Carbohydrate Research 344 (2009) 491–500

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations

Laercio Pol-Fachin [a], Claudia Lemelle Fernandes [a], Hugo Verli [a,b,*]

[a] Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av Bento Gonçalves 9500, CP 15005, Porto Alegre 91500-970, RS, Brazil
[b] Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Av Ipiranga 2752, Porto Alegre 90610-000, RS, Brazil

ARTICLE INFO

Article history:
Received 3 September 2008
Received in revised form 3 December 2008
Accepted 26 December 2008
Available online 3 January 2009

Keywords:
Glycoproteins
Molecular dynamics
Carbohydrates
Glycosidic linkage
GROMOS96

ABSTRACT

Considering the small number of papers assessing the conformational profile of glycoproteins through molecular dynamics (MD) simulations, the current work reports on a systematic analysis of the performance of the GROMOS96 43a1 force field and Löwdin HF/6-31G**-derived atomic charges in the conformational description of glycoproteins. The results substantiate the accuracy of the computational representation of glycoprotein conformational ensembles in aqueous solution based on their agreement to available experimental information, supporting further contributions of computational techniques, mainly MD, in future studies on the characterization of glycoprotein structure and function.

1. Introduction

Glycoproteins and glycopeptides are known to play important roles in many biological events, such as cell adhesion, cell–cell communication, immune response, intracellular targeting, and protease resistance.[1,2] These glycans can be located at various positions of the protein surface, depending on specific consensus sequences, and are linked only through specific amino acid residues. The linkages most frequently found in nature are the N-glycosidic linkages, involving mainly asparagine (Asn) and *N*-acetyl-D-glucosamine (GlcNAc) residues, and the O-glycosidic linkages, mostly including serine (Ser) or threonine (Thr) and, for instance, D-galactose (Gal), *N*-acetyl-D-galactosamine (GalAc), L-fucose (Fuc), or D-mannose (Man) residues. Structurally, these carbohydrate moieties impact several physicochemical properties of proteins, including hydration and polarity, frequently participating in its folding and conformational stabilization.[1,2]

The comprehension of the three-dimensional structure and the dynamical properties of both protein and carbohydrate moieties of glycoproteins is a requirement for a better understanding of the molecular basis of their interaction with each other, with the surrounding environment, and with their molecular targets. In this context, several experimental methods have been applied to study glycosylated proteins and glycans, such as X-ray crystallography and NMR spectroscopy. Concerning the former technique, the conformational flexibility of the glycan at the protein surface usually prevents such molecules from being crystallized,[3,4] but when crystals are available, the electron density is affected by the high thermal motion of the glycan moiety, compromising the accuracy of the geometry beyond the rigid core region of N-glycans.[5] NMR methods provide the solution ensemble of conformations in a set of average three-dimensional models, but an unambiguous determination of the complete conformational space of glycoproteins is hampered because only from one to three contacts across a given glycosidic linkage are usually detected.[6]

Considering the difficulties associated with the experimental determination of glycoproteins structure and conformation, MD simulations have been used to clarify the dynamic aspects of macromolecules. In particular, such simulations support the study of these molecules mimicking their natural environment and describe their conformational properties with a reasonable level of accuracy. On the other hand, only a small number of studies have assessed the profile of glycoproteins and glycopeptides through MD simulations, employing different set of parameters for their glycan parts, such as AMBER, CHARMm, and CVFF.[7–9] In comparison to such parameters, our group has been working on MD conformational representation of carbohydrates, polysaccharides, and glycosaminoglycans[10–13] based on the GROMOS96 43a1 force field,[14] enhanced by Löwdin HF/6-31G**-derived atomic charges.[10,12] This approach supports the adequate description of such molecules in aqueous solutions.

Therefore, the current work intends to evaluate the capability of the GROMOS96 43a1 force field and Löwdin HF/6-31G**-derived

* Corresponding author. Tel.: +55 51 3308 7770; fax: +55 51 3308 7309.
E-mail address: hverli@cbiot.ufrgs.br (H. Verli).

0008-6215/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2008.12.025

atomic charges, and to represent adequately the conformational ensemble of glycoproteins through MD simulations in aqueous solutions. The results of these simulations were compared to experimental data detailing the geometry of the glycosidic linkages between two monosaccharides and between a monosaccharide and an amino acid. Additionally, the effects of glycosylation upon the protein moiety of the glycoproteins were also assessed. Based on the data, the employed methodology may be expected to further contribute to structural and functional studies of glycoconjugates through reliable and accessible methodology.

2. Experimental

2.1. Computational methods

2.1.1. Nomenclature and software

The nomenclature recommendations and symbols as proposed by IUPAC[15] were used. The relative orientation of a pair of contiguous carbohydrate residues is described, for different types of linkages, by two or three torsional angles at the glycosidic linkage. For a (1→X) linkage, where 'X' is '2', '3', '4' or '6' for the (1→2), (1→3), (1→4) or (1→6) linkages, respectively, the ϕ and ψ are defined as shown below:

$$\phi = \mathrm{O5{-}C1{-}OX{-}CX} \quad (1)$$

$$\psi = \mathrm{C1{-}OX{-}CX{-}C(X-1)} \quad (2)$$

For a (1→6) linkage, the ω is defined as shown below:

$$\omega = \mathrm{O6{-}C6{-}C5{-}C4} \quad (3)$$

The manipulation of structures was performed with MOLDEN[16] and VMD,[17] the secondary structure content analyses were performed with PROCHECK,[18] and all the MD calculations and remaining analyses were performed using GROMACS simulation suite[19] and GROMOS96 force field,[14] with all saccharide topologies generated with PRODRG.[20]

2.1.2. Topology construction

The oligosaccharide fragments present in the glycoproteins under the PDB codes 1CDR,[21] 1CDS,[21] 1GYA,[22] 1HD4,[23] and 1FF7[24] (Fig. 1) were described in GROMOS96 43a1 force field through PRODRG server-derived topologies,[20] which were further refined by atomic charges, as described previously.[10,12] Briefly, each monosaccharide was submitted to HF/6-31G** energy minimization, followed by Hessian matrix analyses to characterize them unequivocally as true minima at the potential energy surface. The residues were divided in charge groups, while improper dihedrals were included as necessary to preserve the hexopyranose conformation of each monosaccharide in accordance with its expected form in aqueous solution: $^4C_1$ for D-GlcNAc, $1C_4$ for L-Fuc, $4C_1$ for D-Man, and $^4C_1$ for D-Gal. Additionally, proper dihedrals, as described in GROMOS96 43a1 force field for glucose, were also included in the PRODRG obtained topologies in order to support stable simulations. Concerning the N- and O-linkages between the carbohydrate and amino acid residues, the atomic charges were calculated for the Asn-carbohydrate (GlcNAc $C_1 = 0.110$, Asn $N_\delta = -0.265$, and Asn $H_{N\delta} = 0.155$) or Ser-carbohydrate (Fuc $C_1 = 0.150$, Ser $O_\gamma = -0.320$,

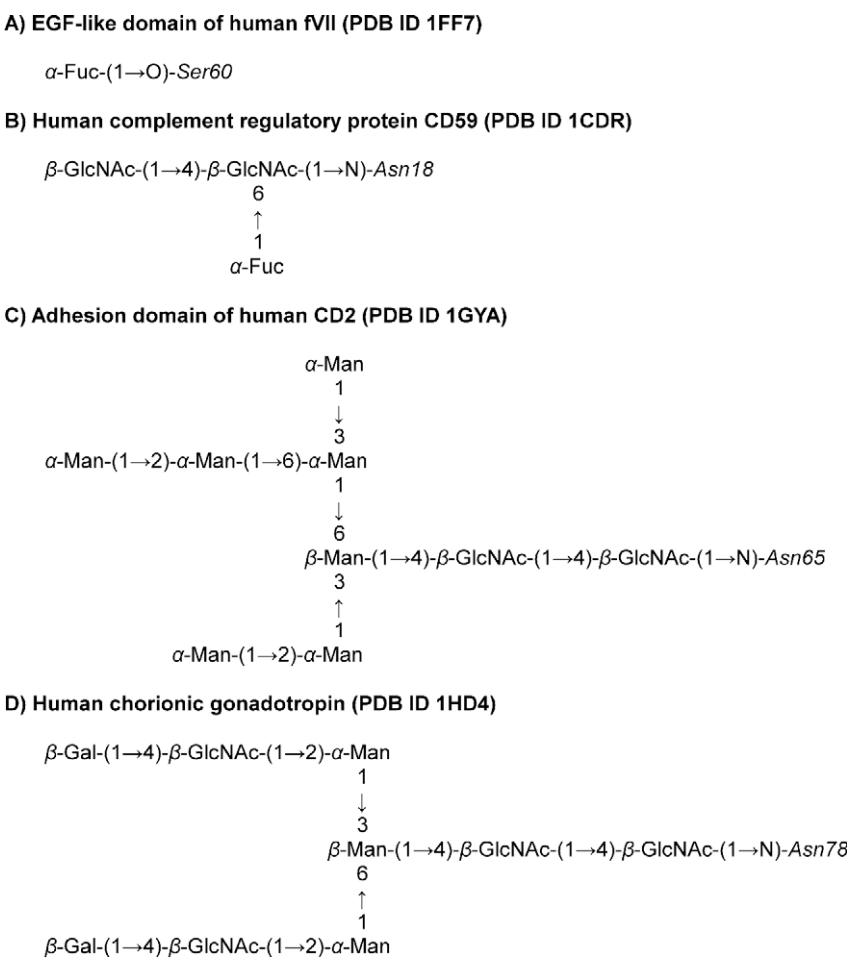


Figure 1. Schematics of the studied glycans and the corresponding glycosylated amino acids in the studied proteins.

and Ser C_β = 0.170) connection charge groups. Such an approach, based on parameters derived from minimum energy conformations, has been successfully applied in previous work in carbohydrate modeling.[10–13,25,26]

2.1.3. MD simulations

The structures of glycoproteins, as obtained from NMR data, were retrieved from the PDB, including the human complement regulatory protein CD59 (CD59), in its non-glycosylated (PDB ID 1CDQ) and glycosylated (PDB codes 1CDR and 1CDS) forms,[21] the α-subunit of the human chorionic gonadotropin (α-hCG), in both non-glycosylated (PDB ID 1DZ7)[27] and glycosylated (PDB ID 1HD4) structures,[23] the first epidermal growth factor-like (EGF-like) domain of the human blood coagulation fVII in its non-glycosylated (PDB ID 1F7E) and glycosylated (PDB ID 1FF7) forms,[24] and the adhesion domain of human CD2 ($hsCD2_{105}$) in its glycosylated (PDB ID 1GYA)[22] and non-glycosylated (PDB ID 1CDB)[28] forms, in a total of five glycosylated (Fig. 1) and four non-glycosylated proteins. These structures were solvated in a rectangular box using periodic boundary conditions and SPC water model.[29] Amino acids ionization was automatically adjusted to the plasma pH, with histidines being non-ionized and at the N_τ–H tautomeric form. Counterions (Na^+ and Cl^-) were added to neutralize the systems, whenever needed. The employed MD protocol was based on previous studies, as described.[10,12] The Lincs method[30] was applied to constrain covalent bond lengths, allowing an integration step of 2 fs after an initial energy minimization using Steepest Descent algorithm. Electrostatic interactions were calculated with Particle Mesh Ewald method.[31] Temperature and pressure were kept constant by coupling glycoproteins, ions, and solvent to external temperature and pressure baths with coupling constants of τ = 0.1 and 0.5 ps,[32] respectively. The dielectric constant was treated as ε = 1, and the reference temperature was adjusted to 310 K. The systems were slowly heated from 50 to 310 K, in steps of 5 ps, each one increasing the reference temperature by 50 K. Each simulation was extended to 50 ns, without any restraint, while a reference value of 3.5 Å between heavy atoms was considered for a hydrogen bond and a cutoff angle of 30° was used between hydrogen–donor–acceptor.[19] As the NMR selected structures include several models in the same PDB file, a total of eight glycoproteins (one model for 1CDS and 1FF7, and two models for 1CDR, 1GYA, and 1HD4), plus four non-glycosylated counter-parts were simulated to achieve a broader conformational description through MD. Details of the simulated models are included in Table 1. Additional simulations of the non-glycosylated proteins were performed with OPLS-AA force field in order to check the pattern of secondary structure as observed in GROMOS96 MD.

3. Results and discussion

3.1. Simulation systems

The analysis of GROMOS96 43a1 force field performance on glycoproteins conformational description was based on the MD simulation of four different proteins: (1) the EGF-like domain of human fVII; (2) the human complement regulatory protein CD59; (3) the adhesion domain of human CD2; and (4) the α-subunit of the human chorionic gonadotrophin. These proteins were simulated in both glycosylated and non-glycosylated forms, taking their NMR-derived model as the starting geometry. Among these molecules, three have N-linked glycans and one has an O-linked glycan. These systems were analyzed taking the NMR-obtained ensemble from each macromolecule as reference, to supply the information of GROMOS96 43a1 force field performance on the description of glycan–protein linkage of different glycosidic linkages composing the glycan structures, and the influence of glycosylation on the protein moiety of each molecule.

Actually, conformational data derived from single crystals of glycoconjugates should be considered carefully since the crystal environment may supply distorted hexopyranose rings and/or be submitted to packing effects, which are also known to occur for other classes of molecules, such as DNA[33] and proteins.[34,35] In contrast, although NMR techniques are not considered to provide an unambiguous determination of the complete conformational space of glycans,[3] it does provide solution conformations for the studied glycans which, in a set of average three-dimensional structures, could provide a realistic sample of glycans-composing disaccharide conformations and, consequently, a better reference for the experimental support of MD in solutions.

3.2. N- and O-glycosidic linkage geometry

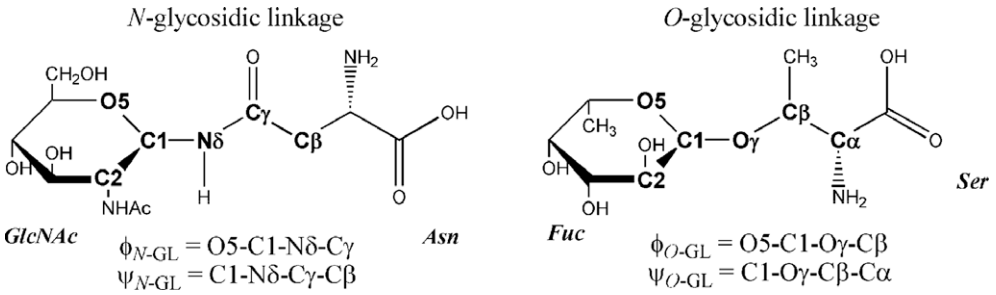
One of the main features determining the orientation of carbohydrates attached to proteins lies in its linkage to specific amino acid residues, such as asparagine and serine/threonine. The N-glycosidic linkages have been extensively studied by both NMR and X-ray methods, while a lower amount of experimental data is available for the O-glycosidic linkages, probably due to the increased flexibility and versatility of these linkages, involving distinct amino acid and monosaccharide residues. Therefore, the performed simulations of glycoproteins were initially analyzed by considering the connection between carbohydrates and proteins, using a set of experimental data as reference, as shown in Table 1 for average geometries and in Figure 2 for the respective conformational behavior of each linkage.

As can be observed, previous analysis of X-ray data indicates that the $\phi_{N\text{-GL}}$ dihedral angle presents higher flexibility compared to the $\psi_{N\text{-GL}}$ dihedral angle, with most X-ray structures presenting the $\phi_{N\text{-GL}}$ with a 80° amplitude (between −140° and −60°) and $\psi_{N\text{-GL}}$ with a 40° amplitude (between 160° and −160°).[36] Additional data, based on ~500 crystal structures obtained from the PDB, indicate a lower flexibility, in which the GlcNAc-(1→N)-Asn linkage exhibits an average $\phi_{N\text{-GL}}$ geometry at ~−100° and $\psi_{N\text{-GL}}$ at ~180°.[37] One of the most recent papers upon such linkages, combining X-ray crystallography and both HF and B3LYP with the 6-31+G* basis set, on models and analogs of GlcNAc-(1→N)-Asn,[38] also observed similar geometries for such dihedral angles. This large amplitude of the $\phi_{N\text{-GL}}$ also suggests the occurrence of multiple conformer populations co-existing in solution and being sustained depending on the crystallographic environment, as indeed observed in MD (Fig. 2A).

Hence, all glycoproteins presented equivalent conformational patterns around the GlcNAc-(1→N)-Asn linkage during the performed simulations, in a behavior apparently independent on the surrounding protein scaffold (Fig. 2A and Supplementary data). The curve representing such conformational patterns include about 81% of the NMR experimental geometries of $\phi_{N\text{-GL}}$, as well as 99% in $\psi_{N\text{-GL}}$ (in a total of 226 geometries for each dihedral angle),[21–23,39–42] indicating a good reliability of the simulations to predict the dynamics of N-linked glycans.

The geometry of O-glycosidic linkages was also evaluated and compared to three NMR studies, containing a fucose residue linked to two amino acid residues, Ser[24] or Thr,[43] as well as an additional structure presenting a 3-*O*-methyl-mannose residue linked to Thr.[44] It may be observed that the $\phi_{O\text{-GL}}$ and $\psi_{O\text{-GL}}$ dihedral angles of these linkages adopt similar geometries regardless of whether Ser and Thr is the involved amino acid residue (Table 1). Different from N-glycosidic linkages, no major flexibility differences could be observed when comparing $\phi_{O\text{-GL}}$ and $\psi_{O\text{-GL}}$ dihedral angles. In addition, the $\phi_{O\text{-GL}}$ presented two major MD conformations, which are both found in NMR data, around −70° (in 1URK and 1FF7 PDB

Table 1
Comparison of the dihedral angles around N- and O-glycosidic linkages, including X-ray, NMR, and MD simulations data



Glycosidic linkage	Structures	Dihedral angle (°)[†]			
		$\varphi_{N\text{-GL}}$	$\psi_{N\text{-GL}}$	$\varphi_{O\text{-GL}}$	$\psi_{O\text{-GL}}$
GlcNAc-(1→N)-Asn	X-ray[a]	−60 to −140	160 to −160	—	—
	X-ray[b]	−99 ± 21	177 ± 12	—	—
	NMR[c]	−91 ± 45	180 ± 7	—	—
	NMR—1CDR[d]	−77 ± 45	180 ± 0	—	—
	MD—1CDR[e,*]	−121 ± 29	−179 ± 14	—	—
	MD—1CDR[f,*]	−126 ± 28	179 ± 13	—	—
	NMR—1CDS[d]	−63 ± 44	180 ± 0	—	—
	MD—1CDS[g,*]	−119 ± 30	−177 ± 14	—	—
	NMR—1GYA[d]	−99 ± 15	179 ± 2	—	—
	MD—1GYA[h,*]	−121 ± 28	178 ± 13	—	—
	MD—1GYA[i,*]	−116 ± 29	−174 ± 14	—	—
	NMR—1HD4[d]	−86 ± 50	177 ± 6	—	—
	MD—1HD4[j,*]	−120 ± 30	−174 ± 14	—	—
	MD—1HD4[k,*]	−116 ± 30	−178 ± 12	—	—
Fuc-(1→O)-Ser	NMR—1FF7[d]	—	—	−67 ± 10	178 ± 10
	MD—1FF7[l,*]	—	—	−110 ± 29	178 ± 22
Fuc-(1→O)-Thr	NMR—1URK[d]	—	—	−104 ± 30	−154 ± 76
3-*O*-methyl-Man-(1→O)-Thr	NMR—2HGO[d]	—	—	72 ± 3	120 ± 22

[†] Values presented as averages ± standard deviation values, obtained from all trajectory points, that is, no distinction of conformer populations was made to support a direct comparison to experimental data.
[*] 50 ns MD averages.
[a] Data from Ref. 36.
[b] Data from Ref. 37, wherein the geometry from most populous conformer is present.
[c] Average dihedral angle values from all models included in PDB codes: 1CDR,[21] 1CDS,[21] 1GYA,[22] 1HD4,[23] 1BZB,[39] 1BYV,[39] 2FN2,[40] 1E88,[41] 1E9J,[23] 2JXA,[42] and 2JX9.[42]
[d] Average dihedral angle values from all models included in PDB codes 1CDR and 1CDS (10), 1URK (15), 1GYA (18), 1FF7 and 2HGO (20) and 1HD4 (26).
[e] Using the model 1 from PDB file 1CDR as a starting structure.
[f] Using the model 8 from PDB file 1CDR as a starting structure.
[g] Using the model 1 from PDB file 1CDS as a starting structure.
[h] Using the model 1 from PDB file 1GYA as a starting structure.
[i] Using the model 12 from PDB file 1GYA as a starting structure.
[j] Using the model 1 from PDB file 1HD4 as a starting structure.
[k] Using the model 11 from PDB file 1HD4 as a starting structure.
[l] Using the model 1 from PDB file 1FF7 as a starting structure. Values of 111 ± 28° were also populated in the simulation.

files) or around 70° (in 2HGO PDB file). Such a pattern is also in good agreement with the distribution of the dihedral angles around the Fuc-(1→O)-Ser linkage of the EGF-like domain, where two main peaks are indeed observed around the $\phi_{O\text{-GL}}$ angle (Fig. 2B). Again, good reliability is observed for O-linked glycoproteins, with 97% of the experimental geometries within the distribution curve for $\phi_{O\text{-GL}}$ and 69% for $\psi_{O\text{-GL}}$ (in a total of 45 geometries for each dihedral angle).[24,43,44]

3.3. Dynamics of glycosidic linkages that compose the glycan part of glycoproteins

Similar to the conformational behavior observed for the linkage between monosaccharides and amino acids, a given disaccharide showed mostly equivalent conformational profiles, with a minor influence of the protein scaffold on glycan conformational preference (Fig. 3 and Supplementary data). As observed, most of the experimental conformations, defined by NMR data, are within the distribution curves obtained from the MD simulations, indicating an adequate reproduction of the glycan conformational ensemble. In addition, considering that NMR methods provide average three-dimensional models, giving rise to ensemble-average properties instead of single solution conformations,[6] it is feasible to observe NMR geometries between main conformational states/peaks. Such characteristics also correlate to the observation of some geometries not included in the main regions of MD distribution curves, as it may not represent distinct conformers but may represent average ones.

Although the correct reproduction of the oligosaccharide geometry is important, the flexibility associated with disaccharides exposed to different environments must also be taken into account. In this context, it is proposed that the number of possible conformations for the dihedral angles of a given disaccharide, when comprising a glycan moiety of a glycoprotein, is considerably reduced at its core compared to the distal ends.[5] This appears to be caused by the conformational restriction promoted by the surrounding carbohydrate residues, although its geometry is similar to the one present in the free carbohydrate.[2] In accordance with such a

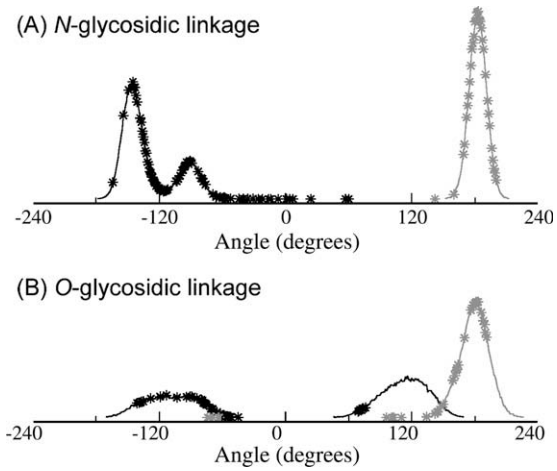


Figure 2. Distribution of the ϕ (black) and ψ (gray) dihedral angles associated with N- and O-glycosidic linkages in 50 ns MD simulations of the studied glycoproteins. Experimental geometries, as obtained from NMR data, are indicated by asterisks (*).

model, we were able to observe differences in the conformational behavior of the same disaccharide, β-GlcNAc-(1→4)-GlcNAc, in different molecular contexts (Figs. 1 and 4). In linear chains, with little or no steric hindrance, as seen in CD59 (PDB code 1CDR, Fig. 4A), a more flexible pattern is observed. In contrast, when such disaccharide is the core of a more complex glycan, as seen hsCD2$_{105}$ and α-hCG (PDB codes 1GYA and 1HD4, respectively, Fig. 4B), additional rigidity occurs, although the main peaks are highly similar, that is, a related set of the main conformational states. Such profiles seem to be influenced by the solvent accessibility around the disaccharides, as a higher solvent accessible surface (SAS) is observed in more flexible glycosidic linkages (Supplementary data). A similar influence of solvent exposure is observed for α-Man-(1→3)-Man depending on its localization at the glycan structure, that is, a given disaccharide will have higher flexibility when more exposed to solvent (Supplementary data). On the other hand, the presence of a linked carbohydrate residue alone does not necessarily influence the surrounding residues accessibility to the solvent. For example, the conformational pattern of the β-GlcNAc-(1→4)-GlcNAc linkage in the two CD59 PDB structures (1CDR and 1CDS) appears to not be influenced by a fucose six-linked to the GlcNAc residue (Supplementary data). This is probably due to the high flexibility of the (1→6)-linkage, which has less steric hindrance with the remaining glycan residues and, consequently, less influence on the solvation of the surrounding glycan.

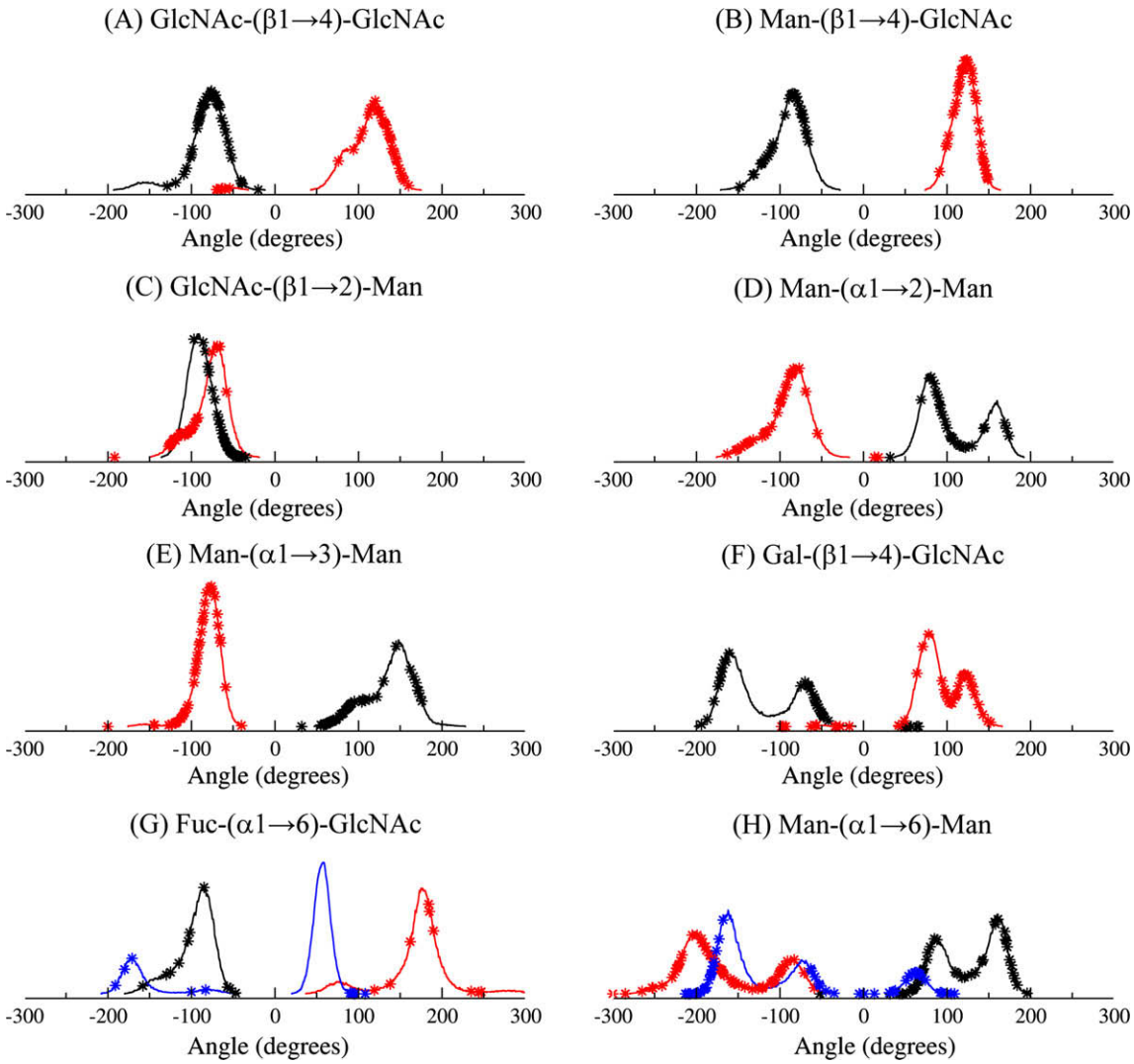


Figure 3. Distribution of the ϕ (black), ψ (red) and ω (blue) dihedral angles associated with specific disaccharides that compose the carbohydrate moieties of the studied glycoproteins in 50 ns MD simulations. Each curve is representative for the conformational behavior of a given disaccharide through the simulated glycoproteins (see Supplementary data). Asterisks (*) in the distribution curves indicate the experimental geometries, as obtained from NMR data, in a total of 836 values.

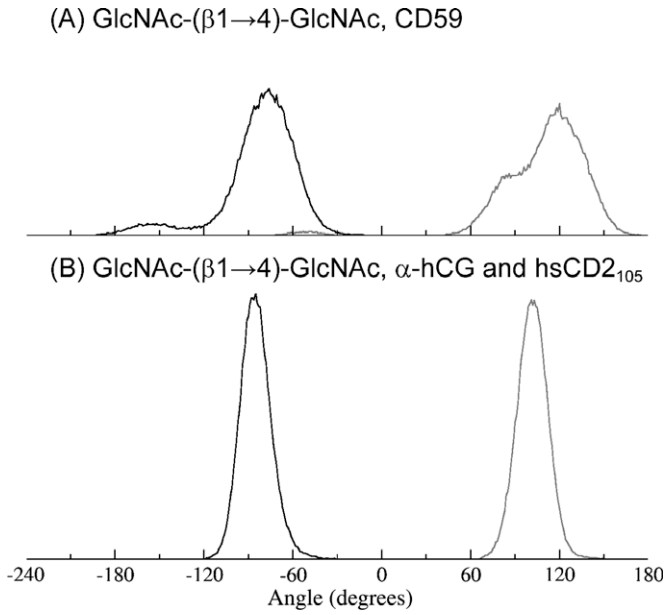


Figure 4. Distribution of the ϕ (black) and ψ (gray) dihedral angles associated with the β-GlcNAc-(1→4)-GlcNAc disaccharide behavior during 50 ns MD on different glycoproteins: (A) Attached to CD59 (PDB ID 1CDR and 1CDS); (B) Attached to both $hsCD2_{105}$ (PDB ID 1GYA) and α-hCG (PDB ID 1HD4).

Among the studies upon the conformational preferences of specific glycosidic linkages, special attention should be given to the description of the ω angle preferred orientations.[45] This bond is reported to adopt three main conformational states, around 180°, 60° and −60°, which is in good agreement with the conformational behavior observed from MD simulations. In contrast, it can be observed that NMR-measured dihedrals with regard to the ω angle of both α-Fuc-(1→6)-GlcNAc and α-Man-(1→6)-Man disaccharides adopt distinct geometries with large amplitudes (Fig. 3). These may indeed represent average values from two or three conformational states, which are observed to occur in solution for such disaccharides through MD.

3.4. Effect of glycosylation upon the protein moiety of glycoproteins

Although the biological functions of glycosylation are still not completely understood,[2,46] it is known that the glycan moieties of glycoproteins are fundamental to many biological processes, and are also involved in protein folding and/or stabilization.[1,2] In previous studies, MD simulations were employed to explore the role of glycosylation on peptides and proteins, such as the C-terminal region of the human prion glycoprotein,[47] the MUC1 peptide,[48] and a glycopeptide derived from a hemaglutinin protein fragment,[49] employing the AMBER force field,[7] and the MHC class I glycoprotein,[50] which used CVFF.[9] In this context, to evaluate simultaneously the glycan conformational description in MD, we have also performed an analysis of the influence of glycosylation on the dynamics and conformation of the protein moiety (Figs. 5 and 6), and these results will be discussed below. Additionally, the non-glycosylated molecules were also simulated with the OPLS-AA force field, with a secondary structure pattern equivalent to that observed with GROMOS96 (Supplementary data), which support the reliability of the obtained results.

3.4.1. EGF-like

According to experimental studies,[24] fucosylation does not significantly affect the structure of the EGF-like motif. In agreement with such an observation, no major differences in the secondary structure pattern between non-fucosylated and fucosylated proteins could be observed in MD simulations, as for RMSD (Fig. 5E) and radius of gyration (Fig. 5I). A high RMSD value for both forms of the EGF-like motif was observed, which appears to be due to the high loop content of this 46 amino acids protein, mainly its C- and N-terminal regions. On the other hand, a difference in flexibility between glycosylated and non-glycosylated proteins is observed around residues 56–59 (Fig. 6A and B), suggesting that the recognition of EGF-like domains by the *O*-fucosyltransferase may be associated with an entropic driven process, as this region lies within a putative consensus sequence for fucosylation in several EGF-like domains.[24]

3.4.2. CD59

The soluble form of CD59 had been studied by NMR methods, revealing a monomeric protein of 77 residues in a non-glycosylated form or with a disaccharide or trisaccharide N-glycan attached to Asn18.[21] Such data indicated that the global conformation of the protein backbone is essentially unchanged upon inclusion of oligosaccharides in the structure calculations. In accordance to these observations, MD simulations of non-glycosylated and glycosylated forms of CD59 showed no major differences between both forms of the protein for RMSD (Fig. 5F) and radius of gyration (Fig. 5J). Additionally, the NMR data pointed to small changes in the mean dihedrals and angular parameters of the protein component upon inclusion of carbohydrate groups, mainly around the side chain of Asn18.[21] Accordingly, the RMSF analyses of the simulated proteins support a decreased flexibility in the region of the N-glycan attachment, around Asn18 (Fig. 6C and D).

3.4.3. $hsCD2_{105}$

Regarding $hsCD2_{105}$, no major differences could be observed between non-glycosylated and glycosylated proteins based on RMSD and radius of gyration (Fig. 5G and K). However, an increased RMSD occurs in both forms of the protein, probably due to high loop content (mainly C- and N-terminal regions, Fig. 5C and G) and to the formation of several β-strands during MD simulations, known to exist in solution.[22,28,51] However, according to PROCHECK analysis, they are present in lower amounts in the NMR structures of both glycosylated and non-glycosylated forms (Fig. 5N).

Additionally, the NMR data suggest that the N-linked glycan at Asn65 is required for adhesion functions, because it stabilizes protein folding by counterbalancing an unfavorable clustering of five positive charges centered around Lys61.[22] Such an effect appears to be observed in MD simulations, because the region around these residues presents decreased flexibility upon glycosylation, as well as around regions comprising residues 22–32, 49–53, 61–67, and the N-terminal (Fig. 6E and F) domain.

Regarding CD2 function, transmembrane variants with mutations in the consensus N-glycosylation sequence Asn65-Gly66-Thr67,[52,53] which prohibit the attachment of the high-mannose N-glycan at Asn65, could be normally expressed on cell surfaces. However, they have shown neither antibody- nor ligand-binding activity, suggesting that the N-linked glycan on $hsCD2_{105}$ plays an important role in the function of this protein. In addition, the NMR studies suggested that protein–carbohydrate interactions may be involved in mediating the CD2 interaction with CD58 which should involve, mainly, the amino acids 32–34, 43–51 and 86–91 of $hsCD2_{105}$.[52,54] Accordingly, the residues around these regions concentrate the main differences in protein flexibility upon glycosylation (Fig. 6E and F), while participating in a series of β-sheets folded in MD apparently promoted by the glycan presence (Fig. 5N), suggesting a possible role of β-strand structures on CD2–CD58 binding.

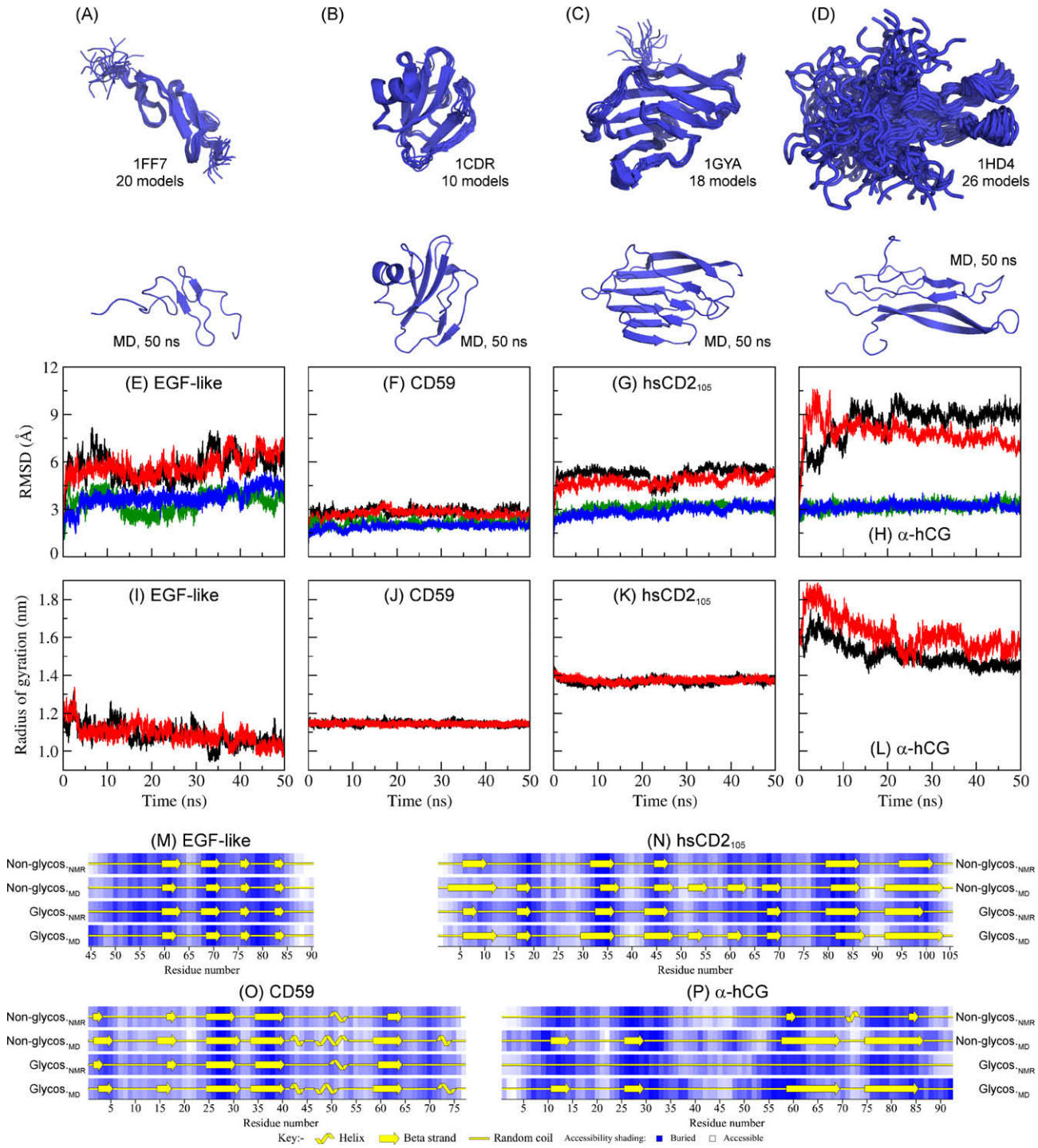


Figure 5. Comparative analysis between the studied glycoproteins and their non-glycosylated counterparts. The superimposed NMR models and the final MD conformation of EGF-like (A), CD59 (B), hsCD2$_{105}$ (C) and α-hCG (D) are presented together with root mean square deviation (RMSD) (E, F, G and H, respectively), in relation to the starting NMR model, for its respective non-glycosylated (black) and glycosylated (red) forms considering all atoms and the radius of gyration (I, J, K and L, respectively). The RMSD profile associated only with the atoms included in secondary structure elements of each molecule are shown for the non-glycosylated (green) and glycosylated (blue) forms. A comparative PROCHECK analysis of secondary structure content, obtained from NMR and final MD conformations of non-glycosylated and glycosylated proteins, is also shown (see Section 2 for further details).

3.4.4. α-hCG

The α-hCG structure has been determined through NMR methods in its glycosylated and non-glycosylated forms and deposited under PDB codes 1HD4[23] and 1DZ7,[27] respectively. Due to its high loop content, increased RMSD values are observed during MD simulations for both forms of the protein (Fig. 5H). This may also be related to the formation of a series of β-sheets in the simulations (Fig. 5P), which is also observed in OPLS/AA simulations of the non-glycosylated protein (Supplementary data). Additionally, some important differences could be observed between non-glycosylated and glycosylated forms of α-hCG based on RMSD (Fig. 5H) and radius of gyration (Fig. 5L). These differences seem to be caused by different loop flexibilities during the MD simulations (Fig. 5H), but especially by the large disordered loop between

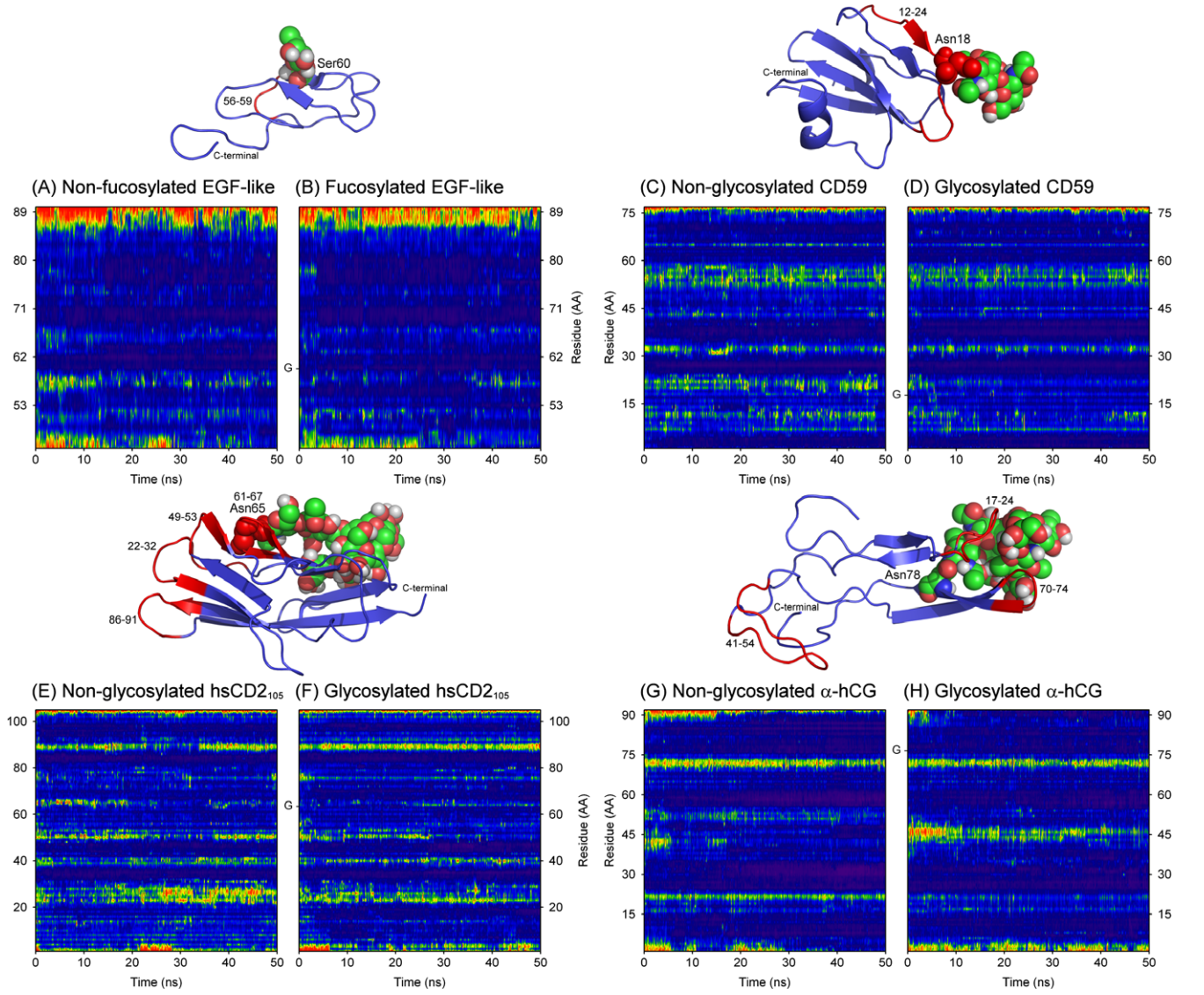


Figure 6. Root mean square fluctuation (RMSF) for the studied proteins, as a function of both residue number and time: EGF-like, in its non-fucosylated (A) and fucosylated (B) forms (from 0.04 to 0.24 nm); CD59, in its non-glycosylated (C) and glycosylated (D) forms (from 0.03 to 0.18 nm); $hsCD2_{105}$, in its non-glycosylated (E) and glycosylated (F) forms (from 0.04 to 0.16 nm); and the α-hCG, in its non-glycosylated (G) and glycosylated (H) forms (from 0.05 to 0.35 nm).

residues 33–58 (Fig. 5P). This loop is known to exist in free α-hCG, based on NMR studies,[55] and seems to adopt increased flexibility in the glycosylated protein (Fig. 6G and H). Such behavior is reproducible in an additional simulation, taking a different NMR model as starting structure under the same PDB code. As a general feature, glycosylation of the hCG α-subunit at Asn78 seems to play an important role in the stabilization of the protein[23] by exerting a protective function through shielding the protein surface from the environment,[55] including residues around 23–26 and 68–70. Therefore, from the MD simulations, we could observe a stabilization of such regions, especially 23–26, as observed in RMSF analysis (Fig. 6G and H).

3.5. Conformation of glycoproteins and their assessment through MD simulations

In recent years, glycobiology has become a critical facet of post-genomic science[56] as many proteins are post-translationally modified by glycosylation or co-translationally added by N-glycans in the endoplasmic reticulum.[46] Glycoslation may alter and regulate the biological activities of such macromolecules. In this context, several structural and functional studies on glycans alone, or on glycopeptides or glycoproteins have been carried out, using both experimental and theoretical methods. Concerning the use of the latter techniques, an assortment of approaches has been employed to model glycan structures when its three-dimensional structure is absent or is considered inappropriate. These include energy minimization of an entire oligosaccharide,[47] submitting the glycan to a simulated annealing protocol,[57] or constructing the carbohydrate moiety of glycoproteins from the disaccharide level,[50] that is, minimized structures of isolated disaccharides used to assemble an oligosaccharide tree. Additionally, several force field parameters have been employed to describe and simulate carbohydrate moieties of glycoproteins.[7–9] However, data regarding to the geometry of glycosidic linkages of such biomacromolecules are not always compared to experimental data[58] and, at times, are not analyzed during the simulated time,[47] that is, a proper populational ensemble characterization.

In contrast to most nucleic acids and proteins, which are linear and have a unique type of linkage, carbohydrates can be branched, linked through one of two anomeric configurations,[4] and there are more than 100 possible monomers (whereas ~10 in glycan composition),

which may be connected by different atoms, increasing its structural diversity.[59] All these properties, together with environmental factors, such as hydration and target proteins, determine the conformational ensemble of oligosaccharides. They are very flexible molecules, as their minimum-composing units, disaccharides, can adopt a dynamic equilibrium between two or more distinct conformations. All these properties are associated to the many functional roles of carbohydrates, making them one of the most challenging classes of molecules for conformational analysis.[6] Indeed, they are considered to have several orders of magnitude higher potential information content than any other biological macromolecule.[59] In this context, standard experimental techniques face several challenges when employed to obtain atomic-level structural information about carbohydrates. For example, X-ray crystallography may be difficult on these highly flexible systems and NMR spectroscopy mainly provides time-averaged conformational data.[6] Therefore, molecular modeling techniques, such as MD simulations, emerge as a promising tool for structural and conformational representation of carbohydrates, as they can describe, with a high level of accuracy in both spatial and temporal components, their geometry, flexibility and interaction with target proteins.[10–13,50,57,60,61]

Several force field parameters have been modified and/or designed to model and simulate carbohydrates, such as AMBER'93,[7] AMBER'06,[62] CVFF,[9] and CHARMm.[8] In this context, a previous paper pointed out that different approaches may be better for simulating different aspects of saccharides.[63] For example, the AMBER'93 force field seemed to be better than CHARMm for describing glycan dynamics and interactions with proteins, but less accurate for describing glycan hydration.[4] This may indicate that there is not a definitive force field for carbohydrates MD simulations, whereas further improvements were achieved with AMBER'06. On the other hand, this is the first systematic work assessing glycoprotein dynamics employing GROMOS96 43a1 force field parameters, together with the GROMACS simulation suite, which was shown to adequately represent the carbohydrate (and protein) moieties, as their NMR-derived conformations were adequately reproduced in a series of 50 ns MD simulations. As well, the simulation time scale achieved in this work totalized 0.4 μs for glycoproteins MD simulations and 0.2 μs for their non-glycosylated counterparts.

Although the modeling and adequate reproduction of glycans conformation are crucial for the adequate description of glycoproteins function, the analysis of N- and O-glycosidic linkages is essential. This may be because this relationship determines the orientation of the whole oligosaccharide relative to the protein and, in turn, influences the functional role of this class of biomacromolecules, such as the exposure of the glycan chain on the cell surface.[38] Among the MD studies on glycoproteins, only a few have explored the geometry of such linkages, including the α-GalNAc-(1→O)-Ser motif in solution,[64] and an antifreeze glycopeptide,[58] both consisting of O-glycosidic linkages with a carbohydrate moiety bound up to a disaccharide. The current work analyzed and properly described the conformations of an O-glycosidic linkage in a fucosylated glycoprotein, and the conformation of N-glycosidic linkages in four different glycoproteins. Among the parameters studied were its geometry and flexibility, indicating that the employed methodology is adequate for the reproduction of glycoprotein conformational ensembles in solution.

4. Conclusions

It is well known that glycoproteins play important roles in many biological events and that their carbohydrate moieties may be directly involved in such functions, including protein stabilization, folding, and interaction with target molecules. However, experimental methods present limitations when applied to carbohydrates. For example, X-ray crystallography is difficult on these highly flexible systems, and NMR tends to provide time-averaged conformations. Therefore, MD simulations emerge as a promising tool for complementing both NMR and X-ray experimental data on glycans and glycoproteins, because it provides a reasonable sampling of the conformational behavior of such molecules in solution.

The employment of molecular modeling techniques may be expected to contribute significantly to the understanding of the many events related to this class of biomacromolecules. Indeed, some force field parameters have been designed for modeling and simulating glycans, being shown to adequately represent experimental geometries. In this context, the GROMOS96 43a1 force field, enhanced by Löwdin HF/6-31G**-derived atomic charges for the glycan moiety, adequately reproduced the conformational ensemble of a series of glycoproteins, as well as the effect of glycosylation over the studied proteins, when compared to NMR data.

Finally, the employed methodology appeared to reproduce adequately the conformational ensemble of a series of glycoproteins containing carbohydrate moieties that are involved in different biological roles. Consequently, upcoming studies using MD simulations upon glycoproteins may be expected to provide a significant contribution to the investigation of their involvement in biological events.

Acknowledgments

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq #420015/2005-1 and #472174/2007-0), MCT, and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), MEC, Brasília, DF, Brazil.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.carres.2008.12.025.

References

1. Varki, A. *Glycobiology* **1993**, *3*, 97–130.
2. Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683–720.
3. Bohne-Lang, A.; von der Lieth, C. W. *Pac. Symp. Biocomput.* **2002**, *7*, 285–296.
4. Petrescu, A. J.; Wormald, M. R.; Dwek, R. A. *Curr. Opin. Struct. Biol.* **2006**, *16*, 600–607.
5. Petrescu, A. J.; Petrescu, S. M.; Dwek, R. A.; Wormald, M. R. *Glycobiology* **1999**, *9*, 343–352.
6. Woods, R. J. *Glycoconjugate J.* **1998**, *15*, 209–216.
7. Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832–3846.
8. Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carbohydr. Res.* **1988**, *180*, 207–221.
9. Hwang, M. J.; Ni, X.; Waldman, M.; Ewig, C. S.; Hagler, A. T. *Biopolymers* **1998**, *45*, 435–468.
10. Verli, H.; Guimarães, J. A. *Carbohydr. Res.* **2004**, *339*, 281–290.
11. Verli, H.; Guimarães, J. A. *J. Mol. Graphics Modell.* **2005**, *24*, 203–212.
12. Becker, C. F.; Guimarães, J. A.; Verli, H. *Carbohydr. Res.* **2005**, *340*, 1499–1507.
13. Pol-Fachin, L.; Verli, H. *Carbohydr. Res.* **2008**, *343*, 1435–1445.
14. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Huenenberger, P. H.; Krueger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Vdf Hochschulverlag, AG Zurich: Switzerland, 1996.
15. IUPAC-IUB Commission on Biochemical Nomenclature, *Pure Appl. Chem.* **1983**, *55*, 1269–1272.
16. Schaftenaar, G. MOLDEN. CAOS/CAMM Center, University of Nijmegen, Toernooiveld 1, 6525 ED NIJMEGEN, The Netherlands, 1997.
17. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
18. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
19. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

20. Schuettelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr., Sect. D* **2004**, *60*, 1355–1363.
21. Fletcher, C. M.; Harrison, R. A.; Lachmann, P. J.; Neuhaus, D. *Structure* **1994**, *2*, 185–199.
22. Wyss, D. F.; Choi, J. S.; Li, J.; Knoppers, M. H.; Willis, K. J.; Arulanandam, A. R. N.; Smolyar, A.; Reinherz, E. L.; Wagner, G. *Science* **1995**, *269*, 1273–1278.
23. Erbel, P. J. A.; Karimi-Nejad, Y.; van Kuik, J. A.; Boelens, R.; Kamerling, J. P.; Vliegenthart, J. F. G. *Biochemistry* **2000**, *39*, 6012–6021.
24. Kao, Y. H.; Lee, G. F.; Wang, Y.; Starovasnik, M. A.; Kelley, R. F.; Spellman, M. W.; Lerner, L. *Biochemistry* **1999**, *38*, 7097–7110.
25. Lins, R. D.; Hünenberger, P. H. *J. Comput. Chem.* **2005**, *26*, 1400–1412.
26. Kräutler, V.; Müller, M.; Hünenberger, P. H. *Carbohydr. Res.* **2007**, *342*, 2097–2124.
27. Erbel, P. J. A.; Karimi-Nejad, Y.; De Beer, T.; Boelens, R.; Kamerling, J. P.; Vliegenthart, J. F. G. *Eur. J. Biochem.* **1999**, *260*, 490–498.
28. Withka, J. M.; Wyss, D. F.; Wagner, G.; Arulanandam, A. R.; Reinherz, E. L.; Recny, M. A. *Structure* **1993**, *1*, 69–81.
29. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
30. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
31. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
32. Berendsen, H. J. C.; Postma, J. P. M.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
33. Jain, S.; Sundaralingam, M. *J. Biol. Chem.* **1989**, *264*, 12780–12784.
34. Eyal, E.; Gerzon, S.; Potapov, V.; Edelman, M.; Sobolev, V. *J. Mol. Biol.* **2005**, *351*, 431–442.
35. Andrec, M.; Snyder, D. A.; Zhou, Z.; Young, J.; Montelione, G. T.; Levy, R. M. *Proteins* **2007**, *69*, 449–465.
36. Imberty, A.; Pérez, S. *Protein Eng.* **1995**, *8*, 699–709.
37. Petrescu, A. J.; Milac, A. L.; Petrescu, S. M.; Dwek, R. A.; Wormald, M. R. *Glycobiology* **2004**, *14*, 103–114.
38. Ali, M. M.; Aich, U.; Varghese, B.; Pérez, S.; Imberty, A.; Loganathan, D. *J. Am. Chem. Soc.* **2008**, *130*, 8317–8325.
39. Hashimoto, Y.; Toma, K.; Nishikido, J.; Yamamoto, K.; Haneda, K.; Inazu, T.; Valentine, K. G.; Opella, S. J. *Biochemistry* **1999**, *38*, 8377–8384.
40. Sticht, H.; Pickford, A. R.; Potts, J. R.; Campbell, I. D. *J. Mol. Biol.* **1998**, *276*, 177–187.
41. Pickford, A. R.; Smith, S. P.; Staunton, D.; Boyd, J.; Campbell, I. D. *EMBO J.* **2001**, *20*, 1519–1529.
42. Vakonakis, I.; Langenhan, T.; Prömel, S.; Russ, A.; Campbell, I. D. *Structure* **2008**, *16*, 944–953.
43. Hansen, A. P.; Petros, A. M.; Meadows, R. P.; Nettesheim, D. G.; Mazar, A. P.; Olejniczak, E. T.; Xu, R. X.; Pederson, T. M.; Henkin, J.; Fesik, S. W. *Biochemistry* **1994**, *33*, 4847–4864.
44. Barthe, P.; Pujade-Renaud, V.; Breton, F.; Gargani, D.; Thai, R.; Roumestand, C.; de Lamotte, F. *J. Mol. Biol.* **2007**, *367*, 89–101.
45. Shefter, E.; Trueblood, K. N. *Acta Crystallogr.* **1965**, *18*, 1067–1077.
46. Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364–2369.
47. Zuegg, J.; Gready, J. E. *Glycobiology* **2000**, *10*, 959–974.
48. Rubinstein, A.; Kinarsky, L.; Sherman, S. *Int. J. Mol. Sci.* **2004**, *5*, 119–128.
49. Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali, B. *J. Am. Chem. Soc.* **2004**, *126*, 8421–8425.
50. Mandal, T. K.; Mukhopadhyay, C. *Biopolymers* **2001**, *59*, 11–23.
51. Wyss, D. F.; Withka, J. M.; Knoppers, M. H.; Sterne, K. A.; Recny, M. A.; Wagner, G. *Biochemistry* **1993**, *32*, 10995–11006.
52. Arulanandam, A. R. N.; Withka, J. M.; Wyss, D. F.; Wagner, G.; Kister, A.; Pallai, P.; Recny, M. A.; Reinherz, E. L. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11613–11617.
53. Recny, M. A.; Luther, M. A.; Knoppers, M. H.; Neidhardt, E. A.; Klandekar, S. S.; Concino, M. F.; Schimke, P. A.; Francis, M. A.; Moebiusll, U.; Reinhold, B. B.; Reinhold, V. N.; Reinherz, E. L. *J. Biol. Chem.* **1992**, *267*, 22428–22434.
54. Peterson, A.; Seed, B. *Nature* **1987**, *329*, 842–846.
55. de Beer, T.; van Zuylen, C. W. E. M.; Leeflang, B. R.; Härd, K.; Boelens, R.; Kaptein, R.; Kamerling, J. P.; Vliegenthart, J. F. G. *Eur. J. Biochem.* **1996**, *241*, 229–242.
56. Turnbull, J. E.; Field, R. A. *Nat. Chem. Biol.* **2007**, *3*, 74–77.
57. Mukhopadhyay, C. *Biopolymers* **1998**, *45*, 177–190.
58. Nguyen, D. H.; Colvin, M. E.; Yeh, Y.; Feeney, R. E.; Fink, W. H. *Biophys. J.* **2002**, *82*, 2892–2905.
59. Imberty, A.; Pérez, S. *Chem. Rev.* **2000**, *100*, 4567–4588.
60. Becker, C. F.; Guimarães, J. A.; Mourão, P. A. S.; Verli, H. *J. Mol. Graphics Modell.* **2007**, *26*, 391–399.
61. Naidoo, K. J.; Denysyk, D.; Brady, J. W. *Protein Eng.* **1997**, *10*, 1249–1261.
62. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.
63. Corzana, F.; Motawia, M. S.; Du Penhoat, C. H.; Pérez, S.; Tschampel, S. M.; Woods, R. J.; Engelsen, S. B. *J. Comput. Chem.* **2003**, *25*, 573–586.
64. Corzana, F.; Busto, J. H.; Jiménez-Osés, G.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A. *J. Am. Chem. Soc.* **2006**, *128*, 14640–14648.