SCIENTIFIC REPORTS

OPEN

Characterization of Conformational Ensembles of Protonated N-glycans in the Gas-Phase

Suyong Re[1], Shigehisa Watabe[1,3], Wataru Nishima[1,2], Eiro Muneyuki[3], Yoshiki Yamaguchi[4], Alexander D. MacKerell Jr.[5] & Yuji Sugita [1,2]

Ion mobility mass spectrometry (IM-MS) is a technique capable of investigating structural changes of biomolecules based on their collision cross section (CCS). Recent advances in IM-MS allow us to separate carbohydrate isomers with subtle conformational differences, but the relationship between CCS and atomic structure remains elusive. Here, we characterize conformational ensembles of gas-phase N-glycans under the electrospray ionization condition using molecular dynamics simulations with enhanced sampling. We show that the separation of CCSs between isomers reflects folding features of N-glycans, which are determined both by chemical compositions and protonation states. Providing a physicochemical basis of CCS for N-glycans helps not only to interpret IM-MS measurements but also to estimate CCSs of complex glycans.

Carbohydrates are ubiquitous components of living organisms. As over 50% of eukaryotic proteins and most of membrane proteins are glycosylated[1], their roles in biological processes, ranging from protein folding to biomolecular recognition events, have received considerable attention[2–4]. The biological activity of a glycoconjugate is often affected by a subtle change in a carbohydrate structure[5]. Carbohydrates have multiple linkage points and take branched structures. In addition, many rotamers arise along flexible linkages. The hydroxyl groups act as both donor and acceptor for hydrogen (H-) bonds, potentially forming a complex network of interactions to stabilize various conformations. Carbohydrates are commonly a complex mixture of various isomers, each having multiple conformations, and analysis of their structures in solution is extremely difficult[6].

Ion mobility mass spectrometry (IM-MS) is an analytical method increasingly used for obtaining information of conformational changes in polyatomic ions, such as unfolding or aggregation[7–9]. This method separates gas-phase ions based on the collision cross section (CCS) that reflects their sizes and shapes. IM-MS is capable of separating stereoisomers with identical mass but different conformations[10]. The method has been increasingly applied for structure determination of carbohydrates[11–14]. Recently, highly sensitive techniques have been reported for separating diastereomers of synthetic carbohydrates as well as those of glycoconjugates[15–21]. For instance, epimeric monosaccharides, which differ in chirality of only one hydroxyl group, can be differentiated[17]. However, conformational diversity as well as chemical heterogeneity of carbohydrates raises questions about the interpretation of IM-MS data. First, the experimentally derived CCS represents either a single species or multiple isomers and/or conformers. Second, the ionization of samples (e.g. protonation or sodiation) modifies their conformational ensemble. To date, limited investigations have been undertaken on gas-phase conformations at a finite temperature, while many studies have focused on solution conformations[22]. Thus, the determinants of CCSs of carbohydrates still remain largely unknown.

We previously reported a combined IM-MS and molecular dynamics (MD) simulation study on the identification of isomeric N-glycan structures[23]. Such a combined approach enabled us to link the measured CCS with structures at atomic resolution[7,24,25]. We measured the drift times of ten pyridylamino (PA)-glycans, including isomeric forms, in the presence of N_2 gas by using traveling wave ion mobility mass spectrometry. The PA group was introduced to enhance detection sensitivity. Doubly protonated PA-glycans ($[M + 2H]^{2+}$) predominated. Then,

[1]RIKEN Theoretical Molecular Science Laboratory, 2-1 Hirosawa, Wako, Saitama, 351-0198, Japan. [2]RIKEN iTHES, 2-1 Hirosawa, Wako, Saitama, 351-0198, Japan. [3]Graduate School of Science and Engineering, Chuo University, 1-13-27, Kasuga, Bunkyo-ku, Tokyo, 112-8551, Japan. [4]Structural Glycobiology Team, Systems Glycobiology Research Group, RIKEN Global Research Cluster, 2-1 Hirosawa, Wako, Saitama, 351-0198, Japan. [5]Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland, 21201, USA. Correspondence and requests for materials should be addressed to Y.S. (email: sugita@riken.jp)
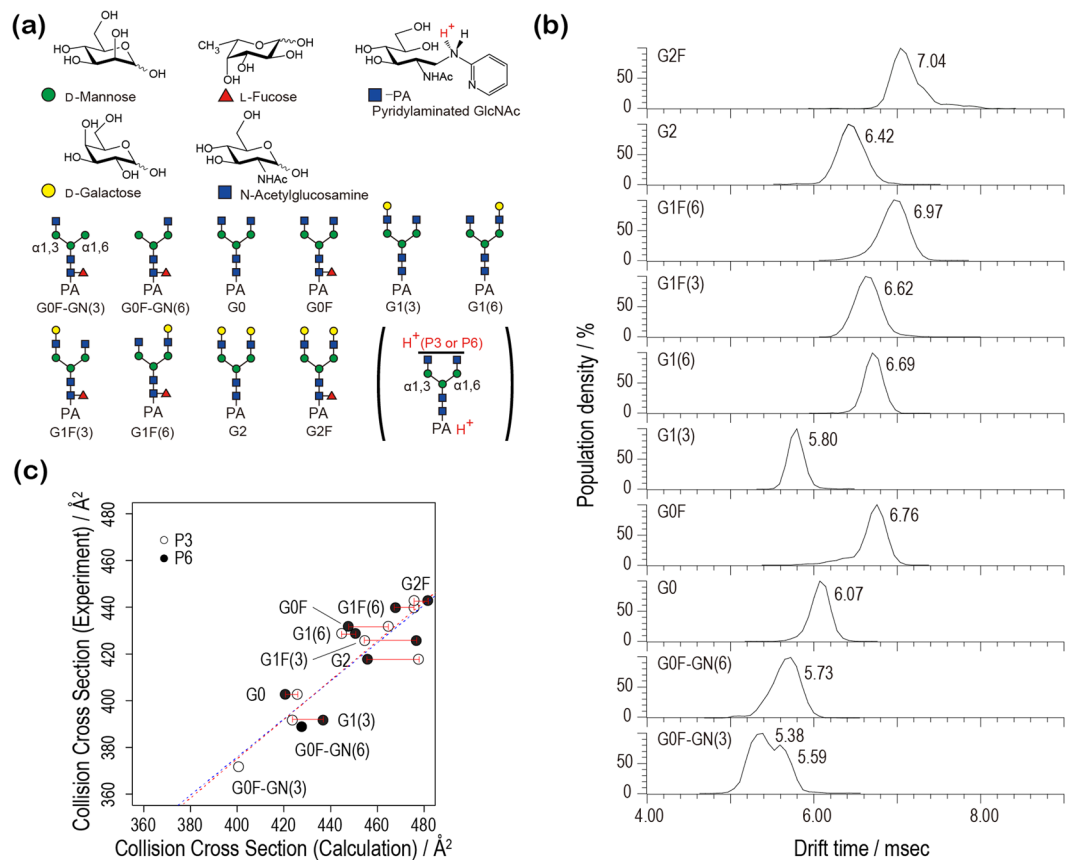
Figure 1. (**a**) Symbolic representation of the ten PA-glycans studied. (**b**) Experimentally observed arrival time distributions of the ten PA-glycans[23]. (**c**) Plot of calculated Collision Cross Sections (CCSs) in N_2 gas against experimental values obtained in N_2 gas. P3 and P6 indicate the different protonation states (P3: H^+ at α1-3 (white) and P6: H^+ at α1-6 branch (black)) (Dotted lines: regression lines with the correlation coefficient of 0.90 (P3, blue) and 0.89 (P6, red), respectively).

CCS values were derived by calibration with polyalanine, and different values were obtained for different isomers. Finally, we performed replica-exchange molecular dynamics (REMD) simulations[26] of isomeric N-glycans. The REMD simulation facilitated transitions between rotameric isomers that rarely occur in conventional MD simulations, allowing us to obtain conformational ensembles with high statistical accuracy. CCS values were calculated using the MOBCAL program[27,28]. The CCS is commonly calculated for a few selected low-energy conformers and compared with experimentally obtained values[29,30]. Our approach was different in that we obtained a population-weighted CCS based on the ensemble of conformations. We showed that the isomeric N-glycans fold to unique shapes, giving distinct CCS values.

In the present study, we computationally investigated CCS determinants of N-glycans to predict CCSs from the structural information. To this end, we extended the previous simulation study in several ways. First, the target system is expanded to ten pyridylamino (PA)-glycans (Fig. 1a). Second, we explicitly model doubly protonated PA-glycans ([M + 2H]²⁺) to directly compare with experiment under the electrospray ionization condition[23]. The protonation sites are not known experimentally. Considering that the amide group is more readily protonated than the hydroxyl group[31] and that protons electrostatically repel each other, we built two models with H^+ on PA and on the distal N-acetylglucosamine (GlcNAc) of either the α1-3 (P3) or α1-6 branch (P6). The consideration of protonation states adds further complexity to the calculations. As force field parameters of protonated species are not available in CHARMM36[32], which is used to perform MD simulations, the missing parameters for protonated PA and GlcNAc were specially developed for the present study. Third, we calculate CCS values in N_2 drift gas using the trajectory method[27,33], while a simplest projection approximation assuming He drift gas condition was used in our previous study[23]. We use the modified version of the MOBCAL program and newly parametrized N_2-based trajectory method algorithm[34]. The Replica-Exchange INter-face (REIN) is used for the REMD simulations[35]. All analyses were done using the trajectories at 300 K. Details of modeling are given in Supplementary information.

Results and Discussion

Experimental collision cross section (CCS) for 10 PA-glycans. The studied system (Fig. 1a) covers representative N-glycan modifications, e.g. core-fucosylation and/or galactosylation, which are known to modulate protein functions. We previously reported the drift time of doubly protonated PA-glycans in N_2 gas by using traveling wave ion mobility mass spectrometry (Fig. 1b)[23]. In the present work, we recalibrated CCS values of

2

Glycan	ion species	m/z	drift time (msec.)[a]	CCS in N_2 (Å^2)
G0F-GN(3)	$[M+2H]^{2+}$	669.8	5.31	371
G0F-GN(6)	$[M+2H]^{2+}$	669.8	5.73	388
G0	$[M+2H]^{2+}$	698.3	6.07	402
G0F	$[M+2H]^{2+}$	771.3	6.76	431
G1(3)	$[M+2H]^{2+}$	779.3	5.80	391
G1(6)	$[M+2H]^{2+}$	779.3	6.69	428
G1F(3)	$[M+2H]^{2+}$	852.3	6.62	425
G1F(6)	$[M+2H]^{2+}$	852.3	6.97	440
G2	$[M+2H]^{2+}$	860.3	6.42	417
G2F	$[M+2H]^{2+}$	933.4	7.04	442
Polyalanine (n)[b]				
10	$[M+2H]^{2+}$	365.0	3.24	not available
11	$[M+2H]^{2+}$	400.5	3.52	296
12	$[M+2H]^{2+}$	436.0	3.79	309
13	$[M+2H]^{2+}$	471.5	4.07	320
14	$[M+2H]^{2+}$	507.0	4.35	333
15	$[M+2H]^{2+}$	542.5	4.69	344

Table 1. Experimentally measured drift times and estimated collision cross sections (CCSs) in N_2 gas for ten doubly protonated PA-glycans. CCS values of doubly protonated polyalanines used for the calibration are also shown for comparison. [a]Ref.[23]. [b]Ref.[36]. A good linearity was found for the drift time (t) and CCS of polyalanine ($CCS = 41.3 \times t + 152$ with $R^2 = 0.996$).

		Calculated CCS in N_2 gas		Calculated CCS in He gas	
	Experiment	P3	P6	P3	P6
G0F-GN(3)[a]	371	400 (7.8)	—	279	—
G0F-GN(6)	388	—	427 (10.1)	—	296
G0	402	425 (5.7)	420 (4.5)	298	292
G0F	431	464 (7.7)	447 (3.7)	328	315
G1(3)	391	423 (8.2)	436 (11.5)	297	309
G1(6)	428	444 (3.7)	450 (5.1)	313	318
G1F(3)	425	454 (6.8)	476 (12.0)	322	338
G1F(6)	440	475 (8.2)	467 (6.4)	340	336
G2	417	477 (14.4)	455 (9.1)	340	325
G2F	442	475 (7.5)	481 (8.8)	342	350

Table 2. Calculated collision cross sections (in Å^2) in N_2 and He drift gas for each of two protonation states, P3 and P6, of ten PA-glycans. The experimental values in the presence of N_2 drift gas are also listed. The percentage difference between the experimental and calculated CCS in N_2 drift gas (%diff = $CCS^{calc} - CCS^{exp}/CCS^{exp} \times 100\%$) are given in parenthesis. [a]The core GlcNAc, rather than that of the $\alpha1$-3 branch, was protonated.

doubly protonated PA-glycans using the absolute CCS values of doubly protonated polyalanine[36] in N_2 gas, rather than the singly protonated species used before[23]. They are listed in Table 1. Five points ($n = 11$, 12, 13, 14, and 15) were used for the calibration since the polyalanine with $n > 15$ becomes triply charged. A good linearity was found for the drift time and CCS of polyalanine as shown in Table 1. We thus consider that the standard curve can be used for the calculation of glycan CCS.

Calculated CCS. The calculated CCS values for the ten glycans in N_2 gas as well as in He gas are listed in Table 2. For both P3 and P6 states, they are plotted against experimental values in Fig. 1c. A single protonation state was considered for G0F-GN(3) (core GlcNAc protonated) and for G0F-GN(6) (see Methods for details). Calculated values coincide with experimental ones, with correlation coefficients of 0.90 (P3 state) and 0.89 (P6 state), and average percentage differences of 7.8% (P3 state) and 7.9% (P6 state), respectively (percentage differences were calculated as $CCS^{calc} - CCS^{exp}/CCS^{exp} \times 100\%$). These differences could arise from inaccuracies either in simulations (such as force field parameters) or in experimental calibration. For instance, the use of other calibrants, such as dextran, may improve the accuracy of experimental CCS values. CCS values for the two protonation states of each glycan are different (12 Å^2 on average), the magnitude depending on type (maximum 22 Å^2 for G1F(3) and G2, and minimum 5 Å^2 for G0). The sharp CCS distributions, which were observed experimentally[23], suggest that dehydration suppresses the dynamics of the sample ion under electrospray ionization conditions. Fig. 2a shows the root mean square fluctuations (RMSFs) of heavy atoms. For most of the glycans, except for G1F(6),
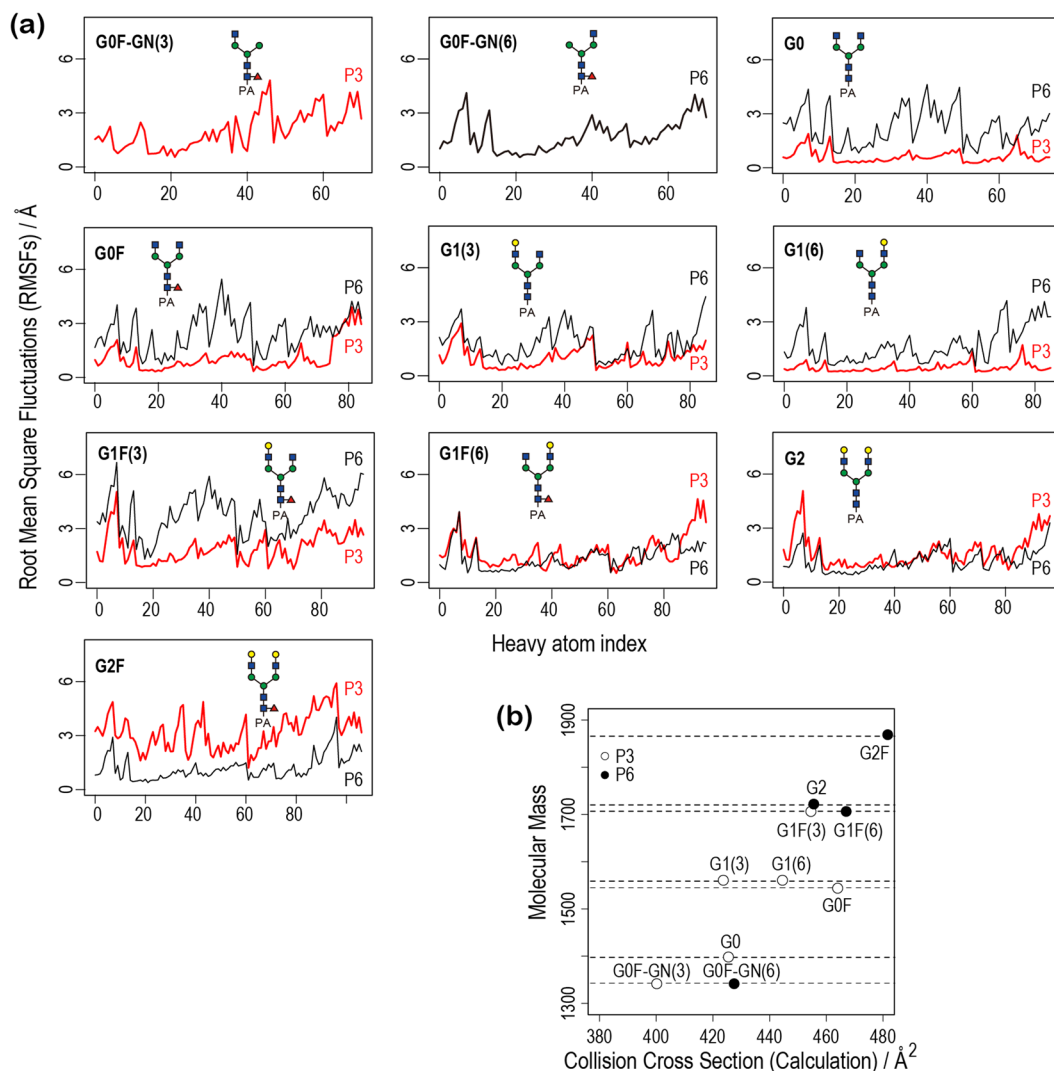
Figure 2. (**a**) RMSFs of heavy atoms calculated for the ten PA-glycans. The protonation states, P3 and P6, are marked using red and black lines, respectively. (**b**) Plot of calculated Collision Cross Sections (CCSs) along molecular mass. P3 and P6 indicate the different protonation states (P3: H$^+$ at α1-3 (white), P6: H$^+$ at α1-6 branch (black)).

G2, and G2F, the P3 state is more constrained than the P6 state. Taking less fluctuating structures, the correlation coefficient between experimental and theoretical values is 0.95. The calculated separations of CCSs for the three isomeric pairs, G0F-GN(3)/G0F-GN(6), G1(3)/G1(6), and G1F(3)/G1F(6), (13 Å^2–27 Å^2) well reproduce the experimental data (14 Å^2–37 Å^2) (Fig. 2b)[23]. Again consistent with experiment, the glycans with the GlcNAc addition or galactosylation at the α1-6 branch (G0F-GN(6), G1(6), and G1F(6)) have a larger cross section than corresponding isomers (G0F-GN(3), G1(3), and G1F(3)).

CCS distribution and conformational ensemble. We examined the calculated CCS distributions of the ten PA-glycans in terms of their conformational ensembles (see Methods for details). We also performed k-means clustering using the MMTSB toolset[37] (threshold: the root mean square deviation (RMSD) = 2.5 Å) and CCSs were calculated for each cluster. The results are summarized in Fig. 3. We found that the CCS distributions are relatively narrow (within the range of 50 Å^2) and mostly unimodal. This implies that CCS reflects a few distinct major conformers, and, consistent with this, only one or two clusters have a large population size. Other IM-MS studies on similar N-glycans also reported that a single conformer dominates[29,30]. Different orientation of the core chitobiose, particularly the PA group, gives a bimodal character to the CCS distributions of G1(3) and G0F (Fig. 4). This contrasts with the unimodal character in the experimentally observed arrival times of G1(3) and G0F (Fig. 1b). The discrepancy could arise because either two conformers interconvert in the experimental timescale or there is an inaccuracy of the force field parameters for the PA group. The major conformations of the ten N-glycans are classified into two folded forms described as compact globular and rod-like shapes (Fig. 5). In the compact globular shape, three arms (Manα1-3, Manα1-6, and core chitobiose) symmetrically bend over the central mannose at the branching point. On the other hand, in the rod-like shape, the Manα1-6 arm preferentially undergoes "backfolding" to the
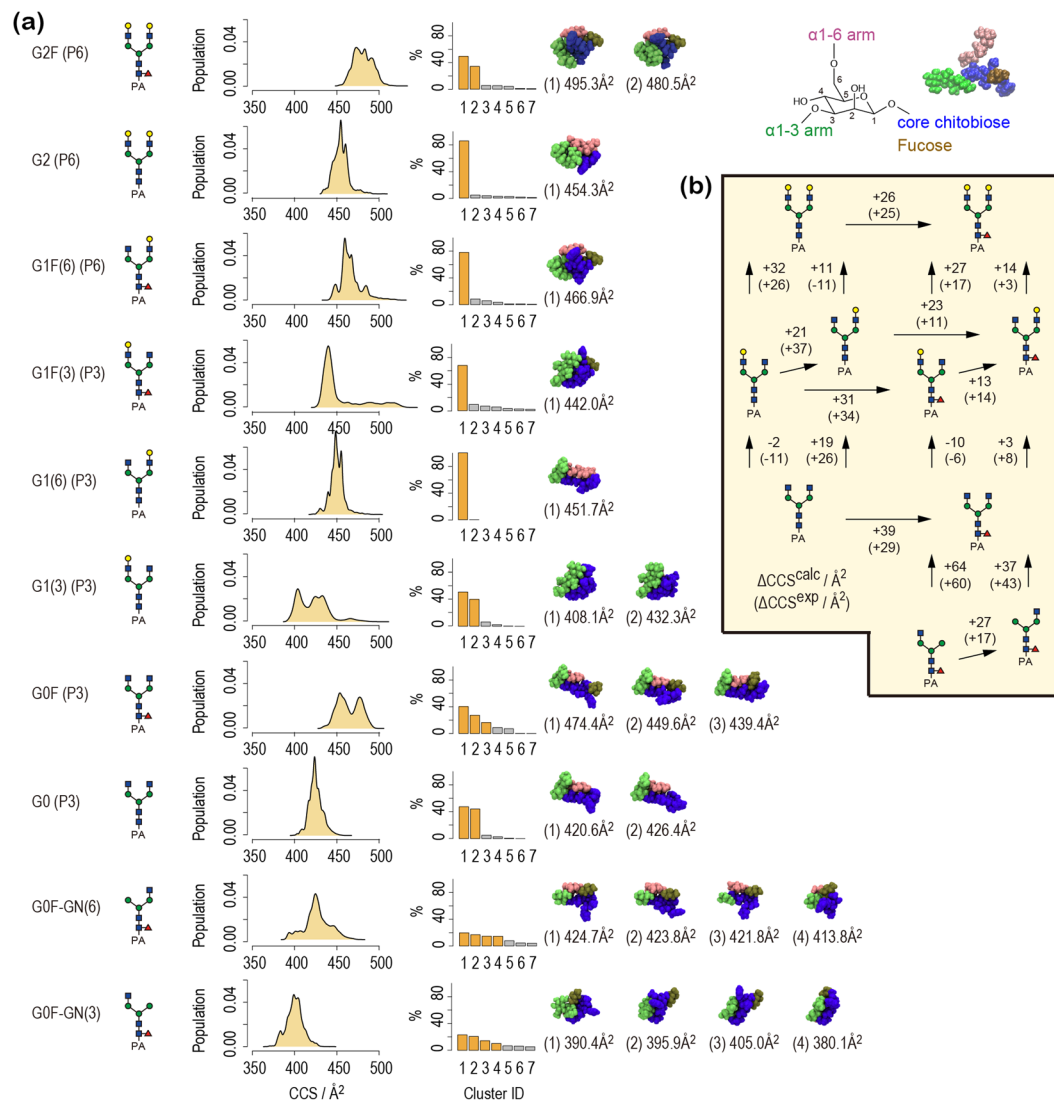
Figure 3. (**a**) CCS distributions, cluster populations that were obtained from k-means clustering using MMTSB toolset (threshold: RMSD = 2.5 Å), and major conformers making up more than 10% of the population (orange bars). The CCS values of major conformers are also given. Blue, green, red, and gold colors for the core chitobiose, α1-3 arm, α1-6 arm, and fucose residue respectively. (**b**) Differences in CCS values (ΔCCS^{calc}) are listed with the corresponding values from experiment (ΔCCS^{exp}).

core chitobiose. The glycans having the α1-3 arm longer than the α1-6 arm (G0F-GN(3), G1(3), and G1F(3)) have compact globular shapes (small CCSs), while other glycans favor rod-like shapes (large CCSs). An earlier study on doubly sodiated N-glycans showed that the backfolding of the α1-6 arm is commonly observed in the lowest energy conformers and folding of α1-3 arm leads to small CCSs[30]. Note that the CCS values of rod-like conformers change monotonically upon increasing the molecular mass by adding GlcNAc or Gal residues, implying that the folding of the α1-3 arm causes an unexpected change in CCS values. Further elaborated modelling of proton positions may be required to resolve the multiple features of glycans more accurately.

Difference in hydrogen bonding pattern. We analyzed H-bond interactions in G1(3) and G1(6), where one galactose is added to G0 (Fig. 5). We used 2,500 snapshots from the trajectory of each system. The following geometric definition was used to define H-bonds: $R_{XY} < 3.5$ Å and $\theta_{HXY} < 30°$, where R_{XY} is the distance between the heavy atoms X and Y, and θ_{HXY} is the angle between the X–H bond and the X–Y vectors. The residue-residue H-bond maps calculated for all isomeric pairs are shown in Fig. 6. In G1(3), which has the compact globular shape, H-bonds are frequently observed between intra-arm proximal residues. In contrast, we found high probabilities of H-bond interactions between distal residue pairs of the α1-6 arm and the core chitobiose, e.g. 2:GlcNAc—5:Gal (82%) and 1:PA-GlcNAc—6:Gal (52%) in G1(6). In all isomeric glycans, the compact globular shapes have less core-α1-6 H-bonds, while the total number of H-bonds is almost preserved (Table 3). Thus, the separation of CCS values is a result of different H-bonding patterns in the isomers. In the studied N-glycans, the flexible Manα1-6 linkage facilitates backfolding of the α1-6 arm to maximize stabilization via H-bonds. When the α1-3
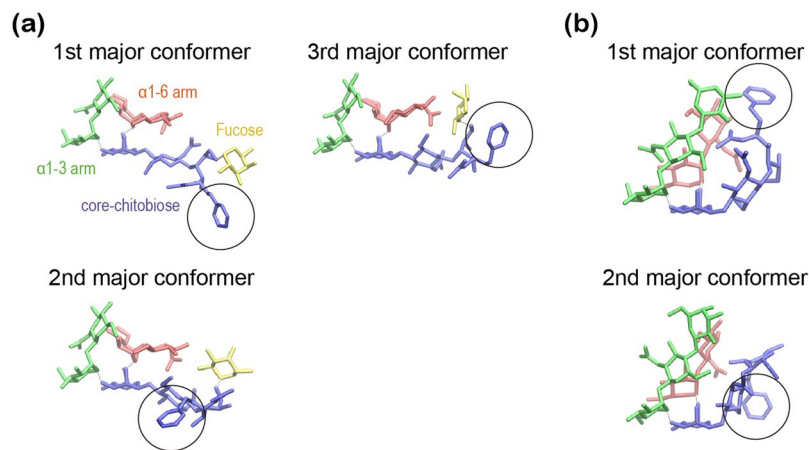
Figure 4. Comparison of major conformers: (**a**) three major conformers in G0F and (**b**) two major conformers in G1(3). Blue, green, red, and yellow colors for the core chitobiose, α1-3 arm, α1-6 arm, and fucose residue, respectively. The pyridylamino (PA) group is marked by a black circle.
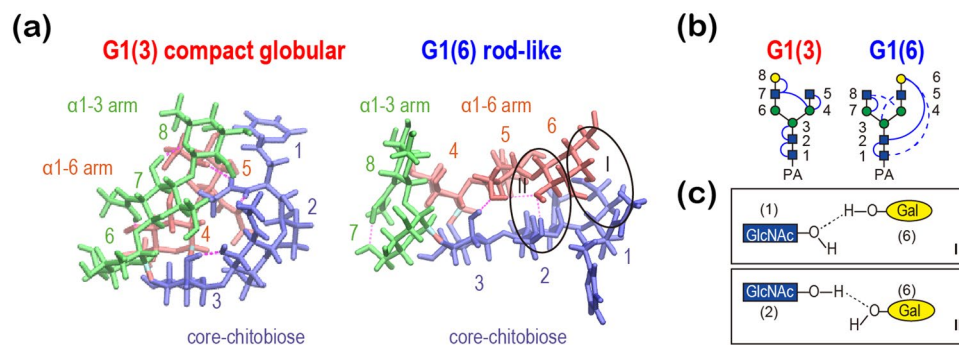

Figure 5. (**a**) Molecular structures of compact globular and rod-like "backfolding" forms. (**b**) Schematic illustration of H-bond networks (solid line for >70% and dashed line for >50% probability). The geometric definition was used to identify H-bonds: $R_{XY} < 3.5$ Å and $\theta_{HXY} < 30°$, where R_{XY} is the distance between heavy atoms X and Y, and θ_{HXY} is the angle between X–H bond and X–Y vectors. (**c**) Key inter-arm H-bonds in "backfolding" structure (H-bond with the first GlcNAc residue (I) and H-bonds with the second GlcNAc residue (II)).

arm is longer than the α1-6 arm, the rather rigid Manα1-3 linkage also undergoes backfolding because the resultant H-bond stabilization compensates for the distortion penalty. Interestingly, our previous simulation showed that N-glycans in solution favor the backfolding structures[38].

Conclusions

MD simulations have revealed that the protonation state as well as the conformational ensemble are important determinants for CCSs of gas-phase N-glycans. In the studied N-glycans, single conformers of either globular or rod-like shapes dominate and their CCSs are sufficiently different for identification of isomeric N-glycan structures. The results provide a physicochemical basis for IM-MS separation of complex glycan isomers as well as facilitating interpretation of experimental spectra. Furthermore, CCSs of new glycan structures can now be computationally predicted by combining simulations and MOBCAL prediction of the spectra. Recent developments in the field of CCS computations, such as IMoS[39,40], make accurate CCS calculations for appreciably larger polyatomic ions feasible. Combining CCSs with *ab initio* MD simulation even allows us to accurately model protonation states[41]. The present work lays the foundation for combining IM-MS and simulation to accurately identify complex glycan structures in living organisms.

Methods

Model preparation. We modeled 2H+ adducts of pyridylamino (PA)-glycans ([M + 2H]²⁺) that were studied in our previous work[23]. The positions of attached protons are not known experimentally. Considering that the amide group is more readily protonated than the hydroxyl group[31] and that two protons repel each other due to electrostatic repulsion, we built two models, both of which have one H+ on PA and a second H+ on the distal N-acetylglucosamine (GlcNAc) of either α1-3 (P3) or α1-6 branch (P6). The structures of protonated GlcNAc (PGLN) and pyridylaminated GlcNAc (PA) are shown in Supplementary Figure S1.
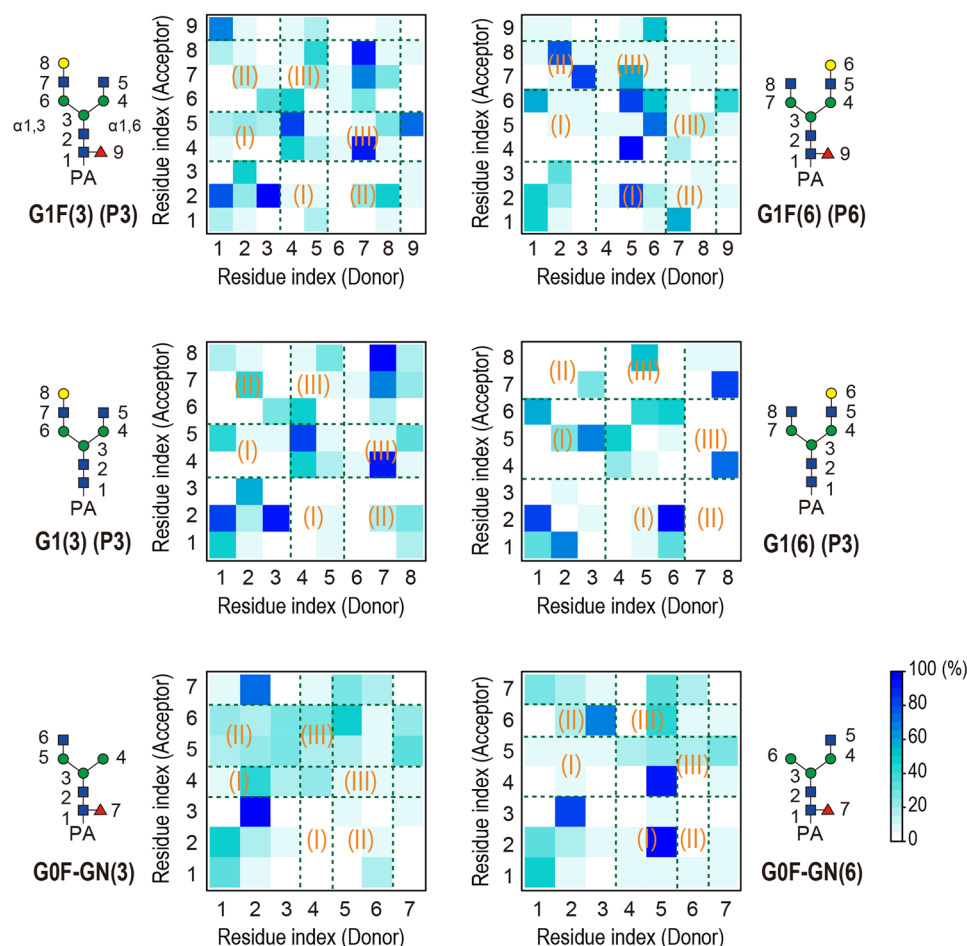
Figure 6. Residue-residue H-bond maps calculated for three pairs of isomeric PA-glycans with selected protonation states. The geometric definition was used to identify H-bonds: $R_{XY} < 3.5\,\text{Å}$ and $\theta_{HXY} < 30°$, where R_{XY} is the distance between heavy atoms X and Y, and θ_{HXY} is the angle between X–H bond and X–Y vectors. The regions of inter-branch interactions are highlighted with Greek numbers (I: core chitobiose-α1-6 branch, II: core chitobiose-α1-3 branch, III: α1-6 branch-α1-3 branch).

	G0F-GN	G1	G1F
CCS (exp.)	−27 (−17)	−21 (−37)	−12 (−15)
Av. Num. of H-bonds	2	1	0
% core-α1-6 H-bonds	−18	−28	−14
% core-α1-3 H-bonds	15	9	−11
% α1-3-α1-6 H-bonds	4	5	11
% Intra-arm H-bonds	−1	13	14

Table 3. Differences in CCS (Å^2) and H-bonds between rod-like and globular shapes ($X_{globular} - X_{rod}$; X = CCS, Average number of H-bonds, and Percentage of each type of H-bond). The experimental CCS values are given in parentheses.

In PGLN, both N-H and C=O groups could be protonated. In order to determine the protonation site, proton affinities were estimated for both groups using quantum chemical calculations at the B3LYP/6-31 G(d,p) level. The proton affinity was estimated by subtracting the electronic energy of protonated PGLN from that of neutral PGLN. The geometries of the two forms (protonated and neutral PGLN) were optimized both in gas-phase and in aqueous solution. The effect of solvent was taken into account by using the polarized continuum model[42,43] (dielectric constant of 78.35). All calculations were performed using the Gaussian09 program package[44]. The results show that the C=O group has a higher proton affinity than the N-H group both in the gas-phase (219.7 kcal/mol for N-H and 229.6 kcal/mol for C=O) and in solution (262.3 kcal/mol for N-H and 271.2 kcal/mol for C=O). Therefore, we decided to protonate the C=O group for PGLN. This proton position is supported by the estimated pKa using the Marvin pKa plugin from ChemAxon (Marvin 16.6.27.0 ChemAxon Ltd. http//www.chemaxon.com) (Supplementary Figure S1). The use

of such a pKa program helps to theoretically model proton positions that are not known experimentally, and may be followed by more expensive quantum chemical calculations to elaborate the model.

Note that we considered a single protonation state for G0F-GN(3), in which the core GlcNAc was protonated. Protonation of GlcNAc of the α1-3 branch in G0F-GN(3) gives a CCS value significantly larger than that of G0F-GN(6). This is inconsistent with experimental evidence[23], which indicates that the CCS of G0F-GN(3) is smaller than that of G0F-GN(6). In G0F-GN(3), the length of the flexible α1-6 branch is exceptionally short compared to the other N-glycans studied, and this could increase surface exposure of the core GlcNAc and the possibility of protonation. Indeed, protonation of the core GlcNAc gives a much better correlation between calculated and experimental CCSs.

Force field parameterization. The parameters for PGLN and PA were newly developed based on the CHARMM36 carbohydrate[32,45,46] and CHARMM General Force Field[47,48]. In order to preserve consistency with existing force fields, we assigned available parameters to fragments (GlcNAc for PGLN and D-Threitol, Alanine dipeptide, Morpholine, and 2-acetamide pyridine for PA) and determined the partial atomic charges and dihedral angle parameters only for atoms proximal to the added proton as shown in Supplementary Figure S2.

For the charges of PGLN, the CHARMM36 values for GlcNAc were used for the glucose moiety. The charges of the acetamide moiety were derived from the electrostatic potential fit (ESP) charges obtained from the quantum chemistry calculation of acetamide using the Merz-Singh-Kollman scheme[49,50] at the MP2/6-31 G(d) level. Similarly, the CHARMM36 charges for D-threitol and the CHARMM General Force Field charges for the alanine dipeptide, morpholine, and 2-acetamide pyridine were assigned to each of the four fragments for PA. Then, we modified the charges around the added proton (NH_2^+ group) according to the ESP charges calculated for PA. Note that we manually modified some values in order to preserve a total charge of +1. The final charges for the two residues are given in Supplementary Table S1.

For dihedral angle parameters, we optimized two missing parameters, $\chi 0$ in PA and $\chi 3$ in PGLN, to reproduce the *ab initio* potential energy surfaces (PESs) of the model compounds shown in Supplementary Figure S3. The *ab initio* PESs were obtained at the MP2/cc-pVTZ//MP2/6-31 G(d) level. The PES was scanned in 30° increments for the corresponding dihedral angles. The parameters were determined by fitting the empirical PES to the entire range of *ab initio* PES using the fit_dihedral.py script[51], with the optimized parameters listed in Supplementary Table S2. The newly determined parameter sets well reproduce the *ab initio* PESs for the rotations of $\chi 0$ and $\chi 3$ in the model compounds, as shown in Supplementary Figure S3. Note that the *ab initio* PESs for the rotations of $\chi 1$ and $\chi 2$ angles of PGLN were reasonably reproduced using the default parameters, although the added proton could affect these angles significantly.

Molecular dynamics simulations. The replica-exchange molecular dynamics (REMD) simulations[26] were performed for 18 PA-glycan systems in total: A single protonation state for G0F-GN(3) and G0F-GN(6), and two protonation states for the rest of the eight PA-glycans. The in-house interface program (REIN: Replica-Exchange Interface version 0.1)[35] was used for REMD simulations. The NAMD program[52] was used for each MD simulation inside of REIN. The initial configurations were constructed using the GLYCAM web portal (Glycam Biomolecule Builder, http://glycam.org)[53]. All the simulations were carried out in vacuum. The Langevin dynamics method was used to maintain the temperature (damping constant = 5 ps^{-1}). Nonbonded interactions were calculated without any truncation, using a large cutoff distance (50 Å). The bonds involving hydrogen atoms were kept rigid using SHAKE. A time step of 2 fs was used in the simulations. Each system was equilibrated using a constant temperature MD simulation at 300 K before the REMD simulation. The final snapshots were used to initiate the subsequent REMD simulations. Sixteen replicas were used to cover the temperatures ranging from 300 K to 1,058 K. To keep the chair conformation of the pyranose rings at high temperatures, six dihedral restraints were applied to the ring. Each replica was simulated for 50 ns with replica exchange trials at every 2 ps. The total simulation time was 0.8 μs (50 ns × 16 = 800 ns) for each system. Only the trajectory at 300 K was used for analyses. The convergence of simulations was assessed by checking the acceptance ratio, replica exchanges, and the relative population sizes of conformers. In Supplementary Figure S4, we show data obtained for G1F(3) as an example. The acceptance ratio is 28% on average. We confirmed random walks both in the replica space and temperature space. We plotted the relative population size of each conformer having different α1-6 linkage orientations. Each population changes drastically in the first 10 ns and converges reasonably within 50 ns. We obtained similar results for the other glycans.

Calculation of collision cross section (CCS). CCS values in N_2 drift gas were calculated using the trajectory method[27,33]. We used the modified version of the MOBCAL program and a newly parametrized algorithm of the N_2-based trajectory method[34]. The structures from REMD simulations were used for CCS calculations without optimization. In the MOBCAL program, we used the same partial charges as for the REMD simulations that we parametrized to reproduce the results of quantum chemical calculations as described in "Force field parametrization". Since the trajectory method is computationally very expensive, we used a conformational clustering to reduce the number of CCS calculations. First, we obtained representative conformers by the k-means method (threshold RMSD: 2.5 Å) implemented in the MMTSB Tool Set[37]. We obtained 13 clusters on average for each glycan. Then, the CCS value was calculated for the center structure of each cluster. We also performed clustering analysis with a small threshold (RMSD: 0.5 Å) to obtain a fine distribution of CCSs. This produces a maximum of 2,625 clusters and the CCS value was calculated for the center structure of each cluster. Finally, the CCS value of each center structure was weighted by multiplying the number of structures involved in the corresponding cluster, giving a fine distribution of CCSs as shown in Fig. 3. In the CCS calculation, we set the number of integration points to 25 in the Monte Carlo integration of the impact factor and orientation[54]. The CCS values in He drift gas were also calculated for comparison.

References

1. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *BBA-Gen. Subjects* **1473**, 4–8 (1999).
2. Dwek, R. A. Glycobiology: Toward understanding the function of sugars. *Chem. Rev.* **96**, 683–720 (1996).
3. Molinari, M. N-glycan structure dictates extension of protein folding or onset of disposal. *Nat. Chem. Biol.* **3**, 313–320 (2007).
4. Varki, A. Sialic acids in human health and disease. *Trends Mol. Med.* **14**, 351–360 (2008).
5. Li, W. Z. *et al.* Core fucosylation of IgG B cell receptor is required for antigen recognition and antibody production. *J. Immunol.* **194**, 2596–2606 (2015).
6. Duus, J., Gotfredsen, C. H. & Bock, K. Carbohydrate structural determination by NMR spectroscopy: Modern methods and limitations. *Chem. Rev.* **100**, 4589–4614 (2000).
7. Wyttenbach, T., Pierson, N. A., Clemmer, D. E. & Bowers, M. T. Ion mobility analysis of molecular dynamics. *Annu. Rev. Phys. Chem.* **65**, 175–196 (2014).
8. Bleiholder, C., Dupuis, N. F., Wyttenbach, T. & Bowers, M. T. Ion mobility-mass spectrometry reveals a conformational conversion from random assembly to beta-sheet in amyloid fibril formation. *Nat. Chem.* **3**, 172–177 (2011).
9. Zhou, M. *et al.* Ion mobility-mass spectrometry of a rotary ATPase reveals ATP-induced reduction in conformational flexibility. *Nat. Chem.* **6**, 208–215 (2014).
10. Fenn, L. S. & McLean, J. A. Structural separations by ion mobility-MS for glycomics and glycoproteomics. *Methods Mol. Biol.* **951**, 171–194 (2013).
11. Zhu, F. F., Lee, S., Valentine, S. J., Reilly, J. P. & Clemmer, D. E. Mannose7 glycan isomer characterization by IMS-MS/MS analysis. *J. Am. Soc. Mass Spectrom.* **23**, 2158–2166 (2012).
12. Zhu, F. F., Trinidad, J. C. & Clemmer, D. E. Glycopeptide site heterogeneity and structural diversity determined by combined lectin affinity chromatography/IMS/CID/MS techniques. *J. Am. Soc. Mass Spectrom.* **26**, 1092–1102 (2015).
13. Hofmann, J. *et al.* Estimating collision cross sections of negatively charged N-glycans using traveling wave ion mobility-mass spectrometry. *Anal. Chem.* **86**, 10789–10795 (2014).
14. Pagel, K. & Harvey, D. J. Ion mobility-mass spectrometry of complex carbohydrates: Collision cross sections of sodiated N-linked glycans. *Anal. Chem.* **85**, 5138–5145 (2013).
15. Hofmann, J., Hahm, H. S., Seeberger, P. H. & Pagel, K. Identification of carbohydrate anomers using ion mobility-mass spectrometry. *Nature* **526**, 241–244 (2015).
16. Hinneburg, H. *et al.* Distinguishing N-acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry. *Chem. Commun.* **52**, 4381–4384 (2016).
17. Both, P. *et al.* Discrimination of epimeric glycans and glycopeptides using IM-MS and its potential for carbohydrate sequencing. *Nat. Chem.* **6**, 65–74 (2014).
18. Gray, C. J. *et al.* Bottom-up elucidation of glycosidic bond stereochemistry. *Anal. Chem.* **89**, 4540–4549 (2017).
19. Kurulugama, R. T., Darland, E., Kuhlmann, F., Stafford, G. & Fjeldsted, J. Evaluation of drift gas selection in complex sample analyses using a high performance drift tube ion mobility-QTOF mass spectrometer. *Analyst* **14**, 6834–6844 (2015).
20. Yang, H. M. *et al.* Identification of structurally closely related monosaccharide and disaccharide isomers by PMP labeling in conjunction with IM-MS/MS. *Sci. Rep.* **6**, 28079 (2016).
21. Martens, J. *et al.* Molecular identification in metabolomics using infrared ion spectroscopy. *Sci. Rep.* **7**, 3363 (2017).
22. Diaz, D., Canales-Mayordomo, A., Canada, F. J. & Jimenez-Barbero, J. Solution conformation of carbohydrates: a view by using NMR assisted by modeling. *Methods Mol. Biol.* **1273**, 261–287 (2015).
23. Yamaguchi, Y., Nishima, W., Re, S. Y. & Sugita, Y. Confident identification of isomeric N-glycan structures by combined ion mobility mass spectrometry and hydrophilic interaction liquid chromatography. *Rapid Commun. Mass Spectrom.* **26**, 2877–2884 (2012).
24. Zhao, Y. J. *et al.* Investigating changes in the gas-phase conformation of Antithrombin III upon binding of Arixtra using traveling wave ion mobility spectrometry (TWIMS). *Analyst* **14**, 6980–6989 (2015).
25. Lanucara, F., Holman, S. W., Gray, C. J. & Eyers, C. E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **6**, 281–294 (2014).
26. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).
27. Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. Structural information from ion mobility measurements: Effects of the long-range potential. *J. Phys. Chem.* **100**, 16082–16086 (1996).
28. Shvartsburg, A. A. & Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **261**, 86–91 (1996).
29. Williams, J. P. *et al.* Characterization of simple isomeric oligosaccharides and the rapid separation of glycan mixtures by ion mobility mass spectrometry. *Int. J. Mass Spectrom.* **298**, 119–127 (2010).
30. Plasencia, M. D. *et al.* Resolving and assigning N-linked glycan structural isomers from ovalbumin by IMS-MS. *J. Am. Soc. Mass Spectrom.* **19**, 1706–1715 (2008).
31. Dewar, M. J. S. & Dieter, K. M. Evaluation of AM1 calculated proton affinities and deprotonation enthalpies. *J. Am. Chem. Soc.* **108**, 8075–8086 (1986).
32. Guvench, O. *et al.* CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J. Chem. Theory Comput.* **7**, 3162–3180 (2011).
33. Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. Mobilities of carbon cluster ions: Critical importance of the molecular attractive potential. *J. Chem. Phys.* **108**, 2416–2423 (1998).
34. Campuzano, I. *et al.* Structural characterization of drug-like compounds by ion mobility mass spectrometry: Comparison of theoretical and experimentally derived nitrogen collision cross sections. *Anal. Chem.* **84**, 1026–1033 (2012).
35. Miyashita, N., Re, S. & Sugita, Y. REIN: Replica-Exchange INterface for simulating protein dynamics and function. *Int. J. Quantum Chem.* **115**, 325–332 (2015).
36. Bush, M. F., Campuzano, I. D. G. & Robinson, C. V. Ion mobility mass spectrometry of peptide ions: Effects of drift gas and calibration strategies. *Anal. Chem.* **84**, 7124–7130 (2012).
37. Feig, M., Karanicolas, J. & Brooks III, C. L. MMTSB Tool Set: Enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.* **22**, 377–395 (2004).
38. Nishima, W., Miyashita, N., Yamaguchi, Y., Sugita, Y. & Re, S. Effect of bisecting GlcNAc and core fucosylation on conformational properties of biantennary complex-type N-glycans in solution. *J. Phys. Chem. B* **116**, 8504–8512 (2012).
39. Larriba, C. & Hogan, C. J. Ion mobilities in diatomic gases: Measurement versus prediction with non-specular scattering models. *J. Phys. Chem. A* **117**, 3887–3901 (2013).
40. Larriba-Andaluz, C. & Hogan, C. J. Collision cross section calculations for polyatomic ions considering rotating diatomic/linear gas molecules. *J. Chem. Phys.* **141** (2014).
41. Arcella, A. *et al.* Structure and dynamics of oligonucleotides in the gas phase. *Angew. Chem. Int. Edit.* **54**, 467–471 (2015).
42. Tomasi, J., Mennucci, B. & Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **105**, 2999–3093 (2005).
43. Scalmani, G. & Frisch, M. J. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.* **132**, 114110 (2010).
44. Gaussian 09, Revision A.02, Frisch, M. J. T. *et al.* (Gaussian, Inc., Wallingford CT 2009).
45. Guvench, O. *et al.* Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **29**, 2543–2564 (2008).

46. Guvench, O., Hatcher, E., Venable, R. M., Pastor, R. W. & MacKerell, A. D. CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Chem. Theory Comput.* **5**, 2353–2370 (2009).
47. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
48. Yu, W. B., He, X. B., Vanommeslaeghe, K. & MacKerell, A. D. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.* **33**, 2451–2468 (2012).
49. Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **5**, 129–145 (1984).
50. Besler, B. H., Merz, K. M. & Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **11**, 431–439 (1990).
51. Guvench, O. & MacKerell, A. D. Automated conformational energy fitting for force-field development. *J. Mol. Model* **14**, 667–679 (2008).
52. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
53. Woods, R. J. *GLYCAM Web; Complex Carbohydrate Research Center*, University of Georgia: Athens, GA, 2005–2011).
54. D'Atri, V., Porrini, M., Rosu, F. & Gabelica, V. Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure. *J. Mass Spectrom.* **50**, 711–726 (2015).

Acknowledgements

Author Contributions

Y.S. designed the research; S.W., W.N. and S.R. conducted the MD simulations, CCS calculations, and trajectory analyses; Y.Y. performed the experiments; M.E. and A.D.M. contributed to the discussion of results and manuscript refinement; S.R., Y.Y., and Y.S. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-20012-0.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.