SIRAH late harvest:

coarse-grained models for protein glycosylation

Pablo G. Garay[1], Matias R. Machado[1], Hugo Verli[2], and Sergio Pantano[1]

[1]Biomolecular Simulations Group, Institut Pasteur de Montevideo, Mataojo 2020, CP 11400 Montevideo, Uruguay.

[2]Programa de Pos-Graduacao em Biologia Celular e Molecular (PPGBCM), Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Av. Bento Goncalves, 9500, Porto Alegre - Brazil – 91509-900.

Abstract

Glycans constitute one of the most complex families of biological molecules. Despite their crucial role in a plethora of biological processes, they remain uncharacterized because of their high complexity. Their intrinsic flexibility, along with the vast variability associated with the many combination possibilities and relatively loose structural character, has hampered their experimental determination. Although theoretical methods have proven a valid alternative to the study of glycans, the large size associated with polysaccharides, proteoglycans, and glycolipids poses significant challenges to a fully atomistic description of biologically relevant glycoconjugates. On the other hand, the exquisite dependence on Hydrogen bonds to determine glycans' structure makes the development of simplified or coarse-grained (CG) representations extremely challenging. This is particularly the case when glycan representations are expected to be compatible with CG models of, for instance, proteins.

We introduce a CG representation able to represent a wide variety of polysaccharides and most glycosylation motifs in proteins, which is fully compatible with the CG SIRAH force field. Examples of application to N-glycosylated proteins, including antibody recognition and Calcium-mediated glycan-protein interactions, highlight the versatility of the enlarged set of CG molecules provided by SIRAH.

Introduction

Glycans are one of the most structurally and functionally diverse molecules in nature. Their structures range from simple mono or disaccharides to highly complex structures composed by long chains, and branched structures that can be linked to proteins and lipids, forming glycoproteins, proteoglycans and glycolipids. This structural diversity is reflected in the wide range of functions they cover, among which we can find, energy storage, ultrastructural stabilization in cell walls of bacteria, plants or fungi, cell recognition and adhesion, protein folding, modulation of immune systems, and cancer, among many others[1]. Indeed, a significant amount of proteins are predicted to be glycosylated to some extent[2,3], illustrating the profound relevance of this family of biomolecules. However, the absence of glycosylation in *E. coli*[4],the most common expression system for recombinant proteins, along with their large flexibility and the difficulty to discern spectroscopic signals from highly hydrated sugar rings has hampered our knowledge about the structure and dynamics of glycans in general, and glycosylated proteins, in particular. Therefore, alternative tools are necessary to study the dynamics and interaction of glycans and glycoconjugates to obtain a better understanding of their functionalities. One of the main tools to complement experimental studies is the use of molecular dynamics (MD) simulations[5–9]. Nevertheless, the computational cost of simulating large glycosylated proteins or polysaccharides chains with large polymerization degrees poses significant challenges to fully atomistic MD simulations. A workaround to this limitation is the development of coarse-grained (CG) models, which reduce the number of atoms and degrees of freedom in the molecular system, making them less computationally demanding but still preserving salient features of the molecules of interest.

Despite a large number of successful examples of CG models developed for almost any biological family of molecules[10], glycans represent undoubtedly one of the most challenging cases. Among other difficulties, the conformational aspects of glycans heavily rely on the accurate description of inter and intra molecular Hydrogen bonds. However, hydroxyl moieties are most frequently lumped into effective interaction points or “beads”. Moreover, the large chemical variety of building blocks, link possibilities, and ramifications make the number of topologies virtually infinite. Indeed, very few examples of CG glycans are available in the literature. Among them, the popular MARTINI force field provided a CG representation for carbohydrates[11], which has been extended to glycolipids[12], and glycosylated proteins[13]. In this case, the parameterization follows the general “4-to-1 mapping rule”, i.e., four heavy atoms are mapped to one single bead, while bonded and non bonded interaction parameters are chosen from the standard MARTINI force field. These CG glycans have been improved to reduce the stickiness observed between MARTINI beads[14]. Following a completely different

strategy Rusu et al.[15] have developed a methodology to develop CG parameters of glycans using a bottom-up approach named PITOMBA. This approach is based on experimental data and fully atomistic simulations using the force field GROMOS 53A6$_{GLYC}$[16]. Another CG force field with representation for linear carbohydrates is the Unified Coarse-Grained Model[17], this force field is based on the consideration that the repeated units of biopolymers are highly polar and their charge distributions can be represented as point multipoles. Carbohydrate parameters for this force field (SUGRES-1P) are still in development[18].

Our group have developed the SIRAH force field (***S****outhamerican* ***I****nitiative for a* ***R****apid and* ***A****ccurate* ***H****amiltonian*), which contains parameters for DNA[19], proteins, solvents and ions[20–22], lipids[23,24], post-transductional modifications (PTMs)[25], and metal ions[26]. In this work, we introduce a CG mapping scheme and parameterization generalizable for arbitrary pyranoses. Using this mapping scheme and the topologies provided here it is possible to perform CG simulations of a series of polysaccharides and virtually all N-glycosylation motifs present in proteins. The versatility of the CG mapping is exemplified by modeling proteins with different N-glycosylations, and glycans mediating protein interactions. We show that our parameterization can also handle Calcium mediated protein-glycan interactions, obviating the need for ad-hoc constraints.

These new moieties considerably expand the range of molecular systems amenable to be studied with our CG force field. Parameter files and topologies to run MD simulations in Amber are available upon request.

Methods

Data Retrieving for parameterization.

The general philosophy for mapping and parameter fitting in SIRAH consists in placing effective beads on atomic positions preserving interaction points deemed important to the correct reproduction of structural properties. Subsequently, statistical information is extracted from the Protein Data Bank (PDB) to derive interaction parameters. For glycans, the coarse-graining is performed keeping only the position of the hydroxyl moieties decorating the rings, and the glycosidic oxygens. This choice preserves the capability of mapping pyranoses with all possible links and ramifications.

To extract statistical structural information regarding abundance, equilibrium distances and angles between beads of each monosaccharide ring, we first looked for data available in the PDB. This data was collected through GLYCOSCIENCES.de[27] with the GlyTorsion tool[28], filter structures with resolution of 0.2 nm or better. We found around 20000 ring for pyranoses, while only 628 structures are

available for furanoses. Among all possible pyranose ring conformations, the most stable and abundant in the PDB is the chair $^4C_1$[29]. Therefore, we decided to restrict the CG modeling to pyranoses in $^4C_1$ conformation. Similarly, we focused on the most abundant D enantiomers, with the exception of α-L-Fucp present in the core complex N-glycosylation[29].

Naming Code and CG mapping.

A clear standard for glycan naming is not strictly followed in all the structures reported in the PDB. Although significant efforts are being devoted to remediate PDB structures (http://wwpdb-remediation.rutgers.edu/documentation/carbohydrate-remediation), ambiguity is still an issue of concern. Throughout this work, we defined a naming code based on the scheme used in GLYCAM[30]. This nomenclature offers the advantage of a non-redundant naming for monosaccharides with different glycosidic bonds. In this sense, it is possible to distinguish between different linkages (1-1, 1-2, 1-3, etc.), anomeric configuration (α or β) and D or L enantiomers.

To preserve the possibility of future multiscale developments[31] and maintain the compatibility with AMBER force fields, the CG monosaccharides in SIRAH are named inverting the order of the characters of the corresponding GLYCAM´s monosaccharide names (i.e., written backward). An example of the three letters code used in SIRAH, α-D-Manp with bond 1-4, is shown in Figure 1A. The first character encodes the anomeric configuration and the ring form, “A” or “B” stand for α or β for pyranose rings. The second character is the monosaccharide´s one-letter-code. Namely, “M” is the code for D-Manp (Figure 1A). Upper and lower case is used for the D and L enantiomers (for instance, “f” is used for L-Fucose). The last character indicates occupied linkage positions. Hence, “4” indicates a bond 1-4 from mannose. The terminal monosaccharides are indicated with “0” (zero).

The CG beads are placed in the position of the oxygens and hydrogens in fully atomistic representations (Figure 1 B). Beads corresponding to Oxygen atoms are named with “A” for [a]xial, “R” for the ecuato[r]ial positions and numbered according to their position in the ring. Oxygen beads in glycosidic linkages are named “L” in axia[l], and “k” in e[k]uatorial positions, respectively. Beads corresponding to Hydrogen atoms are named “SP” in all cases but in 1-6 bonds, where they are named “SX” (Figure 1B). In the case of position 6, a bead is placed on the carbon 6 and another in the position of the hydroxyl´s hydrogen (for instance, see AM6 in Figure 1B). The anomeric configuration of each residue is indicated by the use of lower case letters in the bead names connecting α bonds, while the upper case is used for β bonds (see AM4 (dot line) and BM4 (solid line) in Figure 1A with a2 and A2 beads for α and β bonds, respectively). It is worth noting that this convention applies only to the bead

names, and not the interaction parameters.

This CG mapping and naming schemes preserve the unique identity of each sugar ring and retain the chemical identity of all possible linkages, OH donors/acceptors, and ramifications. Worthy, this also makes possible the unambiguous backmapping to fully atomistic representation. This functionality has been added to SIRAH Tools[32]. Hence, with this new release of SIRAH it is possible to run a CG simulation of N-glycosylated proteins and backmap selected conformations with a naming scheme fully compatible with AMBER/GLYCAM force fields.

All glycan residues presented in this work, including ramifications and non-reducing ends, are shown in Figure 2.

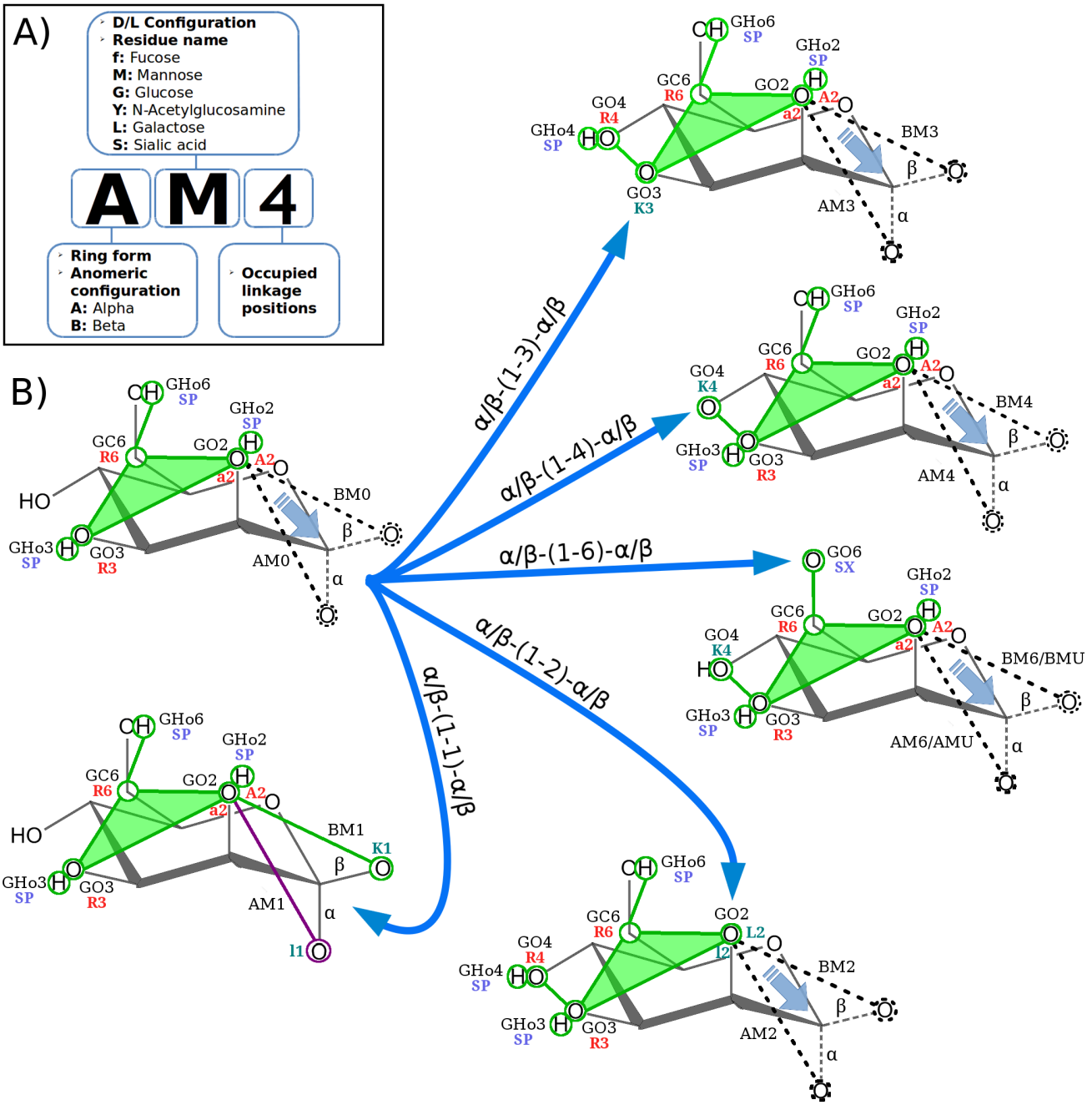

Figure 1. CG representation and naming code. A) Naming and CG scheme for sugar residues using α-D-Manp as example. All rings currently available are listed. **B)** Mapping scheme for all possible mannose linkages. Green lines identify the connectivity between CG beds. Labels near CG beads indicate the bead type and name within the SIRAH force field. Red, blue, and green characters denote the partial charges in each bead (green: 0, blue: positive, and red: negative).

Unlike proteins and nucleic acids, the connectivity between consecutive glycan rings is not immediate to guess because of all the possible linkages. Moreover, the connection between consecutive rings may be difficult to guess from mid-resolution X-ray structures. To the best of our knowledge, there are no

algorithms to automatize the monosaccharides' connection when systems are constructed for CG MD simulations in Amber. To facilitate building the correct connectivity, we list in Table 1 the bead names involved in possible connections between residues.

Residue	i	i+1	Bond
Af0	GO2	None	Non-reducing end
AM0		None	Non-reducing end
AM2		GO2	1-2
AMQ		GO3	1-3
		GO4	1-4
		GO6	1-6
AMV		GO3	1-3
		GO6	1-6
BM0		None	Non-reducing end
BM4		GO4	1-4
BMQ		GO3	1-3
		GO4	1-4
		GO6	1-6
BMV		GO3	1-3
		GO6	1-6
BY0	GNac	None	Non-reducing end
		BND	Asn glycosylation
BY4		GO4	1-4
		BND	Asn glycosylation
BYU		GO4	1-4
		GO6	1-6
		BND	Asn glycosylation

Table 1. Possible connecting beads between the monosaccharides in SIRAH. Note that this follows the PDB convention of numbering monosaccharides from the protein to the non-reducing end.

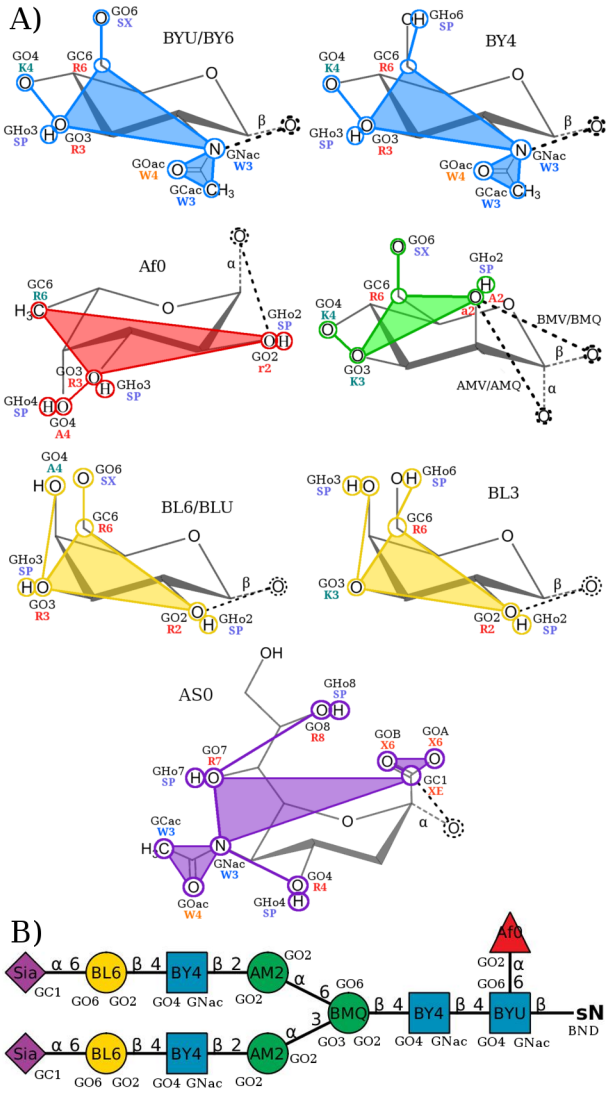

Figure 2. **A)** CG representation of all sugar rings provided in this work. **B)** Example of a complex glycosylation tree in SFNG[33] representation labeled with the SIRAH naming. The beads names involved in glycosidic bonds are noted below each glycan, except for the bonds 1-6, which are written above.

Interaction parameters

To preserve the compatibility with the existing molecules in the force field, we use parameters already developed in the SIRAH (see Supporting Information). To determine the equilibrium values used for bonded interactions, we collected statistical structural information from the PDB, selecting only structures with a resolution of 0.2 nm or better. For the distribution of glycosidic torsional phi (Φ) and psi (Ψ) angles, using the C+1 definition, we first calculated the median of the experimental distribution. Then, for bonds, angles, and torsional angles of each monosaccharide, we selected the value measured in the structure closer to the median. When experimental data is scarce within the selected resolution, as is the case of N-acetylneuraminic Acid (Neu5Ac), the parameters for the monosaccharides were

taken from the templates of GLYCAM[30] (Tables S1, S2 and S3). The inter-monosaccharide parameters are set to allow a wide conformational sampling in each type of glycosidic bond (Table S1, S2 and S3). For the non-bonded part of the potential, we use Lennard-Jones parameters and point charges from similar chemical functional groups existing in different amino acids (Table S4 and S5).

Preparing input files for CG simulations

The first step to run a simulation with SIRAH is to map an atomistic structure to the CG representation. This is performed using a script provided in the SIRAH package[32]. A correct mapping absolutely requires residues names to be written according to the GLYCAM nomenclature. It is important to emphasize that our CG model requires the position of the Hydrogen atoms in the atomistic structure. Furthermore, the Oxygen atom belonging to the glycosidic bond is always assigned to the second residue in each bond, considering the chain's direction from the non-reducing end to the protein. If the glycan is in solution, the chain's direction goes from the non-reducing to the reducing end.

Among other possibilities to generate suitable input files for CG simulations, we suggest the following:

1) If experimental structures are not available use the builder tools for Carbohydrates or Glycoproteins from the webserver GLYCAM (http://glycam.org/) or the tool Glycan Reader & Modeler from CHARMM-GUI[33] (http://www.charmm-gui.org/?doc=input/glycan). In both cases it is possible to produce PDB files with the GLYCAM nomenclature.
2) Starting from an experimental PDB file: use CHARMM-GUI server to convert the names of the monosaccharides to GLYCAM names.

System setup and MD simulations

The system setup always started from experimental structures. The studied systems were: i) the N-glycosylated Receptor Binding Domain (RBD) of SARS-CoV-2 (PDB id.: 6VSB). ii) N-glycosylation interacting with protein and Calcium (PDB id.: 1K9J). iii) N-glycosylation interacting with protein (PDB id: 4UT6). Atomic coordinates were taken from the PDB using the Glycan Reader & Modeler utility of CHARM-GUI. Missing residues were completed when necessary. Non-protein and non-glycan molecules were removed (e.g., water or molecules present in crystallization buffers).

All simulations were performed with the AMBER 18 suite[34]. Protonated structures were mapped to CG with SIRAH tools[32] and converted to AMBER's topologies by the LEAP utility. Each glycosidic bond was manually set according to Table 1. Molecules were centered in an octahedral box filled with pre-equilibrated CG water molecules named WT4. Monovalent salt (NaCl) in a 0.15 M concentration is added[35] by randomly replacing CG ions by WT4 molecules. The GPU code was used to run the

calculations[36]. Non-bonded interactions were treated with a 1.2 nm cutoff and Particle Mesh Ewald (PME) for long-range electrostatics[37,38]. The PME is calculated at every integration step, and the neighbor list is updated whenever any atom had moved more than one-half a non-bonded "skin" of 0.2 nm. A Fourier spacing is defined as a function of the box size and grid density. The PME is calculated at every integration step, and the neighbor list is updated whenever any atom had moved more than one-half a non-bonded "skin" of 0.2 nm. A Fourier spacing is defined as a function of the box size and grid density. The whole system is coupled to a Langevin[39] thermostat with a collision frequency of 50 ps^{-1} and to a Berendsen barostat[40], with a relaxation time of 1 ps.

The simulation protocol used is the following:

1) Solvent and side chains relaxation by 5000 steps of energy minimization imposing positional restraints of 2.4 kcal mol^{-1} $Å^{-2}$ on backbone beads corresponding to the nitrogen and carbonylic oxygen (named GN and GO, respectively), and for the glycan rings (GO2, GNac, GO3 and GC6), for the position 2, 3 and 6 in all rings but for sialic acid. In the latter case, restraints were applied to GC1, GNac and GO7 for the position 1, 5 and 7.

2) Full system relaxation by 5000 steps of unrestrained energy minimization.

3) Solvent equilibration by 5 ns of MD in the NVT ensemble at 300 K, imposing positional restraints of 2.4 kcal mol^{-1} $Å^{-2}$ on the whole protein and glycan.

4) Protein and glycan relaxation by 25 ns of MD in the NVT ensemble at 300 K, imposing positional restraints of 0.24 kcal mol^{-1} $Å^{-2}$ on same beads as step 1.

5) Same as step 4 with positional restrains of 0.12 kcal mol^{-1} $Å^{-2}$.

6) Production simulation in the NPT ensemble at 300 K and 1 bar.

Data analysis and visualizations

The data analysis and plots are made using python packages Ipython[41], Pandas[42], Numpy[43], Matplotlib[44] and Seaborn[45], the analysis of the trajectories is performed using SIRAH tools[32],and MDAnalysis[46,47]. The images of the systems are made with VMD[48] and DrawGlycan-SNFG[49]. RMSD in protein chains is calculated over Cα on structured region and, in glycans are calculated over the beads conforming the ring of the CG monosaccharides. The compaction degree of the glycosylation is calculated as the difference between the initial and average Radii of Gyration (RG) measured during the trajectories taken in percentage. The vdW radii corresponding to SIRAH 2.0 beads are used. Native contacts are calculated between glycans and proteins initially separated 0.8 nm or less. Native contacts are defined

on the experimental structure, and their conservation is measured as the retained percentage along the trajectory. The contacts' accuracy is defined as the ratio between conserved native contacts and the total occurring contacts. Buried Surfaces Areas (BSA), between glycans and proteins, are calculated based on the Surface (Surf) of each molecule. This is, BSA = (A_{surf} + B_{surf}) - C_{Surf}, where A_{surf} and B_{surf} are the surfaces of glycan and protein, respectively, and C_{Surf} is the surface of the complex.

Results and Discussion

In the following paragraphs we apply the developed SIRAH glycans on three different systems: i) the dynamics of an isolated protein domain with three different glycosylation motifs completely exposed to the solvent; ii) Calcium mediated protein-glycan interaction; and iii) antibody recognition of a glycoprotein.

i) N-glycosylated SARS-CoV-2 RBD

The Spike protein of SARS-CoV-2 coronavirus is a trimeric and highly glycosylated protein that protrudes from the membrane of the mature virion[50]. A Receptor Binding Domain (RBD), comprised between amino acids 327 to 532, eventually separates from the compact folding of the Spike, and specifically recognizes the N-terminal segment of the human Angiotensin Converting Enzyme 2 (ACE2)[51]. The RBD contains two glycosylation sites at Asn331 and Asn343[52], which we named for simplicity R1 and R2 respectively. The nature of the attached glycan trees in the intact virions is expected to be variable[53]. As an example of the glycan variety that can be represented by our CG representation, we sought to perform three comparative simulations of the RBD with different glycosylation motifs. Namely, Complex, High-mannose (Manp9) and core of complex glycosylations (core complex, i.e. without the capping galactose and sialic acid, Figure 3A, and S1A, left and right side, respectively).

Each glycosylation motif in the three simulations showed slightly different behaviors. For the sake of briefness we focus mainly on Complex glycosylation (Fig. 3). The data describing Manp9 and core complex is reported in Table 2 and Figure S1. In all three cases, the glycans visited a multiplicity of conformations (Figure 3B, and S1B) separated up to 0.4, 0.6 and 0.7 nm away from their initial positions (RMSD on Table 2), for core complex, Complex and Manp9 motifs, respectively. Such large RMSD values highlight the conformationally unbiased nature of our CG representation. Despite the considerable mobility of the glycans, the protein's flexibility resulted not sensible to the identity of the sugar rings (Figure 3C and S1C). Indeed, a growing number of experimentally reported RBD structures

in the PDB showed no significant differences in the presence or absence of glycosylation. In any of the ramifications studied, the RMSD of the structured part of the protein remained around 0.4 nm (Figure 3C). Such values are well in line with those obtained for a set of simulations of non-glycosylated and post-transductionally modified CG proteins in our force field[22,25].

The main differences observed regard the helix content in the simulation with the Manp9 ramification, with respect to the experimental value (Table 2). This is because the glycosylation in the Asn331 is over a helix that partially losses its structure during the simulation. Comparing the temporal evolution of protein's RMSD with the radius of gyration (RG) of the different ramifications showed no obvious dynamical correlation (compare Figure 3C and 3D). The compaction degree during the trajectories is higher for the Complex glycosylation with reductions of 25% (0.30 nm) and 27% (0.32 nm) for R1 and R2, respectively (Table 2). On the other hand, the core complex glycosylation showed intermediate values of 6.42% (0.05 nm) and 14.17% (0.11 nm), while the Manp9 trees show marginal reductions of 4.6% (0.04 nm) and 3.55% (0.03 nm). Certain compaction is expected because these systems are constructed without the protein that surrounded the RBD domain.

The two ramifications in the core complex glycosylation experienced a random encounter nearly after 1 μs and remained in contact for the rest of the simulation (Figure S1B, right). Nevertheless, the absolute compactions measured in all cases were below 0.72 nm, namely, the approximate size of a Glycine residue. This small variation suggested that despite significant flexibility, the global volume of the glycan trees was conserved.

The small condensation observed implies a glycosylation specific burial of the protein surface. Despite the lack of correlation between the protein and glycans dynamics, we observed clear differences in the protein surface covered by each ramification (Figure 3E). Here the Manp9 ramification, with 11 monosaccharides, has more BSA than the Complex ramification with 12 monosaccharides. That could be explained by the minor compaction degree of the Manp9. This highlights the potentiality of the CG approach to quickly explore different possible scenarios associated with the role of the glycans in shielding the protein surface from the immune system[54].

The raw data corresponding to the simulations' trajectories are part of an initiative for data sharing on SARS-CoV-2 proteins and are freely available from the Zenodo database[55].

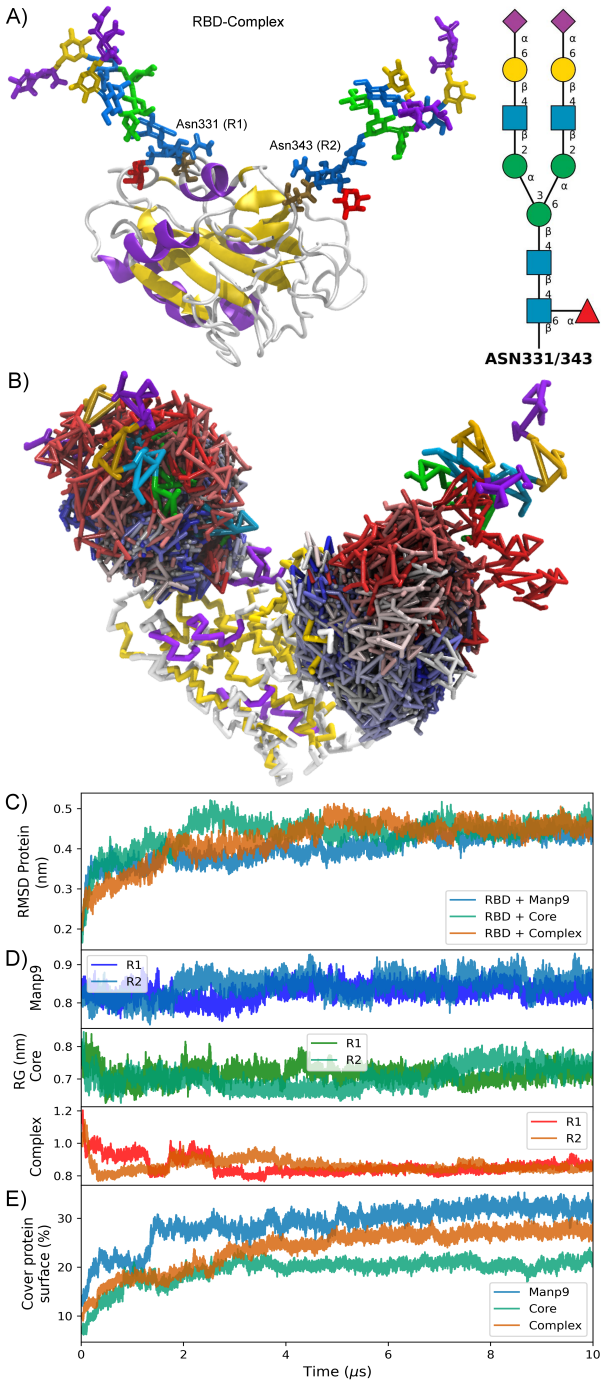

Figure 3. A) Atomistic representation of SARS-CoV-2 RBD with Complex N-glycosylation (left) and SNFG representation of the glycosylation at Asn331 and 334 (right). Protein is represented as cartoon and colored according the secondary structure, helix in purple, extended beta in yellow and coil in white. The glycosylations are represented as licorice and colored according to SNFG, R1 and R2 indicate the glycosylation on Asn331 and Asn343, for clarity the lateral chains were omitted. **B)** CG representation of the system shown in panel A. Glycans are colored according the SNFG for the first frame and by time step each 0.1 μs. **C)** RMSD on Cα carbons of the glycosylated RDB. Blue, green and orange for the Manp9, core complex and Complex glycosylation, respectively. **D)** Radius of gyration for each N-glycosylation site in the individual MD. **E)** Total percentage of buried RBD protein surface by the chains of Manp9, core complex, and complex glycosylations.

ii) N-glycosylation interacting with protein and Calcium

Dendritic Cell-Specific Intracellular adhesion molecule-3 (ICAM-3) Grabbing Nonintegrin Receptor (DC-SIGNR) is a receptor that shares 77% of sequence identity with DC-SIGN, a type II membrane protein with a carboxylic terminal carbohydrate recognition domain (CRD) that belongs to the C-type (Ca^{2+}-dependent) lectin superfamily. The binding of the glycan {β-D-GlcpNAc-(1-2)-α-D-Manp-(1-6) [β-D-GlcpNAc-(1-2)-α-D-Manp-(1-3)]-α-D-Manp)} to ICAM-3[56] is crucially mediated by Ca^{2+}, which is jointly coordinated by 3 Aspartates from the protein and Hydroxyl groups from a Mannose residue (PDB id.: 1K9J). The interaction between α-D-Manp and Ca^{2+} is important because it is present in all C-type lectins[56]. From the simulation standpoint, the tripartite interaction protein-glycan-Ca^{2+} ions shown in Figure 4A constitutes a highly challenging system. The glycan participating in this complex (see inset in Figure 4A) is coordinated by Ca^{2+} and a series of contacts.

During the CG simulation, the glycan experiences some mobility but remained attached to the binding site at all times. As can be inferred from Figure 4B the glycan is more flexible in the terminal sugar rings, which generates lower occupational density. The contact with Phe325 (Figure 4B), which was reported to be crucial for this interaction[56], was fully maintained. Indeed, the RMSD values of the protein and glycan resulted remarkably low for a CG simulation (Figure 4C). The 1-6 glycosidic bond has more degrees of freedom than a 1-3 link. That makes the glycan's "arm 1-6" more mobile in comparison with the "arm 1-3", which is in contact with Ca^{2+}. This movement is visualized in the occupancy volume of the Figure 4B, where the occupancy is centered over the zone where the mannose interacts with Ca^{2+}. In contrast, there is a loss of contacts in the extreme of the arm 1-6, this is something expected and well captured by our model owing to the flexibility of the Ω angle in the 1-6 bond. The RMSD of the glycan drops at the end of simulation when the arm 1-6 rotates to a position closer to the original and recover interactions with Asn323 (Figure 4B and C). The conservation of native contacts, between carbohydrate and protein, dropped to nearly 30% of the original value. Thus, reflecting certain promiscuity in the contacts, which is inherent to the loss of details in the CG approach, and the natural flexibility of the carbohydrate, represented here mostly by the mentioned movement of arm 1-6 (Figure 4D). Nevertheless, measuring the BSA at the protein-glycan interface showed values well comparable to the experimental determination (Figure 4E).

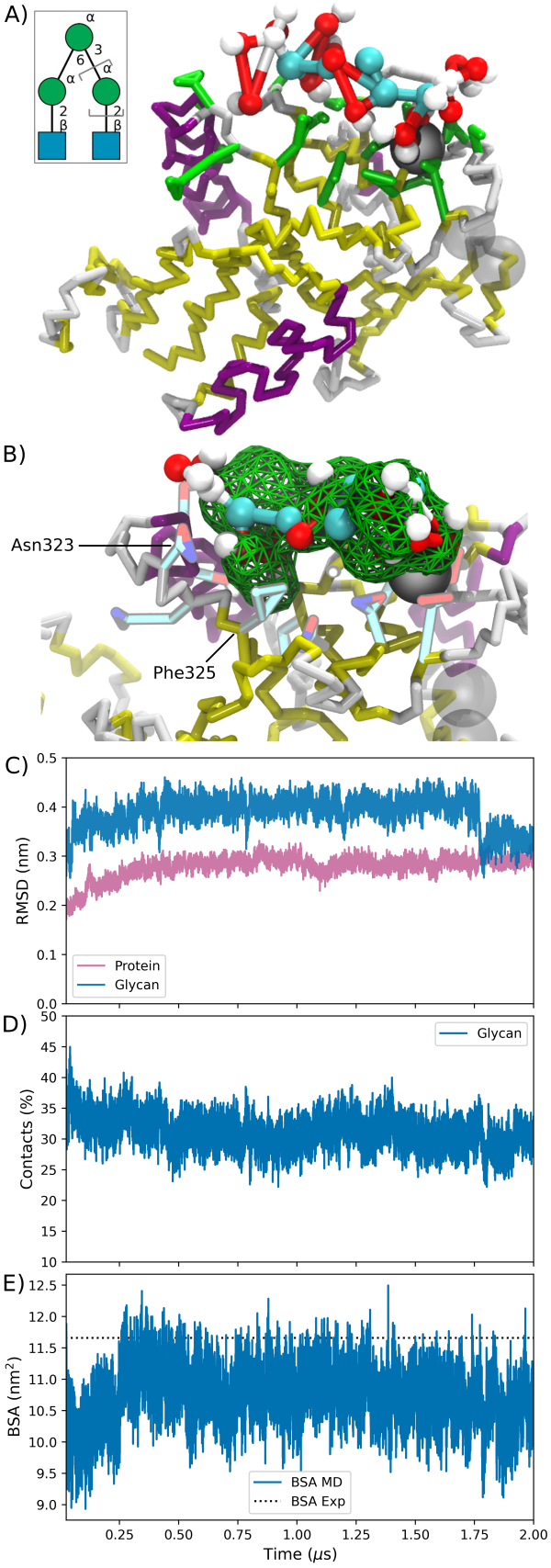

Figure 4. A) CG representation DC-SIGNR in complex with a glycan (PDB id.: 1K9J). The protein backbone is shown in licorice and colored by the secondary structure (helix: purple, sheet: yellow and coil: white). The bound glycan is presented

as balls and sticks. The laterals chains of the protein at 0.4 nm from the glycan are presented as green sticks. Solid and transparent gray spheres represent Ca^{2+} ions interacting and non-interacting with the glycan, respectively. **Inset:** Sequence of the glycan bonded to DC-SIGN in SNFG representation, the mannose residue between square brackets interacts with Ca^{+2}. **B)** Close up on the glycan-protein interaction zone. The green mesh represents the isosurface enclosing the 50% of the occupational density of the glycan over the trajectory. **C)** RMSD for the protein backbone and the ring beads of the glycan. **D)** Glycan-protein native contacts. **E)** BSA of the glycan in the context of the protein.

iii) N-glycosylation interacting with protein

As a final example, we addressed a system with increased size and complexity. We simulated the ectodomain of the homodimer of the Envelope protein (E) from Dengue 2 virus in complex with the Fab fragments of a neutralizing antibody[57]. According to the PDB structure 4UT6, the Dengue´s E protein crystallized as a dimer, with each protomer bearing two different glycosylation motifs, at Asn67 and Asn153. Each antibody presents extensive contacts with both amino acids and glycosylations from each E protein (Fig. 5A).

The inset in the Figure 5A shows the primary sequences of the N-glycosylations at positions 67 and 153, in SNFG representation. For residues Asn153 on chains A and B in the asymmetric unit of structure 4UT6., the primary sequence of the glycans correspond to β-D-GlcpNAc-(1-N)-Asn67 and α-D-Manp-(1-6)[α-D-Manp-(1-3)]-β-D-Manp-(1-4)-β-D-GlcpNAc-(1-4)[α-L-Fucp-(1-6)]-β-D-GlcpNAc-(1-N)-Asn153, respectively.

The results obtained for each N-glycosylation varied depending on the site and the protomer because of the differences in the protein-protein interfaces and the Fab chains. Another source of asymmetry is the ring conformation at the ramification points and the fucoses on the glycosylation at the Asn153, as they are not in $^4C_1$ ring conformation. It is uncertain if this unusual conformation in the X-ray structure results from an artifact in the refinement[29] associated with the poor local resolution or reflects the actual sugar conformation. Regardless the case, this system poses a significant challenge to our CG, as it is based on the most stable sugar ring conformation, namely, $^4C_1$. Therefore, it could not be ruled out that these features could produce unphysically distorted conformations. However, despite some rapid conformational modifications during the minimization/equilibration stages, the global structure and the interface areas remained very close to the crystallographic ones (see below).

During the CG simulation, the E protein's dimer experience a conformational transition that alter both proteins' relative orientation. As shown in Figures 5A and B, the angle between the two protomers (gray line) goes from -172.7 degrees in the crystal packed situation to oscillate around -160° after 400 ns. It is worth noting that this movement drives the complex to a conformation similar to that found in

the structure of the entire virus-like particles, where the same angle is -163.72 degrees[58]. Because of this conformational change, the RMSD of the E proteins raised to relatively high values. Nevertheless, the same quantity measured on the antibody chains remained comparable to other proteins simulated with our force field (Figure 5C).

In order to compare the glycosylations in Asn63-A and Asn153-B against one Fab fragment and the glycosylations in Asn63-B y Asn153-A against the other one, the native contacts and BSA (Figure 5D and E) of each glycosylation site is plotted below his corresponding position in Figure 5A. The change in the angle between both E proteins impacts the contacts between the antibodies and the sugar chains. As shown in Figure 5D (right panel, blue trace), the native contacts of Asn67 in chain A drop during the equilibration phase from 100% to marginal values, while in the same chain the glycosylation on Asn153 remained at relatively higher levels (left panel, blue trace). The opposite behavior is observed for chain B (Figure 5D, green lines). Despite this, the buried surface areas between antibodies and glycans remain well compatible with experimental values (Figure 5E).

For the glycosylations, the interactions between the glycans and the proteins were well maintained during the MD (Table 2 and Figure 5E). Compared with the previous examples, the native contacts showed a decrease in their percentages but the BSA showed a good conservation (Figure 5D). The stabilizing effect of the antibodies on the glycan dynamics is evident form a comparison between the RMSDs measured in the previous examples and this one (Table 2).

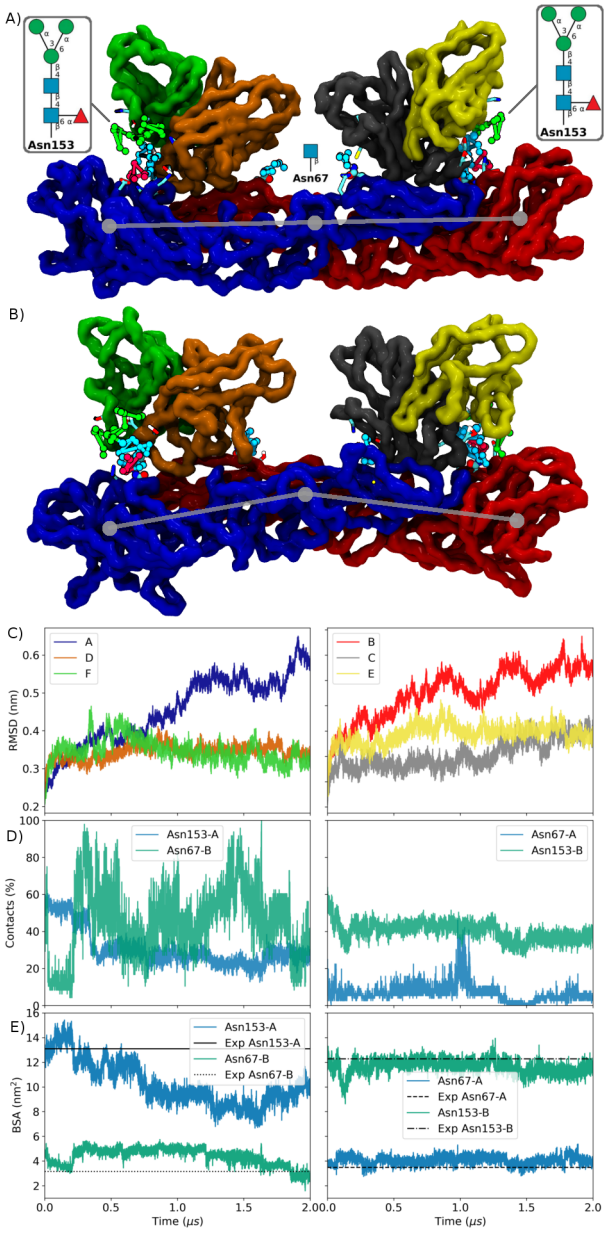

Figure 5. A) Backbone representation of PDB structure 4UT6. Proteins are colored by chain (A: blue, B: red). Chains correspond to proteins and C (gray), D (orange) and E (yellow) chains are Fab fragments. Glycosylations are presented as balls and sticks with SNFG colors. The side chains within 0.4 nm of any glycan are shown and colored by element. The angle between the two proteins is indicated by a gray line. **Insets:** Primary sequences in SNFG representation. **B)** Same as A for the structure at the end of MD. **C)** RMSD of each polypeptide chain in the asymmetric unit of 4UT6. **D)** Native contacts (%) for each glycosylation, the ones for chain A are in blue and the ones for chain B are in green. **E)** BSA for each glycosylation with solid, dash and dash-dot lines indicating the BSA obtained from the MD and the experimental structures, respectively.

PDB / Molecule	Segment / Chain	RMSD (nm)	RG (nm)	BSA (nm^2)	Contacts (%)	Secondary Structure (%)		
						Helix	Extended	Coil
RBD – Man9	R1	0.41 (0.08)	[0.87] 0.83 (0.02)	N/A	N/A	N/A	N/A	N/A
	R2	0.70 (0.13)	[0.87] 0.84 (0.02)					
	Protein	0.39 (0.05)	[1.99] 1.92 (0.06)			[9.2] 5.6 (0.8)	[39.8] 31.8 (2.9)	[51.00] 62.6 (3.2)
RBD – Core Complex	R1	0.40 (0.04)	[0.76] 0.71 (0.02)	N/A	N/A	N/A	N/A	N/A
	R2	0.44 (0.05)	[0.82] 0.70 (0.03)					
	Protein	0.39 (0.03)	[1.99] 1.92 (0.03)			[9.2] 10.5 (0.8)	[39.8] 28.4 (2.3)	[51.00] 61.1 (2.5)
RBD – Complex	R1	0.67 (0.05)	[1.17] 0.85 (0.05)	N/A	N/A	N/A	N/A	N/A
	R2	0.68 (0.05)	[1.16] 0.86 (0.03)					
	Protein	0.43 (0.05)	[1.99] 1.98 (0.04)			[9.2] 10.0 (1.2)	[39.8] 29.3 (2.5)	[51.00] 60.7 (3.1)
1K9J	Glycan	0.39 (0.03)	[0.70] 0.62 (0.03)	[11.66] 10.74 (0.45)	31.23 (2.56)	N/A	N/A	N/A
	Protein	0.28 (0.02)	[2.01] 2.14 (0.03)	N/A	N/A	[19.4] 22.7 (0.4)	[48.4] 27.5 (2.1)	[32.2] 49.8 (2.1)
4UT6 – Glycans	Asn153-A	0.38 (0.07)	[0.71] 0.64 (0.03)	[13.09] 10.25 (1.81)	30.57 (9.80)	N/A	N/A	N/A
	Asn67-A	0.13 (0.01)	N/A	[3.49] 4.05 (0.35)	6.70 (4.71)			
	Asn153-B	0.38 (0.03)	[0.71] 0.58 (0.02)	[12.28] 11.64 (0.55)	40.26 (4.86)			
	Asn67-B	0.13 (0.02)	N/A	[3.12] 4.35 (0.71)	45.17 (18.80)			
4UT6 – Protein	A	0.45 (0.09)	[3.41] 3.31 (0.05)	N/A	N/A	[4.6] 3.0 (0.3)	[58.4] 40.5 (2.1)	[37.1] 56.5 (2.1)
	B	0.42 (0.05)	[3.41] 3.39 (0.02)			[4.8] 2.8 (0.4)	[57.6] 44.7 (1.8)	[37.6] 52.5 (1.9)
	C	0.30 (0.02)	[1.43] 1.51 (0.01)			[6.2] 1.0 (0.9)	[59.2] 42.1 (3.9)	[34.6] 56.9 (4.1)
	D	0.34 (0.02)	[1.43] 1.47 (0.01)			[6.2] 3.7 (0.7)	[60.0] 47.4 (3.1)	[33.8] 48.9 (3.3)
	E	0.34 (0.02)	[1.32] 1.37 (0.01)			[2.8] 0.0 (0.3)	[65.1] 45.9 (3.6)	[32.1] 54.1 (3.6)
	F	0.34 (0.03)	[1.33] 1.37 (0.01)			[2.8] 1.2 (0.9)	[65.4] 47.3 (3.4)	[31.8] 51.5 (3.4)

Table 2. Averages, standard deviations, and experimental values from RMSD, Radius of Gyration (RG), Buried Surface Area (BSA) in protein-glycan interface, Contacts and Secondary structure for the systems presented in this work. Standard deviations are indicated in parentheses, experimental values in square brackets.

Conclusions and outlook

The molecular diversity of the glycan world is virtually infinite. Here we presented a mapping scheme and parametrization of CG glycans within the SIRAH force field, along with a set of examples to illustrate the CG model's capabilities. The current set of parameters include virtually all sugar rings present in N-glycosylations branched or not. The cases of study provide examples of glycans exposed to solvent, acting as a ligand, interacting with Ca^{2+}, mediating glycan-protein, and protein-antibody recognition. The results show very good agreement with experimental data. Although the set of residues parameterized is seemingly small (six pyranoses and one fucose), the mapping scheme developed granted the possibility to consider D or L isomers, α or β linkages and branched chains present in N-glycosylations. For instance, the intricate glycosylation pattern recently reported for the Spike protein of SARS-CoV-2[53] is perfectly amenable with the CG parameters reported here.

Among the limitations, it is important to highlight that interconversions from $^{4}C_{1}$ to $^{1}C_{4}$ and vice versa, are not possible using the present mapping scheme. If different conformations in a sugar ring are needed, they should be included as different residues in the force field. Similarly, the modeling of furanose rings has not been introduced here, although the mapping scheme used here for pyranose rings (Figure 1) could provide a general rule for creating CG parameters for furanoses.

It is important to underline that the current parameterization uses bead types already existing in SIRAH following a line of development inspired in the concept of functional chemical groups. This makes the glycan representation consistent with the existing force field by construction, and opens the possibility to easily expand the current library of residues using the other functional groups already present in SIRAH. For instance, hydroxyl, N-acetyl, propyl, carboxylic, phosphate and sulfate groups[25,26] will allow covering carbohydrates families like polysaccharides glycosaminoglycans and proteoglycans, and other glycoconjugates.

Acknowledgments

P.G.G., M.R.M. and S.P. belong to the SNI program (Sistema Nacional de Investigadores) of ANII (Agencia Nacional de Investigación e Innovación, Uruguay). P.G.G. is beneficiary of a postdoctoral fellowship of CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina). Some of the calculations were carried out using Cluster-UY (site: https://cluster.uy). Some of the graphic cards used in this research were donated by the NVIDIA Corporation. Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The authors acknowledge the National Laboratory for Scientific Computing (LNCC/MCTI, Brazil) for providing HPC resources of the SDumont supercomputer, which have contributed to the research results reported within this paper. URL: http://sdumont.lncc.br

We thank Carlos Simmerling, Darya Stepanenko, and Lauren Raguette for useful discussions and testing of the force field.

References

(1) von der Lieth, C.-W.; Lütteke, T.; Frank, M. *Bioinformatics for Glycobiology and Glycomics: An Introduction*; von der Lieth, C.-W., Ltteke, T., Frank, M., Eds.; John Wiley & Sons, Ltd: Chichester, UK, 2009. https://doi.org/10.1002/9780470029619.fmatter.

(2) Apweiler, R. On the Frequency of Protein Glycosylation, as Deduced from Analysis of the SWISS-PROT Database. *Biochim. Biophys. Acta - Gen. Subj.* **1999**, *1473* (1), 4–8. https://doi.org/10.1016/S0304-4165(99)00165-8.

(3) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. *Sci. Rep.* **2011**, *1*, 1–5. https://doi.org/10.1038/srep00090.

(4) Tan, F. Y. Y.; Tang, C. M.; Exley, R. M. Sugar Coating: Bacterial Protein Glycosylation and Host-Microbe Interactions. *Trends Biochem. Sci.* **2015**, *40* (7), 342–350. https://doi.org/10.1016/j.tibs.2015.03.016.

(5) Esteban, C.; Donati, I.; Pantano, S.; Villegas, M.; Benegas, J.; Paoletti, S. Dissecting the Conformational Determinants of Chitosan and Chitlac Oligomers. *Biopolymers* **2018**, *109* (6), 1–9. https://doi.org/10.1002/bip.23221.

(6) Verli, H.; Guimarães, J. A. Molecular Dynamics Simulation of a Decasaccharide Fragment of Heparin in Aqueous Solution. *Carbohydr. Res.* **2004**, *339* (2), 281–290. https://doi.org/10.1016/j.carres.2003.09.026.

(7) Perić-Hassler, L.; Hansen, H. S.; Baron, R.; Hünenberger, P. H. Conformational Properties of Glucose-Based Disaccharides Investigated Using Molecular Dynamics Simulations with Local Elevation Umbrella Sampling. *Carbohydr. Res.* **2010**, *345* (12), 1781–1801. https://doi.org/10.1016/j.carres.2010.05.026.

(8) Jo, S.; Qi, Y.; Im, W. Preferred Conformations of N-Glycan Core Pentasaccharide in Solution and in Glycoproteins. *Glycobiology* **2015**, No. 785, cwv083. https://doi.org/10.1093/glycob/cwv083.

(9) Patel, D. S.; Qi, Y.; Im, W. Modeling and Simulation of Bacterial Outer Membranes and Interactions with Membrane Proteins. *Curr. Opin. Struct. Biol.* **2017**, *43* (Figure 1), 131–140. https://doi.org/10.1016/j.sbi.2017.01.003.

(10) Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. The Power of Coarse Graining in Biomolecular Simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4* (3), 225–248. https://doi.org/10.1002/wcms.1169.

(11) López, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hünenberger, P. H.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5* (12), 3195–3210. https://doi.org/10.1021/ct900313w.

(12) López, C. A.; Sovova, Z.; Van Eerden, F. J.; De Vries, A. H.; Marrink, S. J. Martini Force Field

Parameters for Glycolipids. *J. Chem. Theory Comput.* **2013**, *9* (3), 1694–1708. https://doi.org/10.1021/ct3009655.

(13) Shivgan, A. T.; Marzinek, J. K.; Huber, R. G.; Krah, A.; Henchman, R. H.; Matsudaira, P.; Verma, C. S.; Bond, P. J. Extending the Martini Coarse-Grained Force Field to N -Glycans. *J. Chem. Inf. Model.* **2020**, *60* (8), 3864–3883. https://doi.org/10.1021/acs.jcim.0c00495.

(14) Schmalhorst, P. S.; Deluweit, F.; Scherrers, R.; Heisenberg, C.-P. P.; Sikora, M. Overcoming the Limitations of the MARTINI Force Field in Simulations of Polysaccharides. *J. Chem. Theory Comput.* **2017**, *13* (10), 5039–5053. https://doi.org/10.1021/acs.jctc.7b00374.

(15) Rusu, V. H.; Baron, R.; Lins, R. D. PITOMBA: Parameter Interface for Oligosaccharide Molecules Based on Atoms. *J. Chem. Theory Comput.* **2014**, *10* (11), 5068–5080. https://doi.org/10.1021/ct500455u.

(16) Pol-Fachin, L.; Rusu, V. H.; Verli, H.; Lins, R. D. GROMOS 53A6 GLYC , an Improved GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J. Chem. Theory Comput.* **2012**, *8* (11), 4681–4690. https://doi.org/10.1021/ct300479h.

(17) Liwo, A.; Baranowski, M.; Czaplewski, C.; Gołaś, E.; He, Y.; Jagieła, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M. A.; et al. A Unified Coarse-Grained Model of Biological Macromolecules Based on Mean-Field Multipole-Multipole Interactions. *J. Mol. Model.* **2014**, *20* (8). https://doi.org/10.1007/s00894-014-2306-5.

(18) Lubecka, E. A.; Liwo, A. A General Method for the Derivation of the Functional Forms of the Effective Energy Terms in Coarse-Grained Energy Functions of Polymers. II. Backbone-Local Potentials of Coarse-Grained O1→4-Bonded Polyglucose Chains. *J. Chem. Phys.* **2017**, *147* (11), 115101. https://doi.org/10.1063/1.4994130.

(19) Dans, P. D.; Darré, L.; Machado, M. R.; Zeida, A.; Brandner, A. F.; Pantano, S. Assessing the Accuracy of the SIRAH Force Field to Model DNA at Coarse Grain Level; Springer, Cham, 2013; pp 71–81. https://doi.org/10.1007/978-3-319-02624-4_7.

(20) Darré, L.; Machado, M. R.; Dans, P. D.; Herrera, F. E.; Pantano, S. Another Coarse Grain Model for Aqueous Solvation: WAT FOUR? *J. Chem. Theory Comput.* **2010**, *6* (12), 3793–3807. https://doi.org/10.1021/ct100379f.

(21) Darré, L.; Machado, M. R.; Brandner, A. F.; González, H. C.; Ferreira, S.; Pantano, S. SIRAH: A Structurally Unbiased Coarse-Grained Force Field for Proteins with Aqueous Solvation and Long-Range Electrostatics. *J. Chem. Theory Comput.* **2015**, *11* (2), 723–739. https://doi.org/10.1021/ct5007746.

(22) Machado, M. R.; Barrera, E. E.; Klein, F.; Sóñora, M.; Silva, S.; Pantano, S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *J. Chem. Theory Comput.* **2019**, acs.jctc.9b00006. https://doi.org/10.1021/acs.jctc.9b00006.

(23) Barrera, E. E.; Frigini, E. N.; Porasso, R. D.; Pantano, S. Modeling DMPC Lipid Membranes

with SIRAH Force-Field. *J. Mol. Model.* **2017**, *23* (9), 259. https://doi.org/10.1007/s00894-017-3426-5.

(24) Barrera, E. E.; Machado, M. R.; Pantano, S. Fat SIRAH: Coarse-Grained Phospholipids To Explore Membrane–Protein Dynamics. *J. Chem. Theory Comput.* **2019**, *15* (10), 5674–5688. https://doi.org/10.1021/acs.jctc.9b00435.

(25) Garay, P. G.; Barrera, E. E.; Pantano, S. Post-Translational Modifications at the Coarse-Grained Level with the SIRAH Force Field. *J. Chem. Inf. Model.* **2020**, *60* (2), 964–973. https://doi.org/10.1021/acs.jcim.9b00900.

(26) Klein, F.; Caceres-Rojas, D.; Carrasco, M.; Tapia, J. C.; Caballero, J.; Alzate-Morales, J. H.; Pantano, S. Coarse-Grained Parameters for Divalent Cations within the SIRAH Force Field. *J. Chem. Inf. Model.* **2020**, acs.jcim.0c00160. https://doi.org/10.1021/acs.jcim.0c00160.

(27) Böhm, M.; Bohne-Lang, A.; Frank, M.; Loss, A.; Rojas-Macias, M. A.; Lütteke, T. Glycosciences.DB: An Annotated Data Collection Linking Glycomics and Proteomics Data (2018 Update). *Nucleic Acids Res.* **2018**. https://doi.org/10.1093/nar/gky994.

(28) Lütteke, T.; Frank, M.; von der Lieth, C.-W. Carbohydrate Structure Suite (CSS): Analysis of Carbohydrate 3D Structures Derived from the PDB. *Nucleic Acids Res.* **2005**, *33* (Database issue), D242-6. https://doi.org/10.1093/nar/gki013.

(29) de Meirelles, J. L.; Nepomuceno, F. C.; Peña-García, J.; Schmidt, R. R.; Pérez-Sánchez, H.; Verli, H. Current Status of Carbohydrates Information in the Protein Data Bank. *J. Chem. Inf. Model.* **2020**, *60* (2), 684–699. https://doi.org/10.1021/acs.jcim.9b00874.

(30) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29* (4), 622–655. https://doi.org/10.1002/jcc.20820.

(31) Machado, M. R.; Zeida, A.; Darré, L.; Pantano, S. From Quantum to Subcellular Scales: Multi-Scale Simulation Approaches and the SIRAH Force Field. *Interface Focus* **2019**, *9* (3), 20180085. https://doi.org/10.1098/rsfs.2018.0085.

(32) Machado, M. R.; Pantano, S. SIRAH Tools: Mapping, Backmapping and Visualization of Coarse-Grained Models. *Bioinformatics* **2016**, *32* (10), 1568–1570. https://doi.org/10.1093/bioinformatics/btw020.

(33) Varki, A.; Cummings, R. D.; Aebi, M.; Packer, N. H.; Seeberger, P. H.; Esko, J. D.; Stanley, P.; Hart, G.; Darvill, A.; Kinoshita, T.; et al. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **2015**, *25* (12), 1323–1324. https://doi.org/10.1093/glycob/cwv091.

(34) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3* (2), 198–210. https://doi.org/10.1002/wcms.1121.

(35) Machado, M. R.; Pantano, S. Split the Charge Difference in Two! A Rule of Thumb for Adding

Proper Amounts of Ions in MD Simulations. *J. Chem. Theory Comput.* **2020**, *16* (3), 1367–1372. https://doi.org/10.1021/acs.jctc.9b00953.

(36) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878–3888. https://doi.org/10.1021/ct400314y.

(37) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.

(38) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593. https://doi.org/10.1063/1.470117.

(39) Wu, X.; Brooks, B. R. Self-Guided Langevin Dynamics Simulation Method. *Chem. Phys. Lett.* **2003**, *381* (3), 512–518. https://doi.org/https://doi.org/10.1016/j.cplett.2003.10.013.

(40) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(41) Perez, F.; Granger, B. E. IPython: A System for Interactive Scientific Computing. *Comput. Sci. Eng.* **2007**, *9* (3), 21–29. https://doi.org/10.1109/MCSE.2007.53.

(42) McKinney, W. Data Structures for Statistical Computing in Python. *PROC. 9th PYTHON Sci. Conf.* **2010**, *445*, 51–56.

(43) van der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13* (2), 22–30. https://doi.org/10.1109/MCSE.2011.37.

(44) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9* (3), 90–95. https://doi.org/10.1109/MCSE.2007.55.

(45) Waskom, M.; Botvinnik, O.; O'Kane, D.; Hobson, P.; Ostblom, J.; Lukauskas, S.; Gemperline, D. C.; Augspurger, T.; Halchenko, Y.; Cole, J. B.; et al. Seaborn: V0.9.0 (July 2018). **2018**. https://doi.org/10.5281/ZENODO.1313201.

(46) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32* (10), 2319–2327. https://doi.org/10.1002/jcc.21787.

(47) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I.; et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In *Proceedings of the 15th Python in Science Conference*; Benthall, S., Rostrup, S., Eds.; 2016; pp 98–105. https://doi.org/10.25080/Majora-629e541a-00e.

(48) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(49) Cheng, K.; Zhou, Y.; Neelamegham, S. DrawGlycan-SNFG: A Robust Tool to Render Glycans and Glycopeptides with Fragmentation Information. *Glycobiology* **2017**, *27* (3), 200–205. https://doi.org/10.1093/glycob/cww115.

(50) Cui, J.; Li, F.; Shi, Z. L. Origin and Evolution of Pathogenic Coronaviruses. *Nat. Rev. Microbiol.* **2019**, *17* (3), 181–192. https://doi.org/10.1038/s41579-018-0118-9.

(51) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. *Science (80-.).* **2020**, *367* (6483), 1260–1263. https://doi.org/10.1126/science.abb2507.

(52) Zhao, P.; Praissman, J. L.; Grant, O. C.; Cai, Y.; Xiao, T.; Rosenbalm, K. E.; Aoki, K.; Kellman, B. P.; Bridger, R.; Barouch, D. H.; et al. Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host Microbe* **2020**, *28* (4), 586-601.e6. https://doi.org/10.1016/j.chom.2020.08.004.

(53) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-Specific Glycan Analysis of the SARS-CoV-2 Spike. *Science (80-.).* **2020**, *9983* (May), eabb9983. https://doi.org/10.1126/science.abb9983.

(54) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **2020**, *6* (10), 1722–1734. https://doi.org/10.1021/acscentsci.0c01056.

(55) Garay, P.; Barrera, E.; Klein, F.; Machado, M.; Soñora, M.; Pantano, S. SIRAH-CoV2 Initiative: Glycosylated RBD. Zenodo June 2020. https://doi.org/10.5281/zenodo.3903522.

(56) Feinberg, H.; Mitchell, D. A.; Drickamer, K.; Weis, W. I. Structural Basis for Selective Recognition of Oligosaccharides by DC-SIGN and DC-SIGNR. *Science (80-.).* **2001**, *294* (5549), 2163–2166. https://doi.org/10.1126/science.1066371.

(57) Rouvinski, A.; Guardado-Calvo, P.; Barba-Spaeth, G.; Duquerroy, S.; Vaney, M.-C.; Kikuti, C. M.; Navarro Sanchez, M. E.; Dejnirattisai, W.; Wongwiwat, W.; Haouz, A.; et al. Recognition Determinants of Broadly Neutralizing Human Antibodies against Dengue Viruses. *Nature* **2015**, *520* (7545), 109–113. https://doi.org/10.1038/nature14130.

(58) Zhang, X.; Ge, P.; Yu, X.; Brannan, J. M.; Bi, G.; Zhang, Q.; Schein, S.; Hong Zhou, Z. Cryo-EM Structure of the Mature Dengue Virus at 3.5-Å Resolution. *Nat. Struct. Mol. Biol.* **2013**, *20* (1), 105–110. https://doi.org/10.1038/nsmb.2463.