Soc*. 115:8421–8426.

- Poppe L, Stuike-Prill R, Meyer B, Halbeek HV. 1992. The solution conformation of sialyl- α (2 \rightarrow 6)-lactose studied by modern NMR techniques and Monte Carlo simulations. *J Biomol NMR*. 2:109–136.
- Re S, Nishima W, Miyashita N, Sugita Y. 2012. Conformational flexibility of N-glycans in solution studied by REMD simulations. *Biophys Rev*. 4:179–187.
- Roën A, Mayato C, Padrón JI, Vázquez JT. 2008. Conformational domino effect in saccharides. 2. Anomeric configuration control. *J Org Chem*. 73:7266–7279.
- Roën A, Padrón JI, Mayato C, Vázquez JT. 2008. Conformational domino effect in saccharides: A prediction from alkyl β -(1 \rightarrow 6)-diglucopyranosides. *J Org Chem*. 73:3351–3363.
- Ryckaert J-P, Ciccotti G, Berendsen HJ. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J Comput Phys*. 23:327–341.
- Sakamoto J, Furukawa K, Cordon-Cardo C, Yin BW, Rettig WJ, Oettgen HF, Old LJ, Lloyd KO. 1986. Expression of Lewis A, Lewis B, X, and Y blood group antigens in human colonic tumors and normal tissue and in human tumor-derived cell lines. *Cancer Res*. 46:1553–1561.
- Sugita Y, Okamoto Y. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett*. 314:141–151.
- Su Z, Wagner B, Cocinero EJ, Ernst B, Simons JP. 2009. The intrinsic conformation of a Lewis antigen: The Lewis \times trisaccharide. *Chem Phys Lett*. 477:365–368.
- Thøgersen H, Lemieux RU, Bock K, Meyer B. 1982. Further justification for the exo-anomeric effect. Conformational analysis based on nuclear magnetic resonance spectroscopy of oligosaccharides. *Can J Chem*. 60:44–57.
- Tiemeyer M, Goodman CS. 1996. Gliotectin is a novel carbohydrate-binding protein expressed by a subset of glia in the embryonic *Drosophila* nervous system. *Development*. 122:925–936.
- Vacquier VD, Moy GW. 1997. The fucose sulfate polymer of egg jelly binds to sperm REJ and is the inducer of the sea urchin sperm acrosome reaction. *Dev Biol*. 192:125–135.
- Woods RJ. 2005. *GLYCAM Web*. Athens, GA: Complex Carbohydrate Research Center, University of Georgia.
- Xia J, Daly RP, Chuang F-C, Parker L, Jensen JH, Margulis CJ. 2007. Sugar folding: A novel structural prediction tool for oligosaccharides and polysaccharides 2. *J Chem Theory Comput*. 3:1629–1643.
- Yuriev E, Farrugia W, Scott AM, Ramsland PA. 2005. Three-dimensional structures of carbohydrate determinants of Lewis system antigens: Implications for effective antibody targeting of cancer. *Immunol Cell Biol*. 83:709–717.