Conformational preferences of triantennary and tetraantennary hybrid *N*-glycans in solution: Insights from 20 μs long atomistic molecular dynamic simulations

Rajarshi Roy, Sayan Poddar, Parimal Kar *

Discipline of Biosciences and Biomedical Engineering, Indian Institute of Technology Indore, Khandwa Road, Madhya Pradesh 453552, India

*Corresponding Author: Parimal Kar | Email: parimal@iiti.ac.in | Phone: +91 731 2438700 (Ext. 550)

Abstract

N-linked glycans on protein surfaces play a crucial role in the recognition procedure as well as the functional process of it. In the current study, we have investigated the conformational dynamics of a triantennary and a tetraantennary hybrid *N*-glycans associated with the HIV glycoprotein using 20 μs long all-atom standard molecular dynamics (MD) simulations. We classify these long sequences in terms of corresponding glycosidic linkages, which samples several minima resulting in the multiple co-existing conformers of the same. In our simulations, all the puckering states ($^4C_1$ to $^1C_4$) of each monosaccharide except mannose are sampled. The change of sequences or the additional of an arm significantly restrict the sampling of the conformational space. Local conformation of each branch also changes upon the change of sequence. The end to end distance of the complex branch changes almost 10 Å in the presence of another branch. All these conformation changes involve the selective formation of inter-residue hydrogen bonds and water-mediated bonding networks. Thus, our study supports the conformer selection mechanism of N-glycans, which helps to understand the protein-glycan interaction. As these glycans show significant interaction with different broadly neutralizing antibodies (bNAbs), our work might help in designing the HIV vaccines.

Keywords: *N*-glycans, Molecular Dynamics, PCA, phi/psi, puckering angle

Introduction

Carbohydrate interactions are one of the most prevalent interactions at the cellular level[1]. More than ten types of monosaccharides are there in our cell, which forms several types of glycan depends on the variability of the linkages and the stereoisomers. The rigid sugar rings are attached via flexible glycosidic linkages having several linkage points make the glycan molecules a multi number co-existing conformer in the solutions[2, 3]. Along with this, several rotamers via those linkages and their potential acceptor and donor hydroxyl group for hydrogen bond increases glycan conformational complexity. All these glycan molecules are involved in several biological processes such as protein quality control[4, 5], protein trafficking[6, 7], protein stability[8-11], immune response[12], glycan-receptor recognition[2, 13-15] as well as in several commercial roles like biocompatible and biodegradable materials[16], design of human vaccine[17, 18] and other pharmaceutical uses[19, 20]. Most of these oligosaccharides are covalently attached to the nitrogen atom of specific asparagine residues in the endoplasmic reticulum (ER) through the N-glycosylation process[21]. These N-glycans are further modified by its length and sequence via the presence of several enzymes in Golgi and ER[5, 22], and exhibit several critical roles[23, 24]. To examine the glycan recognition by the viral pathogen[25, 26], or the recognition of viral glycoproteins via different lectin and antibodies specifically broadly neutralizing antibodies (bNAbs)[27-29] required a high level of structural knowledge of those glycan molecules which is not clear in our current scenario. Therefore, the characterization of structural and conformational variability of the different N-glycans sequence to address the virus-glycan interactions is the need of the hour.

Nowadays, molecular dynamics (MD) techniques are routinely used to investigate the structural dynamics of glycans. Several N-glycans are subjected to NMR spectroscopy for extensive analysis in their solution state[30-33] as well as in the receptor-bound state[34, 35]. However, these experimental techniques are not able to estimate the structure of long glycans because of high flexible linkages or the crowding of NMR spectra[15, 36-38]. On the contrary, MD simulation studies of oligosaccharides provide us with an atomistic structural detail of diverse glycans and give a precise match in several experimental properties[39-42]. With the advancement of the thoroughly parametrized molecular mechanical (MM) force field[43-48] for carbohydrate molecules, classical molecular dynamics (cMD) and enhanced sampling techniques are allowing us to explore several

structural features with higher confidence, which are absent in the experimental techniques. There are several cMD studies of a small fragment of glycans[49-53] as well as long sequences[54-56] in an explicit solvent model, which elucidates several transitions of the glycosidic angle and the sample of the different possible rotamer states. There are several enhanced sampling techniques such as temperature replica exchange MD (T-REMD)[39, 57-61], Hamiltonian replica exchange MD (H-REMD)[62], replica state exchange metadynamics (RSE-MTD)[63], accelerated MD (aMD)[64, 65] have been used to overcome the potential barrier between different conformer of glycans. Several conformers such as extended, half, and tight backfolded were speculated of a biantennary complex N-glycans using REMD technique[57] besides, the effect of bisecting and core fucosylation was also observed using the same[39]. Our previous study also explores a conformation change of Lewis Y (A-Penta saccharide), a blood group antigen, from to open to closed conformer using aMD technique[65]. Recently, the effect of fucosylation, sialyation, and galactosylation in the conformational space of different N-glycan sequences of IgG are also observed using cMD technique[66]. However, in the case of T-REMD, the number of replicas increases in proportional to the square root of the total number of atoms in the system to maintain a sufficient acceptance ratio between the replicas. It will be impossible to achieve the free random walk in the replica space in case of a more extensive system solvated in the explicit solvent model. To get rid of this complexity, we choose cMD for our N-glycan sequences which an unusual hybrid type sequences may exist in the surface of the human immunodeficiency virus (HIV) glycoprotein[67], with a very long timescale run to sample those inaccessible states separated by a relatively higher free energy barrier.

The main criteria for choosing the HIV-1 glycoprotein glycans named *NGlycan1* and *NGlycan2* shown in **Figure 1(A, B)** to characterize its conformational space as those are selectively recognized by different broadly neutralizing antibodies (bNAbs)[67]. In the process of HIV vaccine development, all those glycans may serve as an epitope for HIV vaccines[68]. HIV-1 envelope glycoprotein is a trimeric structure consists of gp120, and gp41 has a critical role in the recognition and the entry and the attachment to the host cells[69, 70]. Also, most of the HIV bNAbs interaction holds an information reservoir that can be used for vaccine and drug design against HIV infection[71, 72]. Interactions between bNAbs and HIV glycoprotein are mostly mediated through the glycans and the glycopeptide[73, 74]. So, the characterizations of the glycans present in the glycoprotein

surface in solution will further help to understand those interactions in more important details. The glycan sequences (*NGlycan1* and *NGlycan2*) which are used in our current study have more than two branches **(Figure 1(A, B))** and comprise of high mannose as well as the complex types of glycan makes a highly complex 3D dimensional structure **(Figure 1(B, D))** for its atomistic characterizations.

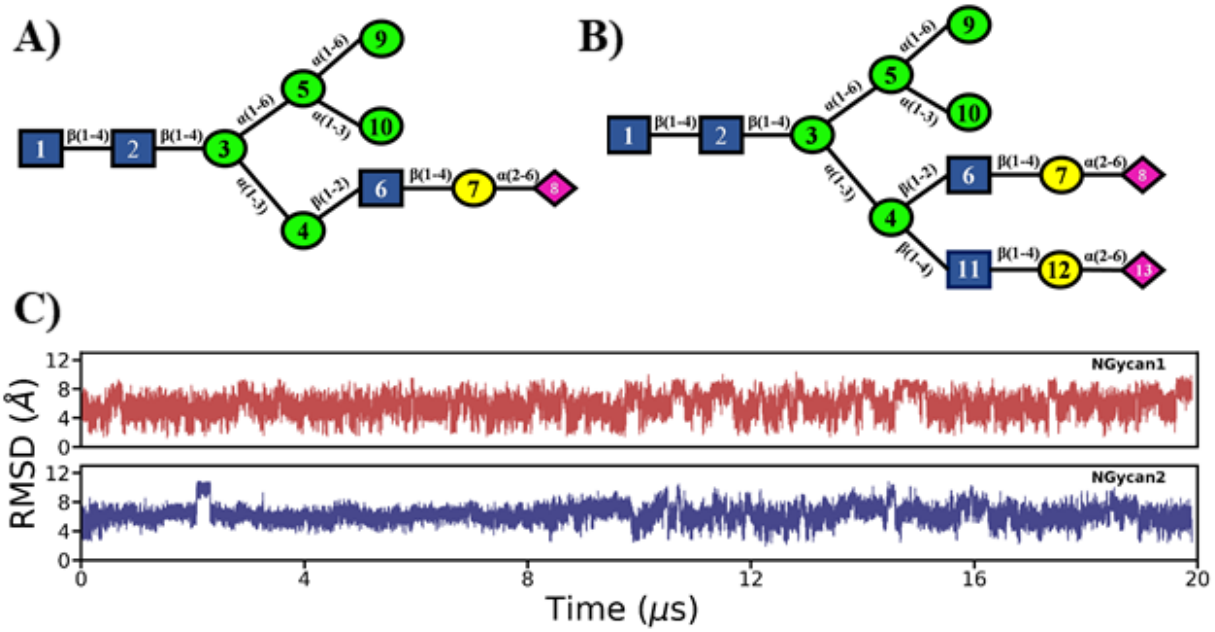

Figure 1: Sequence and structural representation of NGlycan1 and NGlycan2 along with the time evolution of the root-mean-square deviations (RMSDs) of the heavy atom with respect to the initial structure. In 2D representation, the color scheme is used as follows: N-acetyl glucosamine: blue, Mannose: green, Galactose: yellow and Sialic acid: pink. Sequence and nomenclature of A) NGlycan1 and C) NGlycan2 are shown in the line diagram. The time evolution of the 3D conformer is also shown using proper color (light to dark) scheme B) NGlycan1 and D) NGlycan2. E) Root mean square deviation (RMSD) of NGlycan1 (red) and NGlycan2 (Blue) up to 20 μs.

In our study, we conducted a 20 μs long cMD of *NGlycan1* and *NGlycan2* sequences, i.e., tri-antennary and tera-antennary shown in the **(Figure1)**, which are pro-ligands of HIV-1 broadly neutralizing antibodies[67], to fully characterizes its conformational dynamics. For convenience, we will discuss all the monosaccharides of both the sequences by its residue index (see **Figure 1(A, C)**) throughout our paper. As per our knowledge, it is the first computational study which tries to evaluate the conformer space of this amount of long and complex glycan chains. We compare the conformational preferences of these two chains in solution in the light of every local glycosidic

angle, puckering states of each monosaccharide, hydrogen bond (H-bond) interaction network as well as using different global parameters. So, our current study will not only help to understand its 3D space, additionally serve information reservoir to understand the interaction of HIV-glycoprotein and bNAbs.

Computational Methods

MD simulation of glycans

Both the glycans **(Figure 1 (A, B))** structure were modeled using the GLYCAM-Web server (www.glycam.org)[75] with the help of the linkage information from the Shivatare et al. [67]. Energy minimized rotamer structures from the GLYCAM-Web server have been considered for the starting structure of the MD simulation. MD simulation was carried out by the *pmemd.cuda* module[76, 77] of AMBER18[78, 79] with the glycam06j-1[43] carbohydrate forcefield for carbohydrate chain. All the molecules were solvated in a truncated octahedral water box having a 12 Å buffer from the wall by the TIP3P[80] water model. A suitable number of Na+ ions were added to neutralize the systems. Bonds involving with the hydrogen atoms were constrained using the SHAKE algorithm[81], and electrostatic interactions were calculated using the particle-mesh Ewald[82] method. The non bonded cut-off was set to 12 Å, and the temperature was kept fixed at 300k using the Langevin thermostat[83] with a collision frequency of 2 ps^{-1}.

Minimization was performed without restraint in two steps without and with the hydrogen bond vibration restriction. In the first step, steepest descent was run for 500 steps and then 500 steps of the conjugate gradient to minimize the system. The second minimization was also done for 1000 steps of steepest descent followed by 1000 steps of the conjugate gradient. Systems were then heated up to 300 k in a stepwise manner with a 2 kcal $mol^{-1}Å^{-2}$ positional restraint on the glycan chain in the canonical (NVT) ensemble. Consequently, under the NPT ensemble, 1 ns equilibration was carried out at 1 atm pressure using Berendsen barostat[84]. Finally, 20 μs unbiased NPT simulation was carried out with a 2 fs time step for each of the glycans. First, 100 ns data was skipped for the stabilization of both the systems. The last 19.9 μs data, i.e., in total 1990000 frames for each glycan, was used for the analysis purpose discussed in the subsequent section.

Analysis

Most of the analysis was done using the *Cpptraj*[85] module of AMBER18[79]. The atomistic change in the glycan conformation was analyzed in terms of the local glycosidic dihedral angle, which is defined as φ = O_5-C_1-O-C_n, ψ =C_1-O-C_n-$C_{(n-1)}$ and ω = O_6-C_6-C_5-C_4. Using these, we constructed 2D free energy surfaces for understanding the conformational space of each linkage. Generalized pseudo rotation coordinate formulated by Cremer and Pople[86] was used for quantifying the puckering states of the six-membered pyranose rings. We used the IUPAC convection to cluster all the sampled pucker states in our simulation. Further, we also constructed a 1-D free energy surface using amplitude (θ) puckering parameter as a reaction coordinate The conformational free energies in for different coordinates were constructed by the following Boltzmann formula i.e.

$$\Delta G(x,y) = -kTln[\frac{P_1(x,y)}{P_{max}(x,y)}] \quad (1)$$

where, x = reaction coordinate, $P_1(x,y)$ = population of specific bin of the system along reaction coordinate, $P_{max}(x,y)$ = population of the most populated bin of the system along reaction coordinate, k = Boltzmann constant and T = absolute temperature.

The correlated and anti-correlated motions between heavy atoms (without hydrogen) of each carbohydrate residue in the glycan chain, we generated the dynamics cross-correlation map using the cpptraj[85] module of the Ambertools19. We also conducted principal component analysis (PCA) by diagonalizing the covariance matrix constructed by considering only the heavy atoms of carbohydrates. These generate a set of eigenvectors and their corresponding eigenvalues representing the set of principal components (PCs). To obtain the conformation free energy spaces, we also constructed free energy landscapes (FEL) by the first two PCs.

Global conformation was characterized with the help of different sets of end to end distances and angles (discussed in the result section). The occupancy of these parameters was calculated using a kernel density estimation procedure using in-house python scripts. The following criteria for calculating hydrogen bonds between residues of each glycan chains are as follows: distance cut-

off (≤ 3.5 Å) & angle cut-off (≥ 120˚) between the donor and accepter atom. For 3D structure visualization and the data visualization was done using chimeraX[87] and in house python scripts.

Result and Discussion

The results section is divided into three-part, local conformational changes in terms of glycosidic angles, sampling of the rigid ring structure of each six-member rings, and global structural changes. The schematic image of both the glycans is shown in **Figure 1 (A, B),** along with the linkage information. *NGlycan1* and *NGlcyan2* consist of 10 and 13 monosaccharides, respectively. Residue numbers, as well as its names, are also mentioned in **Figure 1**, which will be followed throughout the paper.

Figure 1C displays the time evolution of root mean square deviation (RMSD) of the ring atoms from the well-equilibrated structure for both glycans. The distribution of RMSD for both cases are shown in Supporting Information Figure S1. It is evident from Figure S1 that the distribution is much wider for Nglycan1 compared to Nglycan2. However, both distributions are characterized by a single peak. Thus multiple complex branches restrict conformational variability through long scale dynamics. In the case of Nglycan1, the peak is located at 5.2 Å. The distribution of RMSD for Nglycan2 is characterized by a sharp peak located at 6 Å.

We have also investigated the flexibility of each saccharide by measuring the root mean square fluctuation (RMSF) and shown in **Figure S2 (**Supporting information**)**. The terminal mannose residues and the sialic acids show almost double fluctuation compared to the other residues. Also, the 4th and 5th mannose residue, i.e., branching point of the complex branch and mannose respectively, shows the opposite pattern of RMSF value in *NGlycan1* and NGlycan2 as these points serve the difference between them.

Table 1: Conformational propensities of all the glycosidic linkages (φ, ψ, ω) of NGlycan1 and NGlycan2.

Linkage	NGlycan 1			NGlycan 2		
	φ	*ψ*	*ω*	*φ*	*ψ*	*ω*
GlcNAc(2)-β(1-4)-GlcNAc(1)	-80	105		-80	105	
Man(3)-β(1-4)-GlcNAc(2)	-75	110		-75	115	
Man(4)-α(1-3)-Man(3)	70	-95		70	-125	
Man(5)-α(1-6)-Man(3)	65	-175	55 (gg)	70	-175	-160 (gt)
GlcNAc(6)-β(1-2)-Man(4)	-80	-80		-105	-75	

Gal(7)-β(1-4)-GlcNAc(6)	-75	115		-90	100	
Neu5Ac(8)-α(2-6)-Gal(7)	60	150	-55 (tg)	65	165	-170 (gt)
Man(9)-α(1-6)-Man(5)	65	-170	55 (gg)	65	-170	-70 (tg)
Man(10)-α(1-3)-Man(5)	70	-100		65	-100	
GlcNAc(11)-β(1-4)-Man(4)				-80	105	
Gal(12)-β(1-4)-GlcNAc(11)				-70	120	
Neu5Ac(13)-α(2-6)-Gal(12)				60	150	-60 (tg)

The conformational propensity of different branch and linkages:

Figure 2 and **Figure 3** show the free energy maps corresponding to the glycosidic linkages (α, β) of *NGlycan1* and *NGlycan2*, respectively. The principle minimum of each glycosidic linkage is stated in **Table 1** for better understanding. The common body of these two glycans consists of trisaccharide having GlcNAc (2)-β (1-4)-GlcNAc (1) and Man (3)-β (1-4)-GlcNAc (2) linkages. Both the linkages are conformationally rigid with a single large minimum around φ,ψ = (-80,105) and φ,ψ=(-75,110) respectively, as shown in **Figure 2-3 (A, B)**. This single minimum conformation is guided by two strong hydrogen bond forms among these three residues. The very first one is formed between O5 atom of GlcNAc (2) and O3 atom of GlcNAc (1) having more than 55% time in both the trajectory. The hydrogen bond between Man (3) O5 and O3 atom of GlcNAc (2) also has a strong presence of 35% of its trajectory. Overall, this initial standard sequence maintains its overall rigidity in both sequences throughout our simulation. Apart from these global minima, a prominent secondary minimum in the glycosidic linkage between residue GlcNAc(2) and Man(3) is observed in both the cases having a free energy less than 2 kcal/mol around φ,ψ=(-60,-60) (see **Figure 2-3, B**) which can be justified by the lesser strength of the hydrogen bond compare to the bond between GlcNAc(1) and GlcNAc(2). Details of the hydrogen bonds are mentioned in **Table 2**.

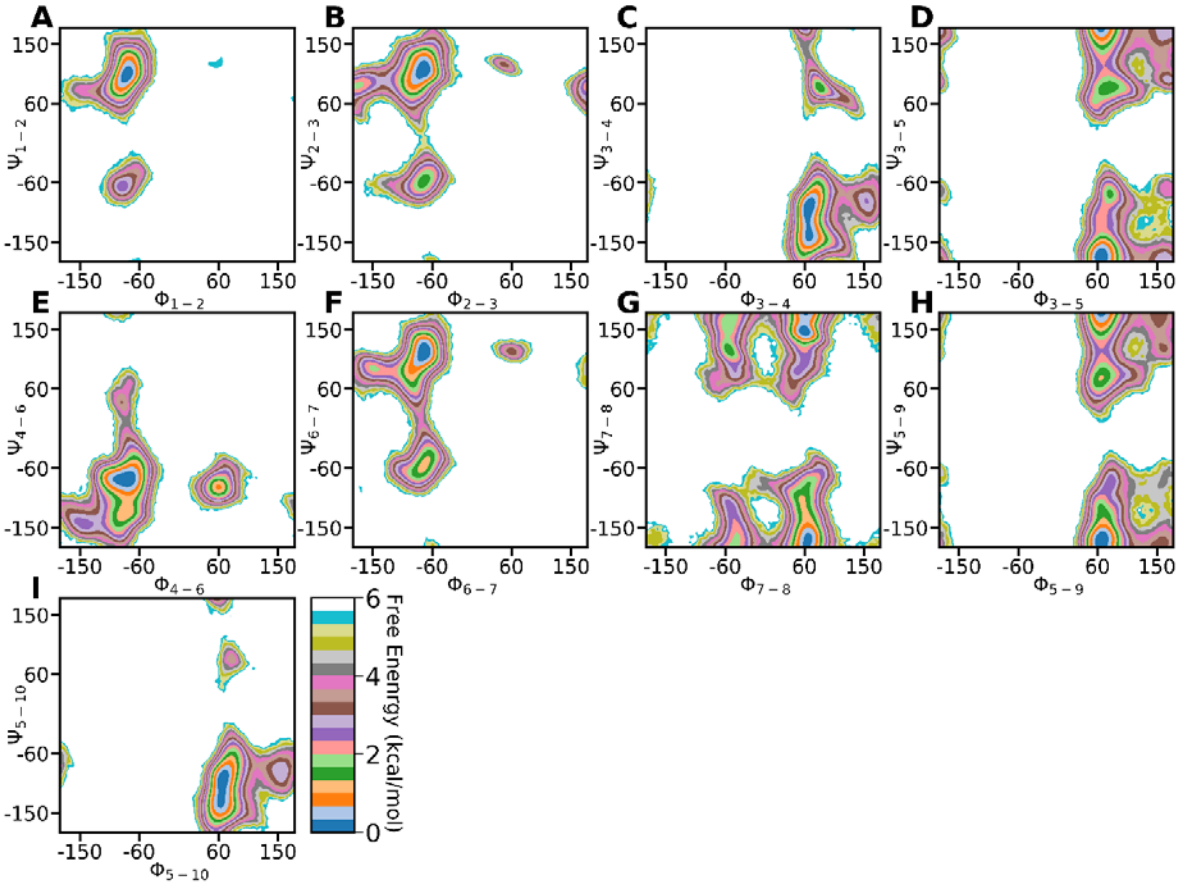

Figure 2: Free energy surfaces (kcal/mol) of the φ/ψ glycosidic angle of NGlycan1 obtained from the MD simulations. A) GlcNAc(2)-β(1-4)-GlcNAc(1), B) Man(3)-β(1-4)-GlcNAc(2), C) Man(4)-α(1-3)-Man(3), D) Man(5)-α(1-6)-Man(3), E) GlcNAc(6)-β(1-2)-Man(4), F) Gal(7)-β(1-4)-GlcNAc(6), G) Neu5Ac(8)-α(2-6)-Gal(7), H) Man(9)-α(1-6)-Man(5), I) Man(10)-α(1-3)-Man(5).

Now we will discuss the flexibility of the mannose (1-6) branch of *NGlycan1* and *NGlycan2* sequences. In both sequences, it comprises of three mannose residues (Man (9)-α (1-6)-Man (5) and Man (10)-α (1-3)-Man (5)) which further connect to Man3 via 1-6 linkage (Man (5)-α (1-6)-Man (3)). Conformational preferences of this branch in terms of glycosidic linkages in both the sequences are comparatively broader in comparison to the previous linkages. Principal minimum of mannose 1-6 branch is observed around (65, -175) (*NGlycan1*) (**Figure 2D**) and φ,ψ=(70, -175) (*NGlycan2*) (**Figure 3D**), whereas other more than one secondary minimum within 2 kcal/mol is observed along with the same φ value bridges within a large basin. Another 1-6 linkage forms between 9th and 5th also show the minimum around the same φ value with a difference in the ψ value. The sampling classification in terms of omega linkage is described along with the other 2-

6 linkage of sialic acid in the next section. There is no such specific hydrogen bond observed in all mannose residues of this specific branch, which explains multiple numbers of minima around both 1-6 linkage. On the other hand, 1-3 linkage of this mannose branch shows a stable and sizeable single minimum around approximately around φ,ψ=(70,-100) (**Figure 2-3 I**).

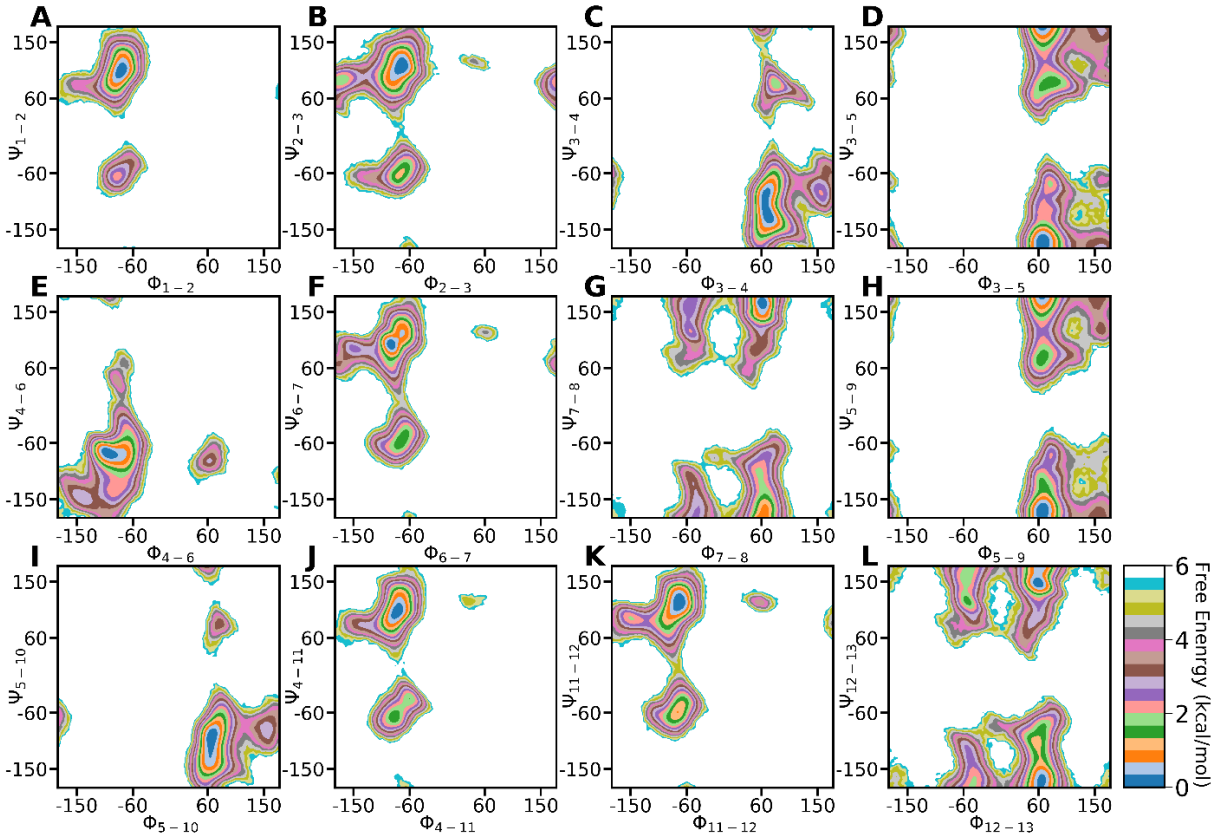

Figure 3: Free energy surfaces (kcal/mol) of the φ/ψ glycosidic angle of NGlycan2 obtained from the MD simulations. A) GlcNAc(2)-β(1-4)-GlcNAc(1), B) Man(3)-β(1-4)-GlcNAc(2), C) Man(4)-α(1-3)-Man(3), D) Man(5)-α(1-6)-Man(3), E) GlcNAc(6)-β(1-2)-Man(4), F) Gal(7)-β(1-4)-GlcNAc(6), G) Neu5Ac(8)-α(2-6)-Gal(7), H) Man(9)-α(1-6)-Man(5), I) Man(10)-α(1-3)-Man(5) J) GlcNAc(11)-β(1-4)-Man(4), F) Gal(12)-β(1-4)-GlcNAc(11), G) Neu5Ac(13)-α(2-6)-Gal(12).

On the other hand, the 1-3 linkage between Man (4)-α (1-3)-Man (3), which is the starting point of complex branch, shows a similar trend in the free energy surface. The difference between *NGlycan1* and *NGlycan2* arises because of the complex branch. *NGlycan1* has one complex branch starting with 1-2 linkage between GlcNAc and Man, whereas *NGlyan2* has two branches having 1-2 as well as 1-4 linkages. The GlcNAc (6)-β (1-2)-Man (4) linkage, which is familiar to both the sequences show a broader minima region in *NGlycan1* sequence along with the other secondary

minima region. The component of the complex branch is the same in both the sequences, which is GlcNAc, Gal, and Neu5Ac, which consist of Gal (7)-β (1-4)-GlcNAc (6)/ and Neu5Ac (8)-α (2-6)-Gal (7). In case of Gal (7)-β (1-4)-GlcNAc (6), the principal minimum is shifted form φ,ψ = (-75,115) (**Figure 3F**) to φ,ψ =(-90,100) (**Figure 3F**) and also area of the principal minimum get decreases in *NGlycan2*. This data suggest that the addition of a secondary arm reduces the flexibility of the existing arm. Also, in *NGlycan2*, the free energy value of the secondary minimum gets increase above 1 kcal/mol in comparison to *NGlycan1*, which is around (-55, -55). Comparatively, the second complex branch of *NGylcan2* shows higher flexibility. The linkage between 11th and 12th residue samples around φ,ψ = (-70,120) (**Figure 3K**) as a principal minimum along with the secondary minimum under 1kcal/mol, which indicates the coexistence of multiple conformations in aqueous solution. Finally, all the terminal linkages involved sialic acids show a higher degree of flexibility irrespective of several branches.

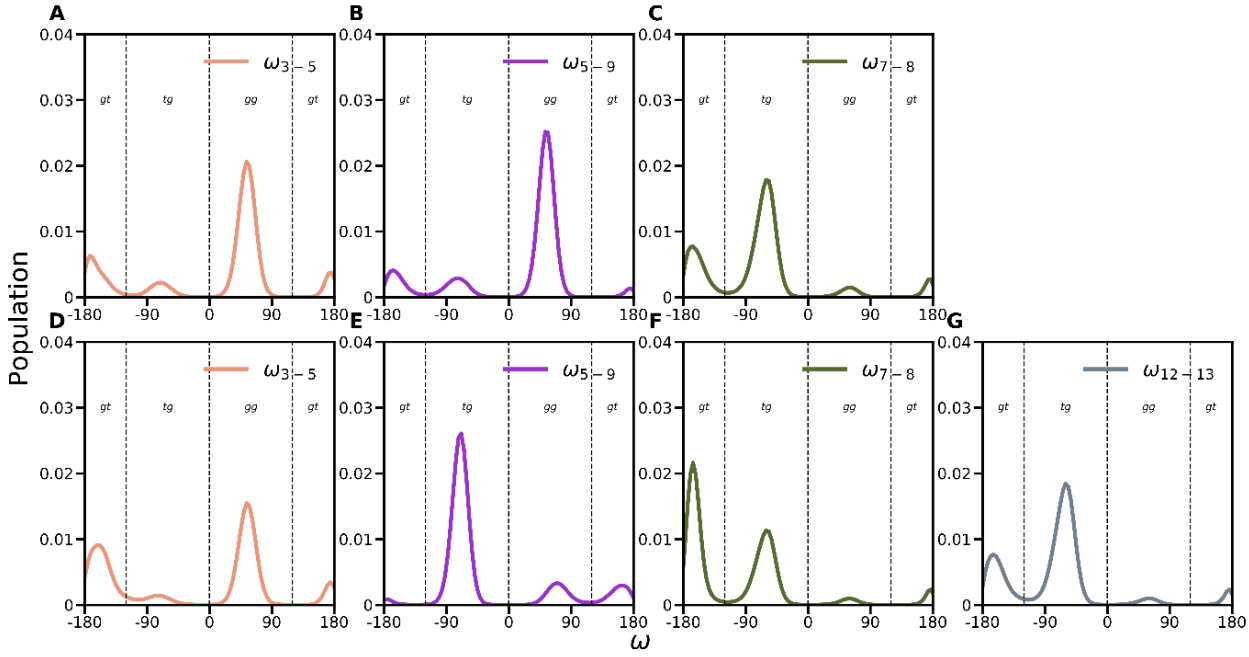

Figure 4: Distribution of ω torsion angle of each 1-6 linkage of NGlycan1 and NGlycan2. Linkages of NGlycan1 are A) Man (5)-α (1-6)-Man (3), B) Man (9)-α (1-6)-Man (5), C) Neu5Ac (8)-α (2-6)-Gal (7). Linkages of NGlycan2 are D) Man(5)-α(1-6)-Man(3), E) Man(9)-α(1-6)-Man(5), F) Neu5Ac(8)-α(2-6)-Gal(7), G) Neu5Ac(8)-α(2-6)-Gal(7).

Figure S2 shows the free energy maps for the orientation of the glycosidic linkage of all 1-6 linkages present in our simulation in terms of ω and ψ. All 1-6 linkage samples over three

conformational clusters, i.e., gauche-gauche, gauche-trans, and trans-gauche, which is shown in **Figure 4**. Mannose branch (Man (5)-α (1-6)-Man (3)) samples over the gauche-gauche cluster in both the sequences, whereas gauche-trans is also found in the case of *NGlycan2* in a significant amount. Interestingly another 1-6 linkage in the mannose branch (Man (9)-α (1-6)-Man (5)) samples different sampling between two sequences. In *NGlycan1*, it stays around in the gauche-gauche cluster, whereas it moves towards trans-gauche. In *NGlycan2*, Neu5Ac (8)-α (2-6)-Gal (7) linkage sample only a very narrow region with high probability which belongs to the gauche-trans region, which because of the result of several strong hydrogen bonds between Man (3) as well as with Man (10). Though another 2-6 linkage of NGlycan2 and the only 2-6 linkage of *NGlycan1* samples the same conformational space as they jumped between trans-gauche and gauche-trans.

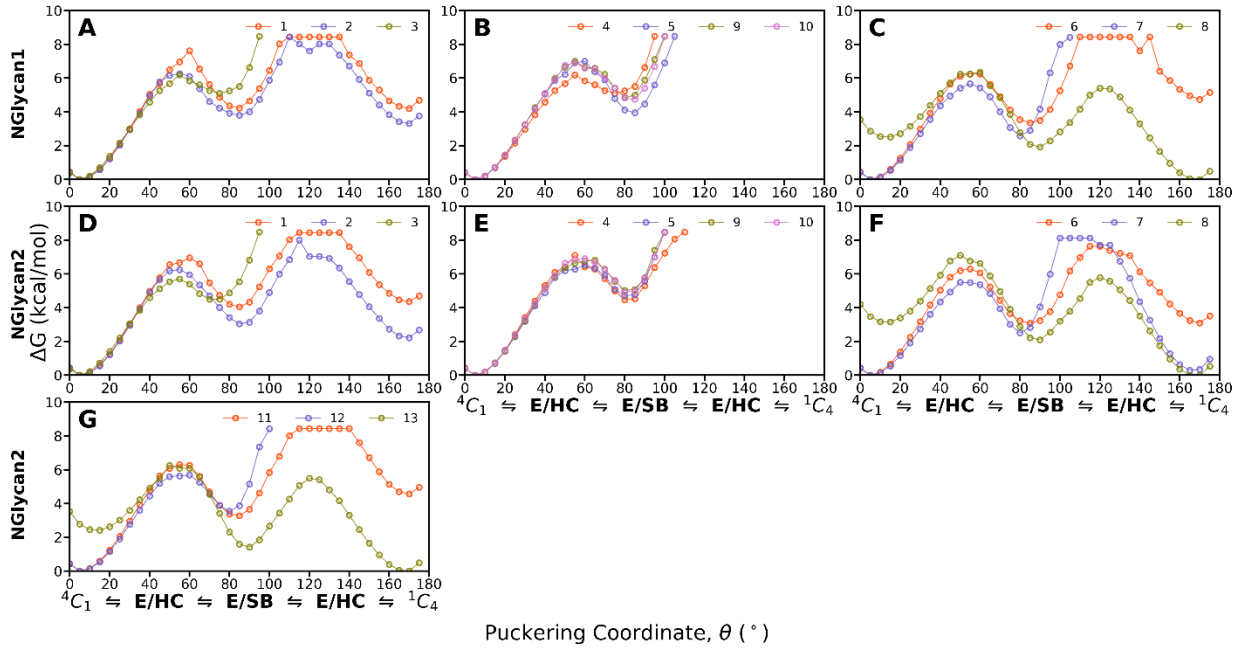

Figure 5: 1D θ puckering profile for each monosaccharides present in NGlycan1 and NGlycan2. A), B), C) corresponds to NGlycan1 whereas D), E), F) and G) corresponds to NGlycan2. Type of monosaccharides corresponding to the residue number are also shown in the image. Same color code is used throughout the paper (See Figure 1 legend).

Comparison of puckering free energy landscapes between NGlycan1 and NGlycan2:

We compute the puckering free energy profile of each monosaccharide present in both sequences, which comprise of N-Acetyl Glucosamine (GlcNAc), Mannose (Man), Galactose (Gal) and sialic acid (Neu5Ac). All these profiles are computed using the Cremer-Pople angle θ, which describe

the change of conformation of the sugar ring from the $^4C_1$ (θ = 0˚) chair form to $^1C_4$ (θ = 180˚) chair form via intermediate boat/skew-boat (θ = 90˚) structure. The free energy profile of all the monosaccharides presented in **Figure 5**. Alongside with the free energy profile, all sampled canonical puckering conformations and the Mercator representations of the Cremer-Pople sphere are also shown in **Figure S4-S5**. The previous study suggests that the excessive of 5 to 10 μs is required to attain the converge free energy profile of the monosaccharides. For that reason, we simulate our structure up to 20 μs to obtain the proper sample of the different puckering states for each sugar molecules[88]. The time evolution of the theta coordinate of the puckering of all monosaccharides is plotted and shown in **FigureS6-S7**. Surprisingly, all mannose present in our study, in the simulation time scale prefer to stay around $^4C_1$ (θ =10˚) along with less probable high energy minima (more than 4 kcal/mol) around θ = 90˚. Figure 5 indicates that the $^1C_4$ state is much higher around 3-4 kcal/mol than the $^1C_4$ ring (principle minima) in the case of GlcNAc. All the GlcNAc residues are also sampled with the skewed boat structure, which is having free energy around 4 kcal/mol. Pucker sampling space for mannose and the GlcNAc residues are indistinguishable in both sequences. Galactose, specifically Gal (7) in NGlycan2, a preferred state in near to $^1C_4$ (θ = 165˚) along with the $^1C_4$ chair in room temperature. Here, $^4C_1$ chair is more than 0.4 kcal/mol from $^1C_4$, which is not observed in the other galactose residues in both the sequences. The time evolution of the θ coordinate of Gal (7) shows the initial conformation of this ring is $^4C_1$ as similar to other Gal (7) residues, which shifted to $^1C_4$ chair after 500 ns of the simulation **(Figure S7)**. Further, it came back to its initial ring structure around 8.3 μs. Thus, this galactose shows multi microsecond rate of a shift in the ring puckering.

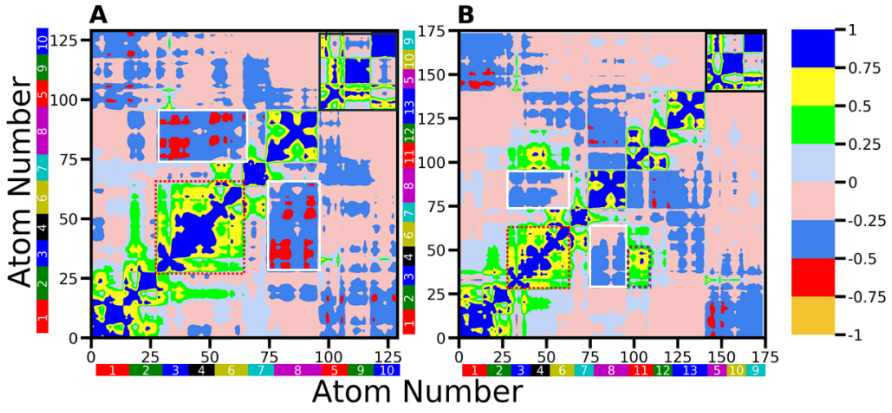

Figure 6: Dyanmic cross-correlation map for A) NGlycan1 and B) NGlycan2. Degree of correlation and anit—correlations are color-coded, as shown in the adjacent color bar. Blue shows

highest degree of correaltion where dark yellow indicates the highest degree of anti-correlation. The color-coded area regions in the X and Y axis show the spread of each specific residue.

Global conformational change in NGlycan1 and NGlycan2:

To estimate the overall dynamics of the N-glycan chain, we calculate the dynamic cross-correlation map throughout the simulation trajectory using all heavy atoms in the system. Higher correlated and anti-correlated regions are shown in the blue and dark yellow colors, respectively. Overall observations indicate the anti-correlation decreases in the more extended sequence, i.e., *NGlycan2*. In *NGlycan1*, terminal sialic acid, Neu5Ac (8) shows a high degree of anti-correlation (white rectangle, **Figure 6A**) with 3,4 and 6^{th} residues, which are the connecting residues between the main branch and complex branch. This anti-correlation destroys because of the additional complex branch (residue 11-13) in *NGlycan2* **(**white rectangle, **Figure 6B)**. We also observed several residues in both cases have a high degree of correlation between themselves only. Those residues are determining several micro-level conjugated structural movements, which are also revealed by the hydrogen bond occupancy and other calculations discussed in the later section. Residues 3^{rd}, 4^{th}, and 6^{th}, i.e., the branching mannose of the complex branch and the initial GlcNAc of the same branch, have a correlated motion (red dotted box) having a Pearson coefficient more 0.5. The other GlcNAc residue (11^{th}) of the second complex branch of NGlycan2 also shows the same level correlation with the group mentioned above (blue dotted box, **Figure 6B**). Surprisingly the residues in the mannose branch, i.e., 5^{th}, $9^{th,}$ and 10^{th} mannose, do not show any type correlation between the branching residues whereas having a high correlation between themselves only in both the sequences (black square, **Figure 6 (A, B)**). Also, the motion of these mannose residues has a slightly anti-correlation relationship with the terminal GlcNAc residues, which is not observed in the case of the terminal residues of the complex branches. From this dynamic cross-correlation, we are not able to see any significant relationship between mannose branch and the complex in both the cases.

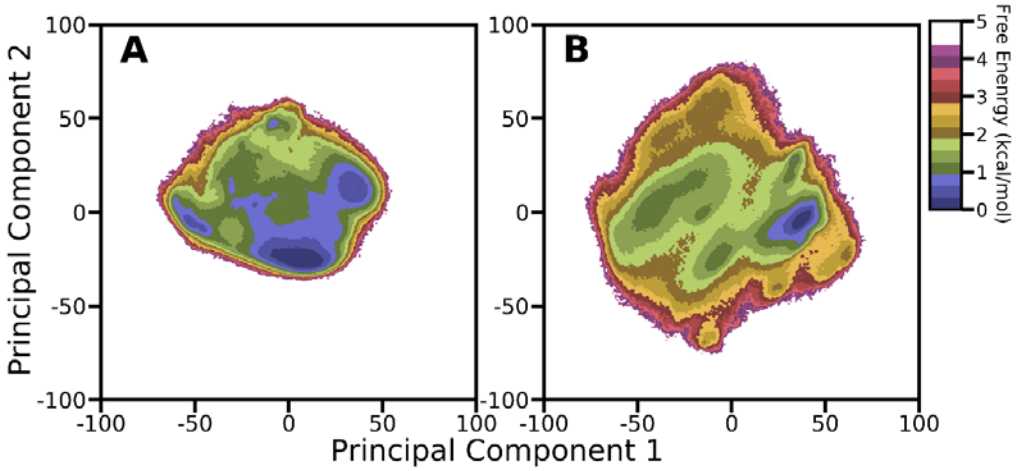

Figure 7: Two-dimensional free energy landscape generated using first and second principal components of the system, A) NGlycan1 and B) NGlycan2.

We also calculate the principal components (PC) on the heavy atoms of N-glycan to extract the structural variation between both the N-glycan sequences by diagonalization of the covariance matrix. Each eigenvector and corresponding eigenvalues were plotted in decreasing order shown in **Figure S8**. The first 10 PCs are responsible for 86% and 82% of the global motion of *NGlycan1* and *NGlycan2*, which is evident because the first few eigenvalues describe the collective motion of local fluctuation. Similarly, in our study, the first two eigenvectors contain 42% and 39% of the overall motion of *NGlycan1* and *NGlycan2*. Free energy landscapes were constructed using the Boltzmann formula using PC1 and PC2 to elucidate the conformational basin of these glycans (see **Figure 7**). *NGlycan2* samples larger conformational space in comparison to *NGlycan1*, whereas the number of minima is more in *NGlycan1*. *NGlycan1* has a principal minimum around (5, -25) along with a secondary minimum (35,15), but both are connected by a very shallow free energy barrier less than 1kcal/mol. There are also a few minima present which separated by a higher free energy barrier indicating the presence of multiple numbers of global conformations at the same temperature. On the contrary, the other sequence has a minimum of around 35, -5, which is surrounded by comparatively higher energetic structures. Overall, it is observed that the mobility of the NGlycan1 much higher because of its smaller size as well as less branching point in the sequence.

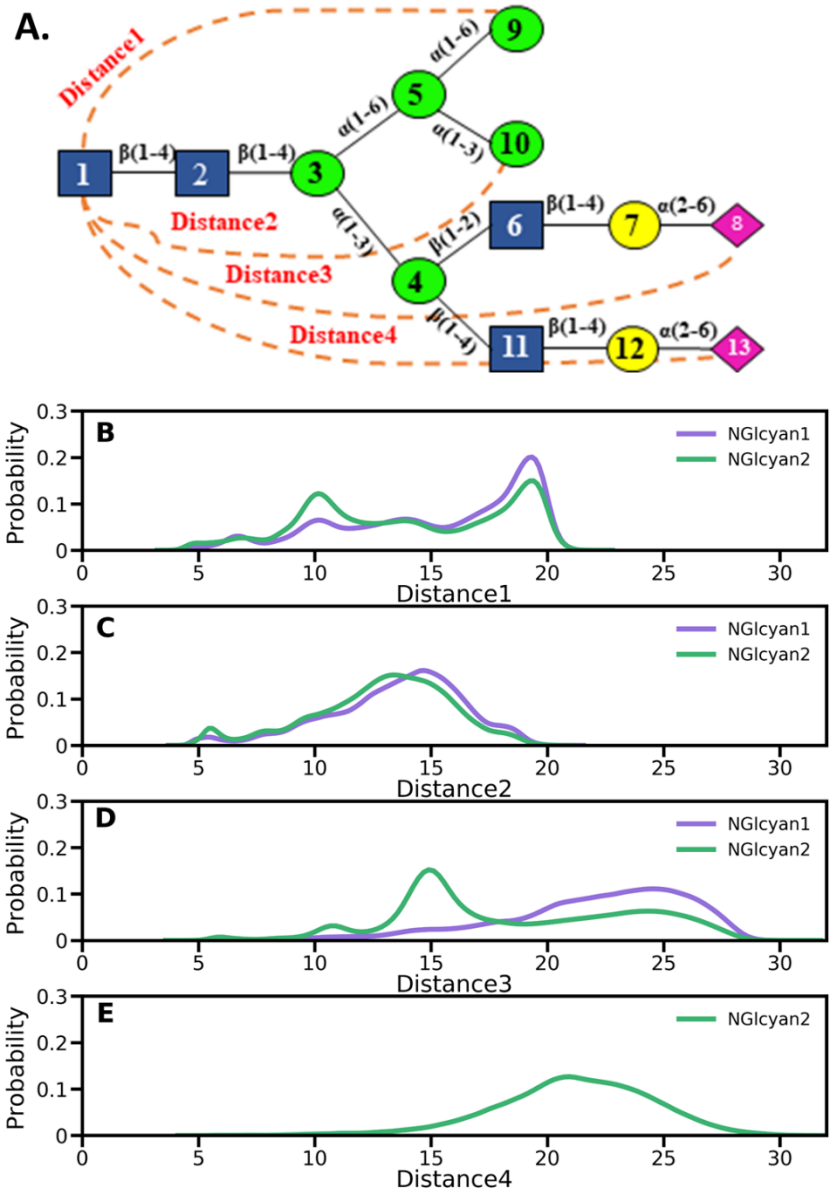

Figure 8: Distribution of end to end distance corresponding to different branches of NGlycan1 and NGlycan2. A) distance between GlcNAc(1) and Man(9) (Mannose branch), B) distance between GlcNAc(1) and Man(10) (Mannose branch, C) distance between GlcNAc(1) and Neu5Ac(8) (Complex branch) and D) distance between GlcNAc(1) and Neu5Ac(13) (Complex branch). NGlycan1 and NGlycan2 data are shown in purple and green colors, respectively.

Distribution of different end to end distances and branching angles of glycan chain:

To examine the movement of different branches and three-dimensional structural features of the N-glycan chain, a different set of ends to end distances (between 1st residue and the other terminal

residues), as well as a set of angles, were calculated. Overall, four different sets of distances were calculated, where the first three are common in both NGlycan1 and NGlycan2. These distances are calculated between the residues as describes, Distance1: GlcNAc (1) and Man (9); Distance2: GlcNAc (1) and Man (10); Distance3: GlcNAc (1) and Neu5Ac (8) and Distance4: GlcNAc (1) and Neu5Ac (13). Distribution of all the distances using kernel density function and the schematic diagram of inter-residue distances are shown in **Figure 8**. Apart from the distances, branching angles were also calculated. Three branching angles are calculated, Angle1: Man(9)-Man(5)-Man(10) (between mannose branches); Angle2: Man(5)-Man(3)-Man(4) (between mannose and complex branch) and Angle3: GlcNAc(11)-Man(4)-GlcNAc(6) (between complex branches). Along with this, we also determine the curvature of the two complex branches by calculating the angles, i.e., Angle4: GlcNAc(6)-Gal(7)-Neu5Ac(8) and Angle5: GlcNAc(11)-Gal(12)-Neu5Ac(13). The distribution of the angles mentioned above is shown in **Figure 9**, alongside with schematic representation. The behavior of the mannose branch mainly depends upon the mannose(9), which is linked by the 1-6 linkage. It fluctuates between two states as obtained from the distance calculation (see **Figure 8A**) around 10.1 Å and 19.2 Å in both the N-glycan having a different level of magnitude. On the other hand, other mannose(10) linked by the 1-3 linkage shows a broader distribution around 14 Å -15 Å irrespective of the glycan chain (see **Figure 8B**). The branching angle of this mannose branch (angle1) also shows a similar bi-modal distribution in both the N-glycan. So the mannose branch fluctuates in two states irrespective of the number of complex branches, whereas the magnitude of the peaks is different in both the glycan chain. In NGlycan1, state around 110˚ has double occupancy over the state around 160˚, whereas both the state has similar occupancy in the other sequences.

The primary branching angle, i.e., angle2, shows a broad peak around 110˚, indicating a steady structure of Penta-saccharide core in both the cases. There is no significant difference after the addition of one new complex branch in the case of NGlycan2 also. Finally, we explore the conformational landscape of the complex branch in terms of distance and angle calculation. The difference between the used sequence of our work lies in the complex branch only, so it is the key determinant factor of structural classification between these two sequences. Therefore, a single complex chain (β1-2) arm, i.e., in *NGlycan1*, shows a broader distribution in the end to end

distance (distance4) ranging from 15 Å to 30 Å indicating a higher level of flexibility (see **Figure 8D**).

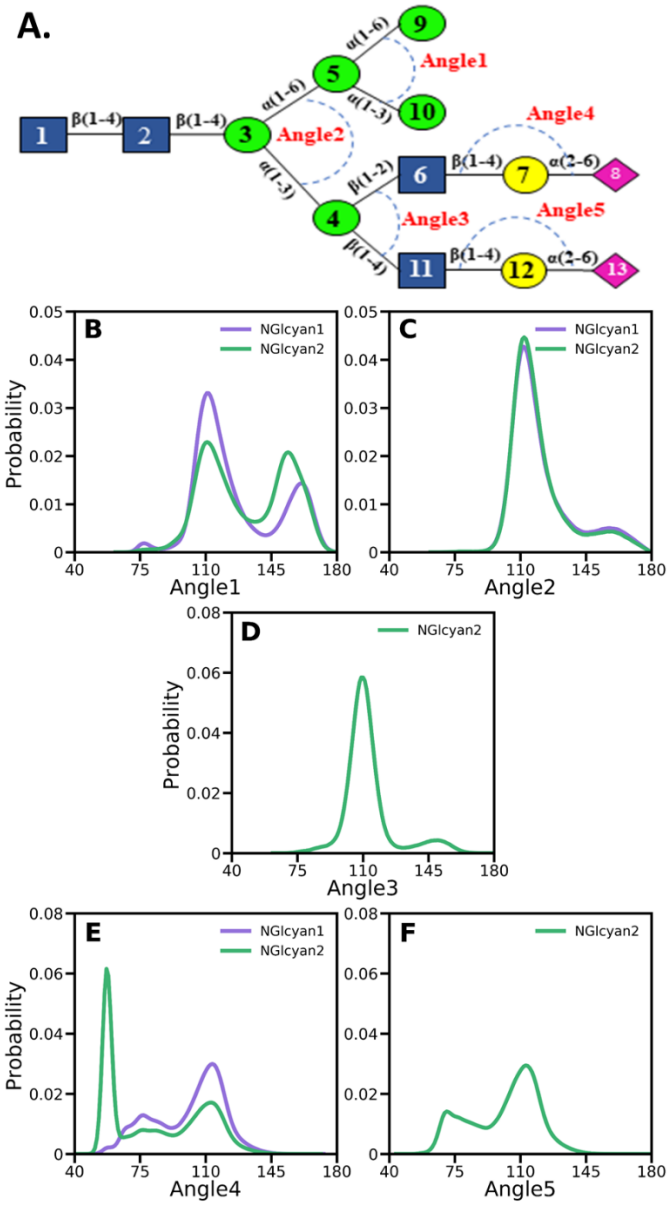

Figure 9: Distribution of branching angles and the angle form by each complex branch. Distribution of branching angles described as follows: A) Man(9)-Man(5)-Man(10), B) Man(5)-Man(3)-Man(4), C) GlcNAc(11)-Man(4)-GlcNAc(6) and angle form by complex branch are as follows: D) GlcNAc(6)-Gal(7)-Neu5Ac(8), E) GlcNAc(11)-Gal(12)-Neu5Ac(13).

On the other hand, the addition of β1-4 arm (*NGlycan2*) restricts the motion (around 15 Å) of an existing arm (β1-2) makes a compact structure rather than a linear structure. Also, the second arm,

i.e., β1-4, stays in the same distribution as previously observed in the β1-2 arm of NGlycan1 (see **Figure 8D**). Further, the distribution of angle4 and angle5 (see **Figure9 (D, E)**), the angle between the tri-residue of each complex branch, shows the same asymmetrical distribution of two complex branches of *NGlycan2*. The angular distribution of β1-2 arm of *NGlycan1* and β1-4 arm of *NGlycan2* has a peak around 115Å along with a little hump around 70Å which specifies the flexible nature of mentioned arm. On the other hand, the angle between GlcNAc(6)-Gal(7)-Neu5Ac(8) of *NGlycan2* shows a sharp peak of around 57.5 Å. So this compact structure of this segment achieved not by the backward movement of the whole arm, only because of the curved conformation of the complex arm. The nature of the conformation in these complex branches changes upon the addition of the other arm, which influences the structural modification. As we extend our simulation over 20 μs, so it concludes that this nature is not a transient and interplay between different hydrogen bond networks stabilize throughout the time scale.

Table 2: Occupancy of hydrogen bonds between residues in each chain during MD simulation.

Acceptor	Donor	Distance (Å)	Angle (º)	Occupancy (%)
NGlycan1				
GlcNAc2@O5	GlcNAc1@O3	2.76	157.41	56.23
Man3@O5	GlcNAc2@O3	2.79	155.90	33.75
Gal7@O5	GlcNAc6@O3	2.78	154.85	25.35
Neu5Ac8@O5N	Neu5Ac8@O7	2.79	157.33	14.78
Neu5Ac8@O1B	Gal7@O4	2.70	164.89	10.21
NGlycan2				
GlcNAc2@O5	GlcNAc1@O3	2.76	157.31	55.36
GlcNAc11@O5	Man4@O3	2.75	156.90	45.74
Man3@O5	GlcNAc2@O3	2.79	155.99	33.90
Man3@O4	Neu5Ac8@O7	2.77	161.62	27.36
Gal7@O5	GlcNAc6@O3	2.79	153.25	23.63
Gal12@O5	GlcNAc11@O3	2.78	155.02	22.23
Gal7@O6	GlcNAc6@O3	2.82	157.36	22.12
Neu5Ac8@O5N	Neu5Ac8@O7	2.78	156.79	15.90
Neu5Ac13@O5N	Neu5Ac13@O7	2.79	157.41	15.36
Neu5Ac8@O1A	Man10@O6	2.69	163.92	10.23

Inter residue and water-mediated hydrogen bond networking:

Conformation of oligosaccharides relies on the hydrogen bonds, which are already established by different previous studies[39, 51, 89]. We performed the hydrogen bond (H-bond) analysis in both the N-glycan chain and the listed the all critical inter-residue hydrogen bond in **Table 2**. As the water molecules play a critical role in the carbohydrate's structures, we also investigate the propensity of the water-mediated hydrogen bond between residues, and the essential bonds are listed in **Table 3**. For a better understanding, we combine both the hydrogen bonds in a schematic diagram and shown in **Figure 10**. Solid and dashed lines are used to show the inter-residue and water-mediate hydrogen bond. As already discussed, the tri-saccharides stay stable with a higher percentage of inter-residue and water mediate h-bonds (see **Table 2 & Table 3**), which remain the same in both the N-glycans. The interaction pattern of the mannose branch was mainly depending on the water-mediated h-bonds having a percentage between 25% to 10% with GlcNAc (2) and Man (3). There are few other water-mediated h-bond presents in between the mannose branch that yielded other less probable conformational states. Instead of having a one or two steady bond, several h-bonds (mainly water-mediated) creates a complex network interaction hence conserve the flexibility of it.

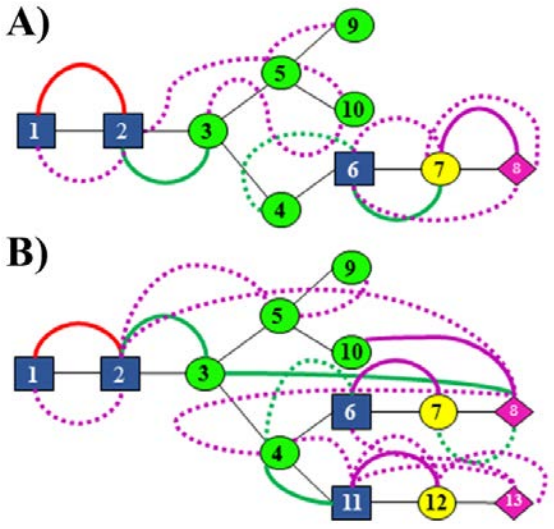

Figure 10: Schematic diagram of inter-residue H-bond pattern (including water-mediated H-bond). A) NGlycan1, B) NGlycan2. The same color scheme was used throughout the paper in all residues. Different color and line-type were used to indicate the percentage and types of H-bond. Solid line type was used for inter-residue, and a dashed line was used for water-mediated H-bond. Red: ≥ 50%, Green: 50% to 25%, Violet: 25% to 10%.

The complex branch shows a steady H-bond between the GlcNAc and galactose residue with 25.35% and 23 % (average) in *NGlycan1* and *NGlycan2*, respectively. Along with Gal (6) and Neu5Ac (8) of NGlycan1 also has a direct h-bond between their oxygen atoms (10.21%) indicate its linear conformation. This branch also has water-mediated h-bond to its neighboring residues (see **Figure 10A**). A similar type of H-bond interaction was also observed in the β1-4 branch of *NGlycan2*. On the contrast, there was a strong hydrogen bond between O5 atom to GlcNAc (11) and O3 atom of Man (4). Now the hydrogen bond profile of the β1-2 branch of NGlycan2 was quite different in comparison to the β1-4 arm. Terminal sialic acid of this branch (β1-2) has two strong hydrogen bonds between Man (3) and Man (10) with a percentage of 27.36 % and 10.23 %, respectively. This data is in agreement with our distance and angle analysis of the same branch. There is also a water-mediated hydrogen bond (21.09%) in between Man (4), and Neu5Ac (8) helps to stay its compact nature of the branch. We also observed a water-mediated bonding network involving GlcNAc (2), Man (3), and Neu5Ac (8) around 6.70 % throughout our timescale, explaining the same shrinking nature of the branch.

Table 3: Occupancy of water mediate hydrogen bonds between residues in each chain during MD simulation.

Residue1	Residue2	Occupancy (%)
NGlycan1		
Man4	GlcNAc6	29.52
GlcNAc6	Neu5Ac8	21.52
Gal7	Neu5Ac8	19.19
GlcNAc6	Gal7	15.69
GlcNAc1	GlcNAc2	14.44
Man3	Man10	13.63
Man5	Man9	12.80
GlcNAc2	Man10	10.33
NGlycan2		
Gal7	Neu5Ac8	48.21
Man4	GlcNAc6	26.22
GlcNAc11	Neu5Ac13	23.99
Man4	Neu5Ac8	21.09
Gal12	Neu5Ac13	20.51
GlcNAc11	Gal12	18.54
Man4	GlcNAc11	16.57
GlcNAc2	Man5	15.43
GlcNAc1	GlcNAc2	15.14

Man5	Man9	13.36
GlcNAc2	Neu5Ac8	10.79
GlcNAc6	Neu5Ac13	10.21

Radial distribution function g(r):

Alongside the water-mediated hydrogen bonds, we also calculated pairwise correlation function or the radial distribution function g(r) of water molecules with respect to water molecules of each monosaccharide for both the sequences and shown in **Figure 11(A, B)**.

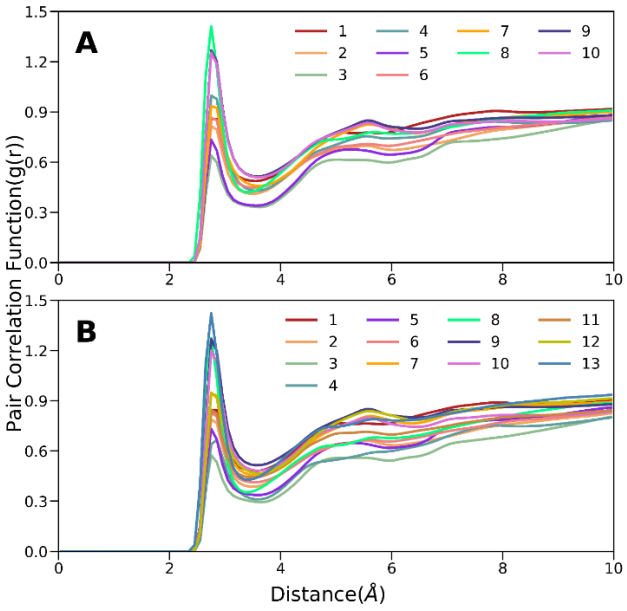

Figure 11: Pairwise correlation function of the water molecules present in the simulation as a function of distance from the oxygen atoms of each monosaccharide residue. Each of the monosaccharides is shown in a different color. A) NGlycan1 and B) NGlycan2. The same number scheme is followed (see **Figure 1**).

To reduce the complexity, we only calculated the overall g(r) of all oxygen atoms of a given monosaccharide rather than for all individual hydroxyl groups. The overall pattern of the distribution is similar for both the *N*-glycans, and the water molecules in the first coordination shell were structured indicate by the sharp first peak. This nature was already observed in other carbohydrate studies[65, 90]. Terminal residues i.e. mannose and sialic acid of both the branches, show the higher intensity of the peak because of its greater exposure to water molecules, whereas

the initial GlcNAc (1) has a lower intensity of its first peak. This is may be due to the presence of the strong hydrogen bond between GlcNAc (1) and GlcNAc (2). In both the N-glycan, Man(3), and Man(5) show the least intensity of the first peak because being a branching residue, their hydroxyl groups are less exposed to water molecules. This observation was also seen in the case of Man(4), after the addition of one more complex branch in *NGlycan2*, i.e., the value of g(r) changes from 1 to 0.6. Thus, our study reveals the potential role of water molecules in the global carbohydrate conformations.

Conclusions

Here, we have performed standard MD simulation of two *N*-glycans in solution for 20 μs long time scale to examine the preferred dynamics and change in the conformational spaces by the addition of a complex branch. These total 40 μs trajectory (2 sequences) was used to analyze the different level of motion starting from the flexibility of glycosidic bonds, sampling of all possible puckering states to the conformational shift in macro-level i.e. movement of different branches relative to each other. Our simulation length is able to sample all possible conformational space corresponding to linkages as well as to show the steady transition of the puckering form of monosaccharides. As the puckering states are separated by a high energy barrier and achieving that in cMD ensure the convergence of our simulation procedures. We also observed that the flexibility of the glycan chain reduces upon increasing the branching. A more in-depth analysis of the flexible nature will help us to understand the interaction between glycans and bNAbs. Experimental techniques already showed the different levels of avidity of PG9, PG16, PGT128, like antibodies with the number and branch of mannose and sialylated arm[67]. So, the estimation angle and distance reveal the change in movement complex branch upon increasing the number of branches. Finally, we estimate the hydrogen bonding network extensively, which has a pivotal role in achieving different conformational space. Our simulation provides the structural details of these two sequences, which will further help to characterize protein-carbohydrate interaction. Our study can be extended by attaching the minimum conformation of N-glycans to their respective glycoproteins and study the interaction of different HIV antibodies. This information may serve the initial requirement of potent HIV vaccine design as well as the fundamental understanding of the protein-glycan interaction and dynamical aspect of glycoproteins.

Acknowledgments:

This work was supported by the Department of Biotechnology, Govt. of India (grant number BT/RLF/Re-entry/40/2014, DBT-Ramalingaswami Re-entry Fellowship), and Department of Science and Technology, Govt. of India (grant number ECR/2017/000010). RJ thanks the Indian Institute of Technology Indore for financial assistance.

Supporting Information:

The Supporting Information is available free of charge on the XXXX website:
Probability distribution of RMSD, PMF of ψ/ω glycosidic angle, Mercator representations of puckering conformations, Time evolution of puckering coordinates.

Disclosure Statement:

No potential conflict of interest was reported by the authors.

Author Information:

*Corresponding Author: Parimal Kar | Email: parimal@iiti.ac.in | Phone: +91 731 2438700 (Ext. 550)

Reference

1. D. B. Werz, R. Ranzinger, S. Herget, A. Adibekian, C.-W. von der Lieth and P. H. J. A. c. b. Seeberger, 2007, **2**, 685-691.
2. B. J. Hardy, *Journal of Molecular Structure: THEOCHEM*, 1997, **395**, 187-200.
3. R. J. Woods, *Chem. Rev.*, 2018, **118**, 8005-8024.
4. A. Helenius and M. Aebi, *Annu. Rev. Biochem*, 2004, **73**, 1019-1049.
5. G. Z. Lederkremer, *Curr. Opin. Struct. Biol.*, 2009, **19**, 515-523.
6. Y. Guo, H. Feinberg, E. Conroy, D. A. Mitchell, R. Alvarez, O. Blixt, M. E. Taylor, W. I. Weis and K. Drickamer, *Nat. Struct. Mol. Biol.*, 2004, **11**, 591.
7. X. Shi and R. M. Elliott, *J. Virol.*, 2004, **78**, 5414-5422.
8. E. K. Culyba, J. L. Price, S. R. Hanson, A. Dhar, C.-H. Wong, M. Gruebele, E. T. Powers and J. W. Kelly, *Science*, 2011, **331**, 571-575.
9. M. M. Chen, A. I. Bartlett, P. S. Nerenberg, C. T. Friel, C. P. Hackenberger, C. M. Stultz, S. E. Radford and B. Imperiali, *Proceedings of the National Academy of Sciences*, 2010, **107**, 22528-22533.
10. S. R. Hanson, E. K. Culyba, T.-L. Hsu, C.-H. Wong, J. W. Kelly and E. T. Powers, *Proceedings of the National Academy of Sciences*, 2009, **106**, 3131-3136.
11. R. Solá, J. Rodriguez-Martinez and K. Griebenow, *Cell. Mol. Life Sci.*, 2007, **64**, 2133-2152.
12. R. Jefferis, *Trends Pharmacol. Sci.*, 2009, **30**, 356-362.
13. H.-J. Gabius, S. André, J. Jiménez-Barbero, A. Romero and D. J. T. i. b. s. Solís, 2011, **36**, 298-313.
14. H.-J. Gabius, *Die Naturwissenschaften*, 2000, **87**, 108-121.
15. M. R. Wormald, A. J. Petrescu, Y.-L. Pao, A. Glithero, T. Elliott and R. A. Dwek, *Chem. Rev.*, 2002, **102**, 371-386.
16. A. M. Slaney, V. A. Wright, P. J. Meloncelli, K. D. Harris, L. J. West, T. L. Lowary and J. M. Buriak, *ACS applied materials & interfaces*, 2011, **3**, 1601-1612.
17. D. R. Burton, P. Poignard, R. L. Stanfield and I. A. Wilson, *Science*, 2012, **337**, 183-186.
18. A. Bernardi, J. Jiménez-Barbero, A. Casnati, C. De Castro, T. Darbre, F. Fieschi, J. Finne, H. Funken, K.-E. Jaeger and M. Lahmann, *Chem. Soc. Rev.*, 2013, **42**, 4709-4727.
19. R. K. Shukla and A. Tiwari, *Critical Reviews™ in Therapeutic Drug Carrier Systems*, 2011, **28**.
20. P. Valverde, A. Ardá, N.-C. Reichardt, J. Jiménez-Barbero and A. Gimeno, *MedChemComm*, 2019, **10**, 1678-1691.
21. F. Schwarz and M. Aebi, *Curr. Opin. Struct. Biol.*, 2011, **21**, 576-582.
22. M. Aebi, R. Bernasconi, S. Clerc and M. Molinari, *Trends Biochem. Sci*, 2010, **35**, 74-82.
23. M. Takahashi, Y. Kuroki, K. Ohtsubo and N. Taniguchi, *Carbohydr. Res.*, 2009, **344**, 1387-1390.
24. Y. Zhao, Y. Sato, T. Isaji, T. Fukuda, A. Matsumoto, E. Miyoshi, J. Gu and N. Taniguchi, *The FEBS journal*, 2008, **275**, 1939-1948.
25. S. Shanker, L. Hu, S. Ramani, R. L. Atmar, M. K. Estes and B. V. Prasad, *Curr. Opin. Struct. Biol.*, 2017, **44**, 211-218.
26. R. Raman, K. Tharakaraman, V. Sasisekharan and R. Sasisekharan, *Curr. Opin. Struct. Biol.*, 2016, **40**, 153-162.
27. D. Shcherbakov, A. Bakulina, L. Karpenko and A. Ilyichev, *Acta Naturae (англоязычная версия)*, 2015, **7**.
28. I. J. Krauss, *Glycobiology*, 2016, **26**, 813-819.
29. B. G. Pierce, Z.-Y. Keck, P. Lau, C. Fauvelle, R. Gowthaman, T. F. Baumert, T. R. Fuerst, R. A. Mariuzza and S. K. Foung, *Proceedings of the National Academy of Sciences*, 2016, **113**, E6946-E6954.
30. S. Homans, R. Dwek and T. Rademacher, *Biochemistry*, 1987, **26**, 6571-6578.

31. S. Homans, R. Dwek and T. Rademacher, *Biochemistry*, 1987, **26**, 6553-6560.
32. S. Homans, R. Dwek, J. Boyd, M. Mahmoudian, W. Richards and T. Rademacher, *Biochemistry*, 1986, **25**, 6342-6350.
33. E. W. Sayers and J. H. Prestegard, *Biophys. J.*, 2000, **79**, 3313-3329.
34. A. W. Barb and J. H. Prestegard, *Nat. Chem. Biol.*, 2011, **7**, 147.
35. C. Weller, J. Lustbader, K. Seshadri, J. Brown, C. Chadwick, C. Kolthoff, S. Ramnarain, S. Pollak, R. Canfield and S. Homans, *Biochemistry*, 1996, **35**, 8815-8823.
36. A. Almond, B. O. Petersen and J. Ø. Duus, *Biochemistry*, 2004, **43**, 5853-5863.
37. T. Lütteke, *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, 2009, **65**, 156-168.
38. V. Slynko, M. Schubert, S. Numao, M. Kowarik, M. Aebi and F. H.-T. Allain, *JACS*, 2009, **131**, 1274-1281.
39. W. Nishima, N. Miyashita, Y. Yamaguchi, Y. Sugita and S. Re, *J. Phys. Chem. B*, 2012, **116**, 8504-8512.
40. M. Wehle, I. Vilotijevic, R. Lipowsky, P. H. Seeberger, D. Varon Silva and M. Santer, *JACS*, 2012, **134**, 18964-18972.
41. E. Fadda and R. J. Woods, *Drug Discovery Today*, 2010, **15**, 596-609.
42. E. L. Wu, O. Engström, S. Jo, D. Stuhlsatz, M. S. Yeom, J. B. Klauda, G. Widmalm and W. Im, *Biophys. J.*, 2013, **105**, 1444-1455.
43. K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods, *J. Comput. Chem.*, 2008, **29**, 622-655.
44. O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell Jr, *J. Chem. Theory Comput.*, 2009, **5**, 2353-2370.
45. O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, R. W. Pastor and A. D. Mackerell Jr, *J. Comput. Chem.*, 2008, **29**, 2543-2564.
46. E. Hatcher, O. Guvench and A. D. MacKerell Jr, *The Journal of Physical Chemistry B*, 2009, **113**, 12466-12476.
47. E. P. Raman, O. Guvench and A. D. MacKerell Jr, *The Journal of Physical Chemistry B*, 2010, **114**, 12981-12994.
48. H. S. Hansen and P. H. Hünenberger, *J. Comput. Chem.*, 2011, **32**, 998-1032.
49. H. Kim, Y. Choi, J. Lee, K. Jeong and S. Jung, *Bull. Korean Chem. Soc*, 2009, **30**, 2723.
50. K. N. Kirschner and R. J. Woods, *Proceedings of the National Academy of Sciences*, 2001, **98**, 10541-10545.
51. A. Almond, *Carbohydrate Research*, 2005, **340**, 907-920.
52. L. Perić-Hassler, H. S. Hansen, R. Baron and P. H. Hünenberger, *Carbohydr. Res.*, 2010, **345**, 1781-1801.
53. F. Corzana, M. S. Motawia, C. H. Du Penhoat, S. Perez, S. M. Tschampel, R. J. Woods and S. B. Engelsen, *J. Comput. Chem.*, 2004, **25**, 573-586.
54. K. Naidoo, D. Denysyk and J. Brady, *Protein Eng.*, 1997, **10**, 1249-1261.
55. R. J. Woods, A. Pathiaseril, M. R. Wormald, C. J. Edge and R. A. Dwek, *Eur. J. Biochem.*, 1998, **258**, 372-386.
56. S. Andre, T. Kožár, S. Kojima, C. Unverzagt and H.-J. Gabius, *Biol. Chem.*, 2009, **390**, 557-565.
57. S. Re, N. Miyashita, Y. Yamaguchi and Y. Sugita, *Biophys. J.*, 2011, **101**, L44-L46.
58. Y. Yamaguchi, W. Nishima, S. Re and Y. Sugita, *Rapid Commun. Mass Spectrom.*, 2012, **26**, 2877-2884.
59. S. Jo, Y. Qi and W. Im, *Glycobiology*, 2015, **26**, 19-29.
60. S. Re, W. Nishima, N. Miyashita and Y. J. B. r. Sugita, 2012, **4**, 179-187.
61. S. Re, S. Watabe, W. Nishima, E. Muneyuki, Y. Yamaguchi, A. D. MacKerell and Y. Sugita, *Sci. Rep.*, 2018, **8**, 1644.

62. M. Yang and A. D. MacKerell Jr, *J. Chem. Theory Comput.*, 2015, **11**, 788-799.
63. R. Galvelis, S. Re and Y. Sugita, *J. Chem. Theory Comput.*, 2017, **13**, 1934-1942.
64. I. Alibay, K. K. Burusco, N. J. Bruce and R. A. Bryce, *J. Phys. Chem. B*, 2018, **122**, 2462-2474.
65. R. Roy, B. Ghosh and P. Kar, *ACS omega*, 2020, **5**, 3932-3942.
66. A. M. Harbison, L. P. Brosnan, K. Fenlon and E. Fadda, *Glycobiology*, 2018, **29**, 94-103.
67. V. S. Shivatare, S. S. Shivatare, C.-C. D. Lee, C.-H. Liang, K.-S. Liao, Y.-Y. Cheng, G. Saidachary, C.-Y. Wu, N.-H. Lin and P. D. Kwong, *Journal of the American Chemical Society*, 2018, **140**, 5202-5210.
68. H. L. Robinson, *Clin. Pharmacol. Ther.*, 2018, **104**, 1062-1073.
69. A. Merk and S. Subramaniam, *Curr. Opin. Struct. Biol.*, 2013, **23**, 268-276.
70. M. A. Checkley, B. G. Luttge and E. O. Freed, *J. Mol. Biol.*, 2011, **410**, 582-608.
71. D. R. Burton and L. Hangartner, *Annu. Rev. Immunol.*, 2016, **34**, 635-659.
72. R. Kumar, H. Qureshi, S. Deshpande and J. Bhattacharya, *Therapeutic advances in vaccines and immunotherapy*, 2018, **6**, 61-68.
73. Y. Qi, S. Jo and W. Im, *Glycobiology*, 2015, **26**, 251-260.
74. J. Tian, C. A. López, C. A. Derdeyn, M. S. Jones, A. Pinter, B. Korber and S. Gnanakaran, *PLoS Comput. Biol.*, 2016, **12**, e1005094.
75. R. Woods, *Complex Carbohydrate Research Center. Athens, GA: University of Georgia*, 2005.
76. A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2012, **8**, 1542-1555.
77. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878-3888.
78. D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668-1688.
79. D. A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, I. T.E. Cheatham, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J.Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R.Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao and D. M. Y. a. P. A. Kollman, 2018, **University of California, San Francisco**.
80. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
81. J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327-341.
82. T. Darden, D. York and L. Pedersen, *The Journal of chemical physics*, 1993, **98**, 10089-10092.
83. M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*, Oxford university press, 2017.
84. H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola and J. Haak, *The Journal of chemical physics*, 1984, **81**, 3684-3690.
85. D. R. Roe and T. E. Cheatham III, *Journal of Chemical Theory and Computation*, 2013, **9**, 3084-3095.
86. D. t. Cremer and J. Pople, *JACS*, 1975, **97**, 1354-1358.
87. T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris and T. E. Ferrin, *Protein Sci.*, 2018, **27**, 14-25.
88. B. M. Sattelle and A. Almond, *Glycobiology*, 2011, **21**, 1651-1662.
89. C. R. Ellis, B. Maiti and W. G. Noid, *Journal of the American Chemical Society*, 2012, **134**, 8184-8193.
90. M. Jana and S. Bandyopadhyay, *Physical Chemistry Chemical Physics*, 2012, **14**, 6628-6638.