Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

CHAPTER NINE

Glycoprotein molecular dynamics analysis: SARS-CoV-2 spike glycoprotein case study

João Victor Paccini Coutinho[a,†], Janaina Macedo-da-Silva[a,†], Simon Ngao Mule[a], Thales Kronenberger[b,c,d,e], Livia Rosa-Fernandes[a], Carsten Wrenger[a], and Giuseppe Palmisano[a,f,*]

[a]Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil
[b]Department of Internal Medicine VIII, University Hospital Tuebingen, Tuebingen, Germany
[c]Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Eberhard-Karls-Universität, Tuebingen, Germany
[d]Cluster of Excellence iFIT (EXC 2180) "Image-Guided and Functionally Instructed Tumor Therapies", University of Tuebingen, Tuebingen, Germany
[e]Tuebingen Center for Academic Drug Discovery & Development (TüCAD2), Tuebingen, Germany
[f]Faculty of Science and engineering, Macquarie University, Sydney, NSW, Australia
*Corresponding author: e-mail addresses: palmisano.gp@gmail.com; palmisano.gp@usp.br

Contents

Abstract

Molecular Dynamics (MD) is a method used to calculate the movement of atoms and molecules broadly applied to several aspects of science. It involves computational simulation, which makes it, at first glance, not easily accessible. The rise of several

† These authors contributed equally to the work.

Advances in Protein Chemistry and Structural Biology, Volume 131
ISSN 1876-1623
https://doi.org/10.1016/bs.apcsb.2022.05.004

automated tools to perform molecular simulations has allowed researchers to navigate through the various steps of MD. This enables to elucidate structural properties of proteins that could not be analyzed otherwise, such as the impact of glycosylation. Glycosylation dictates the physicochemical and biological properties of a protein modulating its solubility, stability, resistance to proteolysis, interaction partners, enzymatic activity, binding and recognition. Given the high conformational and compositional diversity of the glycan chains, assessing their influence on the protein structure is challenging using conventional analytical techniques. In this manuscript, we present a step-by-step workflow to build and perform MD analysis of glycoproteins focusing on the SPIKE glycoprotein of SARS-CoV-2 to appraise the impact of glycans in structure stabilization and antibody occlusion.

Glycoproteins Molecular dynamics
SARS-CoV-2 Protein structure
Glycans Antigen-Antibody interaction

Abbreviations

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a highly pathogenic virus and the agent responsible for the pandemic that began in late 2019 (Baloch, Baloch, Zheng, & Pei, 2020), which has already led to the death of more than 6 million people in the world and 434 million cases (WHO Coronavirus COVID-19 Dashboard, n.d.). Like other viruses in the *Coronovidae* family, SARS-CoV-2 also encodes the SPIKE (S) glycoprotein, which has an important role in infection, including virus entry into the target cells (Ou et al., 2020). This glycoprotein is also essential in the host's immune evasion, representing the main target of therapeutic interventions (Duan et al., 2020). Multiple SARS-CoV-2 variants of concern arose since the beginning of the pandemic, some of which initiated a new wave of infections. These variants include delta (Planas et al., 2021), omicron (Wolter et al., 2022), P1 (Naveca et al., 2021), among others (Hacisuleyman et al., 2021) and have mutations mainly in the S protein, which facilitates successful infection with higher transmission rate. The most relevant mutations in protein S are D614G, present in all variants of Concern (VOC) and it is responsible for accumulating higher viral load and infectivity (Volz et al., 2021). On the other hand, the N501Y mutation

promotes new stabilizing hydrogen bonds between the receptor binding domain (RBD) and ACE-2 receptor, with ~10 times higher binding affinity (Liu, Wei, et al., 2021; Liu, Zhang, et al., 2021). The presence of HV69-70del is required to induce this effect. The absence of this deletion does not allow immune evasion in B.1.1.7 strains (Meng et al., 2021). The L452R, T478K, and P681H/R mutations promote an increase in infectivity, possibly due to an increase in the positive electrostatic potential of the surface and greater steric impediments (Pascarella et al., 2021; Tian, Sun, Zhou, & Ye, 2022). In contrast, the E484K mutation promotes less antibody neutralization and greater escape from the immune system (Lippi, Mattiuzzi, & Henry, 2022; Liu, Wei, et al., 2021).

Another additional layer of structural complexity comes from the conformational changes induced by post-translational modifications (PTMs). Protein glycosylation is a PTM that can have profound effects on the structure and function of a glycoprotein (Macedo-da-Silva, Santiago, Rosa-Fernandes, Marinho, & Palmisano, 2021; Reily, Stewart, Renfrow, & Novak, 2019; Schjoldager, Narimatsu, Joshi, & Clausen, 2020). N-linked glycosylation (N-X-S/T motif, where X is any amino acid residue except proline) is characterized by the addition of oligosaccharides of variable size and complexity to the side chains of asparagine, whereas in O-linked glycosylation, sugar addition occurs at the serine and threonine residues (Aebi, 2013; Hanisch, 2001). The glycan chains can impact in different ways on the protein proprieties, such as stabilizing loops and conformations (Wormald & Dwek, 1999), providing occlusion against antibodies (Ab) of a host cell (Watanabe et al., 2020), promoting correct folding of the protein (Dalziel, Crispin, Scanlan, Zitzmann, & Dwek, 2014), or enhancing lectin-based cell-adhesion (Lasky, 1991). The glycan composition and occupancy of glycosites can vary significantly. This, combined with the high flexibility of glycans, makes a definition of a glycosylated structural model very challenging (Lis & Sharon, 1993; Nagae & Yamaguchi, 2012). Both N- and O-glycosylation represents a key process in viral proteins in multiple aspects of their pathobiology. Since viruses have no metabolic pathways to perform this modification by themselves, they hijack the host-cell glycosylation machinery to their benefit. Glycosylation has been shown to control the infection of several viruses (Sugrue, 2007; Vigerust & Shepherd, 2007). The role of O-linked glycosylation during viral infection is poorly understood. There are reports in the literature that indicate CD99 binding to herpes simplex virus type 1 B glycoprotein

(Wang et al., 2009) and the role of the glycan shield and occlusive neutralization during viral infection. Moreover, O-linked glycan epitopes have be shown to be highly immunogenic in Gammaherpesvirus (Machiels et al., 2011). On the other hand, N-glycan alterations during viral infections have been extensively described, from host cell entry, particle release, and immune evasion as seen in the SPIKE protein of Coronavirus, influenza HA, Ebola GP, and Lassa-virus GPC (Hastie et al., 2017; Kwon et al., 2015; Lee et al., 2014; Mohan, Li, Ye, Compans, & Yang, 2012; Walls et al., 2016; Watanabe, Bowden, Wilson, & Crispin, 2019; Zhao et al., 2016). Due to that, viral glycoproteins represent the major components of the viral envelope and have important roles associated to host cell recognition, attachment, infection, invasion, transmission and immune evasion (Banerjee & Mukhopadhyay, 2016; Cook & Lee, 2013; Watanabe et al., 2019).

Currently, mass-spectrometry-based approaches are the primary choice for the analytical characterization of both site and structure-specific glycosylation (Oliveira, Thaysen-Andersen, Packer, & Kolarich, 2021; Pasing, Sickmann, & Lewandrowski, 2012). This enables a more accurate protein modeling in its biological context. In general, the enrichment of glycopeptides from complex samples is performed by applying strategies based on TiO_2 (Larsen, Jensen, Jakobsen, & Heegaard, 2007; Palmisano et al., 2010), lectin affinity chromatography (Yang & Hancock, 2004), and hydrophilic interaction liquid chromatography (HILIC) (Ongay, Boichenko, Govorukhina, & Bischoff, 2012), followed by glycan release or analysis of intact glycopeptides (Parker, Gupta, Cordwell, Larsen, & Palmisano, 2011; Thaysen-Andersen & Packer, 2014; Thaysen-Andersen, Packer, & Schulz, 2016). For N-linked, the enzyme N-glycosidase F (PNGase F) is usually used, which cleaves between GlcNAc and asparagine residues of N-linked glycoproteins and glycopeptides (Kaji et al., 2003). The release of O-linked occurs mainly by reductive alkaline β-elimination (Wilkinson & Saldova, 2020). The isolated portion of N/O-glycans can be derivatized by permethylation (Kang, Mechref, & Novotny, 2008), and then analyzed by mass spectrometry techniques, as well the glycopeptides, which includes matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF-MS) (Reiding, Blank, Kuijper, Deelder, & Wuhrer, 2014) or by electrospray ionization (ESI-MS/MS) (Morelle & Michalski, 2007; Palmisano, Larsen, Packer, & Thaysen-Andersen, 2013; Palmisano, Melo-Braga, Engholm-Keller, Parker, & Larsen, 2012). Ion mobility mass spectrometry allows access to the native glycocode giving information on the monosaccharide composition and glycosidic linkages (Schindler

et al., 2017). Moreover, novel software solutions have been applied for a comprehensive characterization of thousands of intact glycopeptides in different biological matrices (Alocci et al., 2019; Campbell, 2017; Chernykh, Kawahara, & Thaysen-Andersen, 2021; Kawahara et al., 2021). Native MS allows to transmit large protein ions such as intact glycoprotein complexes which allow the determination of subunits stoichiometry, ligand binding and infer glycoprotein function (Struwe & Robinson, 2019; Wu & Robinson, 2022). Recently, a new method, termed limited deglycosylation assay, was introduced to probe 3D conformational changes of glycoproteins on a proteome-wide scale using limited amount of sample (Mule et al., 2021). This method takes advantage of the accessibility-dependent PNGaseF cleavage to quantitative study changes in glycoprotein conformational changes. X-ray crystallography and cryoEM have limitation in solving the detailed structure of intact glycoproteins due to stoichiometry, heterogeneity and flexibility of the glycan chains. This is known as the glycosylation problem (Chang et al., 2007; Davis & Crispin, 2010). Another approach used in the structural study of glycoproteins is nuclear magnetic resonance (NMR) spectroscopy, a versatile, quantitative, and non-destructive analytical technique (Unione, Ardá, Jiménez-Barbero, & Millet, 2021). NMR provides absolute quantification information for both glycan and glycoprotein. However, the application of the technique has limitations that include the need for large amounts of samples, glycoprotein size and access to adequate labeling (Valverde, Quintana, Santos, Ardá, & Jiménez-Barbero, 2019). NMR overcomes one of the main challenges encountered in glycoproteomics, which consists of differentiating the stereochemistry of a glycan, with isobaric species such as glucose (Glc), galactose (Gal) and mannose (Man) in a single analysis (Hargett et al., 2021). Taken together, the analytical toolbox available to the scientific community allows structural characterization of glycoproteins and correlation with their functions.

The complexity of glycans has always been a challenge to be overcome in molecular modeling, mainly due to the great variety of monomer units that can be organized into stereo and regiospecific-linked oligosaccharides within each glycosylation site. Initial techniques applied to carbohydrate modeling, allowed the elucidation of glycans' conformational and spatial preferences (Lemieux & Koto, 1974; Peters, Meyer, Stuike-Prill, Somorjai, & Brisson, 1993). Subsequently, the development of force fields capable of refining carbohydrate modeling during MD analysis changed the scope of oligosaccharide chain analysis in these simulations, taking into account their fluctuating nature and their interactions with solvent

molecules (Woods, 2018). Development of reliable force-field parameters from experimental data suffers from the lack of training data, given the shortage of 3D structures of branched carbohydrates. These groups are often not resolved or removed from the initial crystal structures, due to considerable thermal fluctuations and flexibility. Among the computational tools to perform MD analysis of glycans and evaluate the force fields, there are AMBER (Pearlman et al., 1995), GLYCAM (Kirschner et al., 2008) and CHARMM (Vanommeslaeghe et al., 2010). MD simulations can reveal several features about the mechanisms and biological functions of glycoproteins and the effects that glycans have on them, as well as evaluate optimal conformations for binding lectins (Cheong, Shim, Kang, & Kim, 1999) or elucidate protein stability in a given conformation (Casalino et al., 2020).

In view of the great importance of the SPIKE glycoprotein in the success of SARS-CoV-2 infection, understanding the structure and impact of adding glycans to this protein can provide important insights to elucidate the mechanisms of viral infection, and highlight possible therapeutic targets. Based on this, we present a step-by-step bioinformatics pipeline to perform the MD of the SARS-CoV-2 SPIKE protein, including carbohydrate structures, solvent, and antibody accessibility. The analysis of the energy states of the performed simulation indicated that it correctly followed the parameters entered and allowed to evaluate and quantify the differential protection of the glycan shield in residues exposed to the solvent. In addition, it was also possible to identify a wide variety of protein-glycan hydrogen bonds and evaluate their interactions within the structure. MD of glycoproteins presents some challenges due to the increase in atoms and in the waterbox size, together with the extension and mobility of the glycans. The MD simulation steps reported here can be applied to other glycoproteins to evaluate various effects that glycosylation can impose on structure and accessibility with potential implications on the biological activities.

2. Methods

2.1 MD simulation workflow applied to SARS-CoV-2 SPIKE glycoprotein

A step-by-step bioinformatic pipeline to perform MD analysis is shown in Fig. 1. This workflow was applied to a structural model of the SPIKE glycoprotein published by Woo et al., 2020. The models were based on

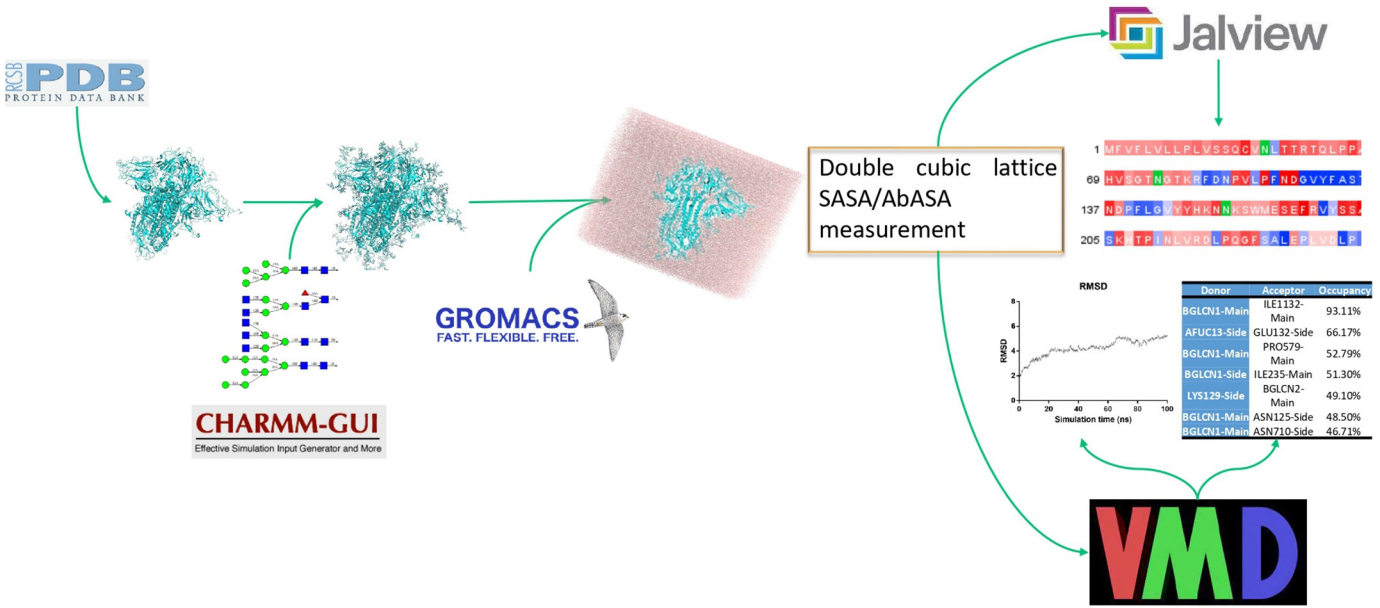

Fig. 1 Computational workflow and software tools used to perform molecular dynamics and evaluate residue specific exposure to solvent and antibody (Ab). The protein of interest structure is downloaded from PDB, in-silico glycosylated and processed to obtain MD simulation using CHARMM-GUI tool. Then, GROMACS is used to run both the simulation (*e.g.*, 100 ns as used in this study) and perform the solvent accessible surface area and average antibody accessible surface area (SASA/AbASA) calculations. VMD software is used to visualize trajectories and RMSD (root mean square deviation) and RMSF (root mean square fluctuation) calculations and Jalview to visualize the calculated proprieties of the protein.

the PDB-6vxx (https://doi.org/10.2210/pdb6vxx/pdb) crystallographic structure and presented a structural resolution of 2.80 Å (Walls et al., 2020). The chosen structure presents a range of modeled residues, which covers most of the head portion of SPIKE, disregarding the membrane and intracellular domains, not considered for this workflow. The detailed structural characterization of the glycosylation sites and glycan occupancy was based on Watanabe et al. (Watanabe et al., 2020; Woo et al., 2020). Mass spectrometric data were used to determine the most frequent glycan structures in each glycosite. A structural model of the protein of interest and site-specific glycan structures are needed to perform MD simulations as described below.

In general, it is advisable to use high-resolution structures whenever available. When selecting a structure for simulations, it is important to analyze the coverage of the structural motifs including the conserved ones, as well as to compare the conformation of individual amino-acids from the catalytic site. The information regarding catalytic/relevant residues could be found in the literature and obtained using site-directed mutagenesis, or from the conserved known motifs of the protein class of interest. The use of multiple structures as starting points as well as multiple simulation replicas as an ensemble is highly advised.

2.2 Computational tools to model glycoproteins

There are multiple computational platforms to model a glycoprotein, such as GLYCAM, Amber, Glycosilator, among others. For this work CHARMM-GUI was used (Jo et al., 2017; Jo, Kim, Iyer, & Im, 2008; Park et al., 2019), due to its web-based intuitive interface for the addition of glycans, an integrated tool for modeling missing residues belonging to unmodelled/low confidence regions, and MD model preparation. It is relevant to mention that full missing loops can be treated by *ab initio* or homology models, despite this being out of the scope of this manuscript (Rodriguez, Chinea, Lopez, Pons, & Vriend, 1998).

The CHARMM-GUI tool can be used to model a large array of biologically relevant structures, such as lipids, glycans, glycoproteins, lipopolysaccharide, solutions and others. For the modeling of glycoproteins, the "Glycan Reader and Modeler" was used, Fig. 2A (Jo, Song, Desaire, MacKerell, & Im, 2011; Park et al., 2019, 2017). PDB models without the oligosaccharide chains can be directly downloaded from the Research Collaboratory for Structural Bioinformatics (RCSB), but can also be uploaded by the user. A relevant note should be made regarding

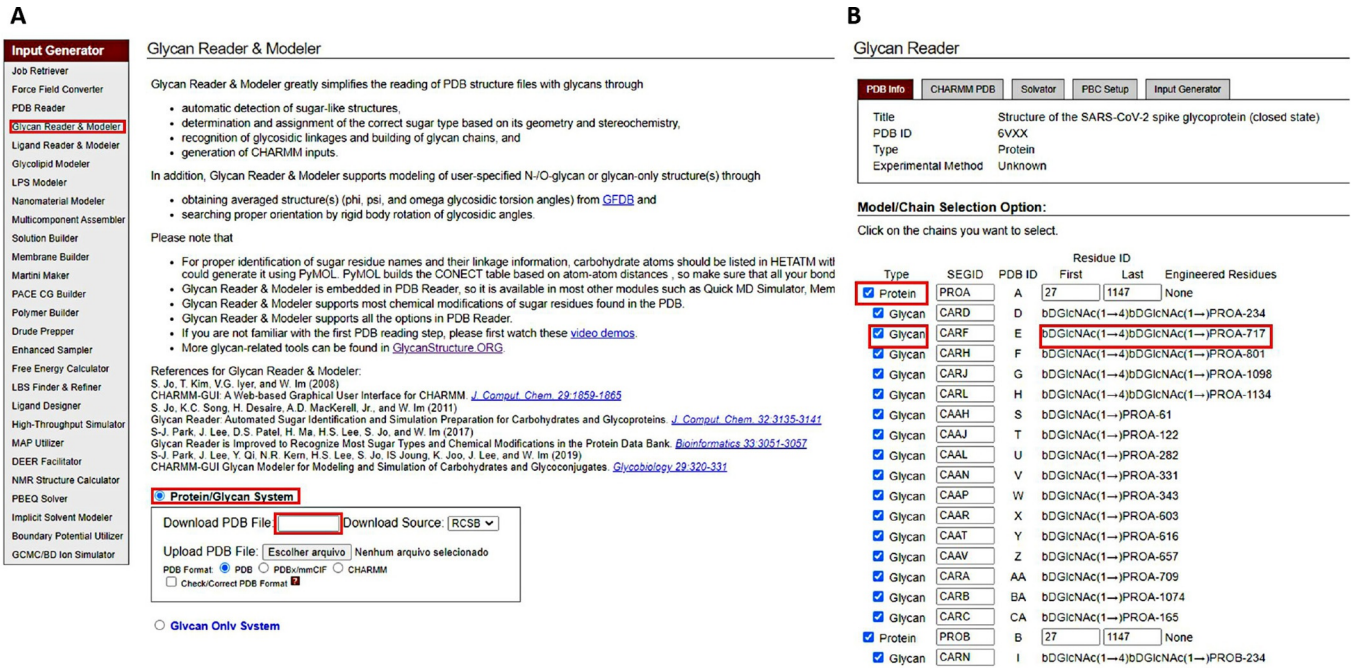

Fig. 2 Defining the model to build the glycoprotein of interest. (A) Depiction of the CHARMM-GUI "Glycan Reader and Modeler" tool used to build glycoproteins. The PDB protein model code is inputted in the "Download PDB File" field on the "Protein/Glycan System." (B) The tool will proceed to download and identify protein chains, connected glycans, ligands, experimental method for structure acquisition and name. The user is able to select all the elements of the model of interest before proceeding for glycoprotein modeling. The glycans code will be displayed by the glycosylated residue site and may be edited on further steps.

uploading glycosylated models. The glycans should be listed as hetero atoms (HETATM) with a CONECT table indicating the bonds among the glycans and the ASN side chain for N-linked glycoproteins. These changes can be added to the PDB by using PyMol software to build a new .pdb file with the coordinates (Fig. 2A).

Once the model is uploaded, the software will detect the protein sequence, chains, missing residues, disulfide bridges, connected glycans, and ligands, all of which might be edited/added/removed in the "Glycan reader" step by deselecting the marked boxes (Fig. 2B).

Upon advancing to the next section, the user will be prompted to edit/fix the protein. Any missing residues and disulfide bridges can be edited before MD simulations and the CHARMM-GUI can perform this modeling in an integrated manner (Fig. 3A and B). The glycans can also be edited in the glycosylation section, where a pop-up menu will bring an editable glycan builder, in which the user can input the appropriate glycans of interest by adding the sugars with the correct linkage (Fig. 3C). With the disulfide bridges added, missing residues modeled and glycans inserted, the glycosylated model is ready to be generated. The CHARMM-GUI will build the structure with the requested properties and prompt the user to the solvation tool, to build a waterbox for the solvent and ions of the protein.

The standard parameters found in the tool will generate a waterbox to match the protein size with an additional 10 Å to the edge and place a physiological concentration of KCl using the Monte-Carlo method, to correct for electrostatic interactions (Fig. 4A).

Then, after setting the periodic boundary conditions (Fig. 4B) and selecting the "force field" CHARMM36m (Huang et al., 2016) to be used for topology generation and inserting standard parameters for the equilibration of the system, the CHARMM-GUI will generate input files for several MD simulation programs as described below (Fig. 4C). The selection of the forcefield of choice can vary on the simulation specificities, but CHARMM36m was chosen due to its improved accuracy for backbone conformational ensembles, besides improved H-bond J coupling. Selecting the GROMACS box as "input generation option," Fig. 4C, will add a folder with all the different files needed for GROMACS MD simulations:

- MD ready model, with glycans, ions and periodic boundaries;
- Topology files;
- Molecular dynamics parameter files (.mdp) for minimization, equilibration and a 100 ns md run.

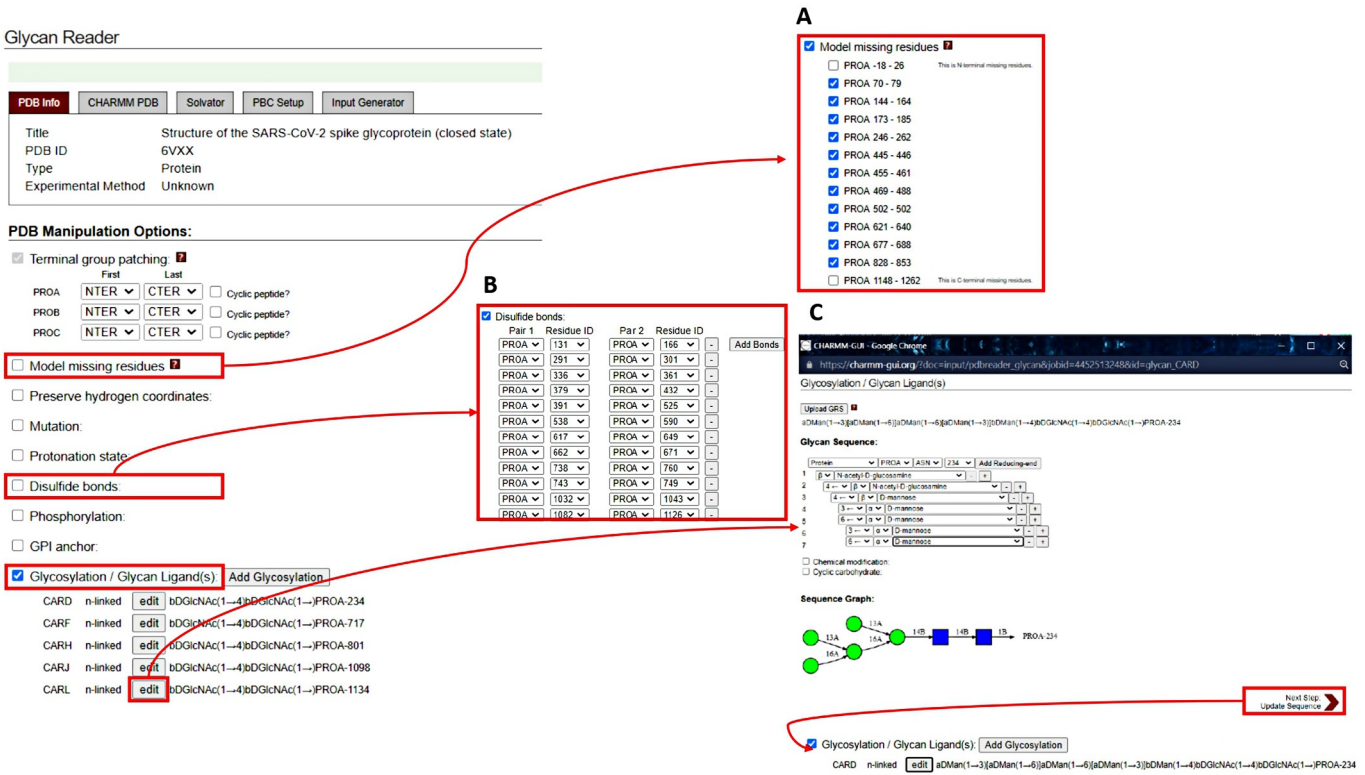

Fig. 3 Editing steps of CHARMM-GUI Glycan reader. The information and selected chains of the downloaded model are displayed and available to be edited at this moment. Preceding the simulation, multiple factors must be accounted for the glycoprotein during its modeling, such as modeling missing residues from low resolution regions of the model, for which the user can use the built-in tool to mark missing residues to model (A), checking for the presence of all structural disulfide bridges and adding missing ones (B), adding explicit hydrogens in the model and, finally, adding and editing glycans (C). Upon selecting a glycan to edit, a pop-up window emerges, allowing for precise glycan editing parameters, including sugar monomer selection and glycosidic bond position. Upon updating the sequence, the glycan code will be updated to reflect the new glycan incorporated in the model.

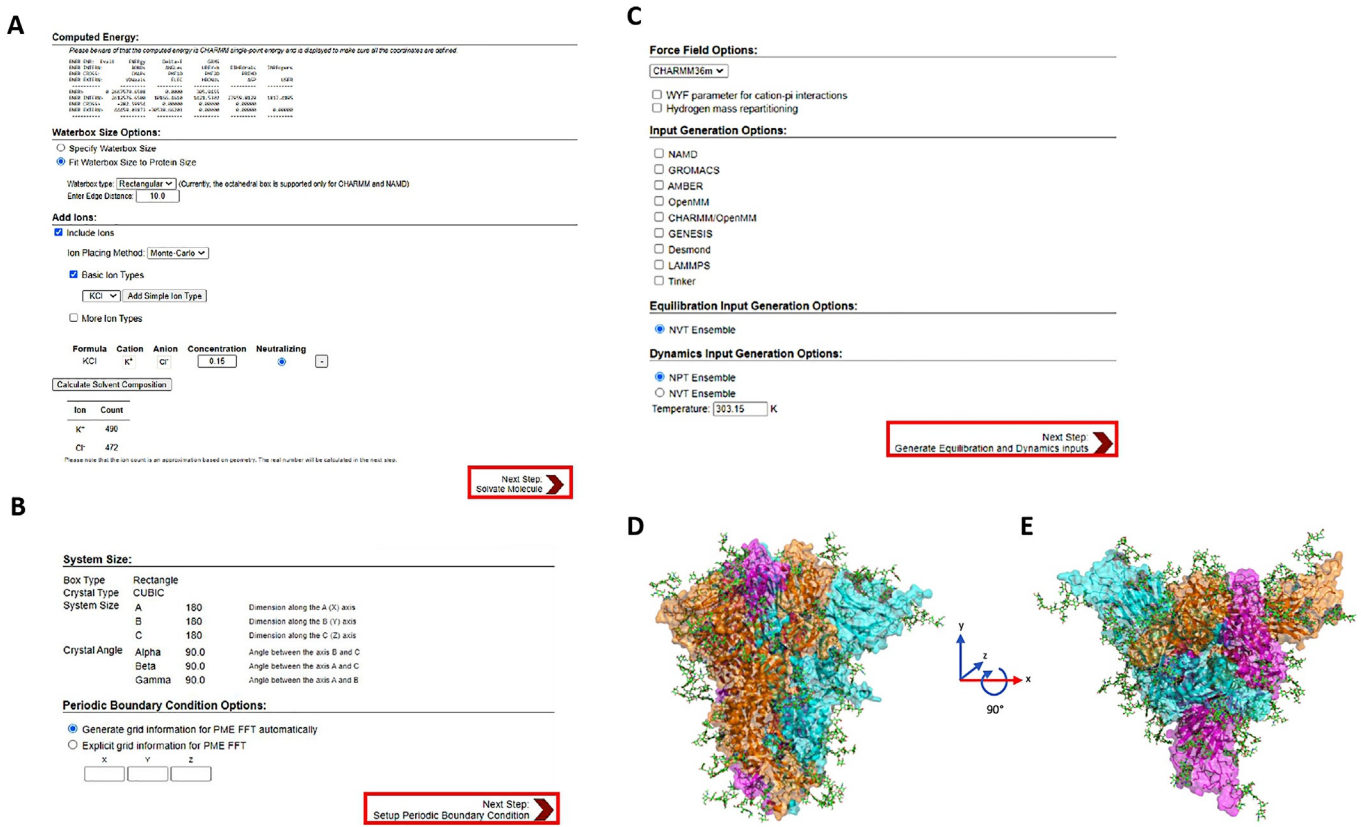

Fig. 4 See figure legend on opposite page.

The final model for MD simulation can be visualized using Pymol or other molecular visualization tools such as Chimera, Coot or VMD. The file is located inside the GROMACS folder named "step3_input.gro" (Fig. 4D and E).

2.3 MD simulation parameters

There are several MD programs available for use, such as AMBER, CHARMM and LAMMPS. In this guide, we will present the use of GROMACS software (Bauer, Hess, & Lindahl, 2022a, 2022b; Van Der Spoel et al., 2005) for our simulations. The .mdp output from CHARMM-GUI can be used as input for the simulation.

In ordsser to proceed with the MD simulation, the generated model needs to be compiled with all the topologies, restrains, run parameters, output options and pressure/temperature coupling algorithms parameters in a singular file, named the .tpr file. This compilation is done by the `gmx grommp` command as follow:

-f: Molecular dynamics parameters file (.mdp)
-o: name of the output portable binary run input file, necessary for the MD run (.tpr)
-c: structure file, the generated glycoprotein model (.gro; .pdb; .tpr; etc)
-p: topology file (.top)
-n: index file, used to set/create a new set of atoms or just using the GROMACS default groups.ndx

Fig. 4 Solvation and molecular dynamics input generation. After editing the structure for MD preparation, a solvent box with a physiological concentration of ions is added to the protein to both neutralize and add ionic strength to the system (A). The waterbox size, ion position method, salt type and concentration parameters can be selected in this section. In this step is critical to make sure that the waterbox encompass the whole protein and the glycans. The Periodic boundary conditions are set to define the "system size" during the simulation (B). Lastly, CHARMM-GUI generates the topology files using the selected "Force field" and generates inputs for an array of commonly used molecular dynamics programs, including parameter files for both equilibration and dynamics (C). The output structure of glycosylated protein with the solvent and ions occluded is presented as side view (D) and top view (E). Each chain of the protein is represented in either cyan, magenta or orange (water molecules were occluded to facilitate the visualization).

It is important to mention that the files: .mdp, .gro/.pdb, .top and others are obtained from the CHARMM-GUI output. The .mdp file contains detailed information on the parameters for MD that can be run in the default mode or adapted to best fit the simulation needs.

All of the .mdp options and their respective inputs can be checked in the GROMACS documentation (https://manual.gromacs.org/documentation/2018/user-guide/mdp-options.html). A full discussion is outside the scope of this article, so we will focus on selected parameters/options to be tailored to the simulation needs. The parameters can be divided in five categories: (a) preprocessing and run options; (b) output control; (c) neighbor search; (d) bonds and constrains and e) pressure and temperature coupling.

(a) Molecular dynamics preprocessing and run options

The preprocessing and run parameters define directories to include in the topology beyond the .itp files inside the .top file (using the -include input) and add position restrictions for user-defined atoms (using the -define input). The run options include the `integrator` algorithms for running the simulation. Among the options available for the `integrator` there are: the `md` algorithm, that uses the leap-frog method for integrating equations of motion, or the `steep` algorithm, for energy minimization using the steepest descent, which would also depend of a `emtol` input for tolerance of energy variation before declaring the system minimized. Lastly, the `nsteps` and the `dt` parameters can be set specifying the number of molecular dynamic steps to be taken in the simulation and the time step for integration in picoseconds.

(b) Molecular dynamics output files control options

The output parameters must be set in order to specify the frequency GROMACS will write information in the trajectory and log files. These parameters are set as an integral number for the number of steps that elapse in the simulation before writing: `nstlog`—energies to the log file; `nstvout`—velocities to trajectory file; `nstxcout`—coordinates to trajectory; `nstfout`—forces to trajectory; `nstcalcenergy` is used to set the number of steps between calculating energies. The `nstcalcenergy` is relevant only with `integrator` set to dynamics simulation algorithms and must be a multiple of `nstenergy`, a parameter for writing energies to the energy file.

(c) Neighbor searching, electrostatics and interatomic interactions

The interatomic interactions are accounted during the simulations by determining the interacting atomic vicinity using a list of user defined distances and other parameters. They account for both short and long distances interactions.

These parameters can be used as default or modified according to the user needs. For example, the `nstlist` is the frequency of update of the neighbor list; `rlist` sets the cutoff for short-range interactions (in nm); `rvdw` distance set to apply to van der waals interactions; `Cutoff-scheme` sets the algorithm to search for the short range interactions list, based on the `rvdw` and `rlist` values set; `vdwtype` set the algorithm for van der waals interactions search using both `rlist` and `rvdw`; `rvdw-switch` and `vdw-modifier` work in tandem to apply a smooth transition between short and vdw interations, based on the algorithm chosen for vdw modifier and the distance set in `rvdw-switch`, which must be smaller than the one set for `rvdw`. Lastly, `colombtype` sets the algorithm for electrostatics interactions and `rcoulomb` sets the distance for the Coulomb electrostatic interactions cut-off.

(d) Setting up bonds and constrains options

This class of parameters set constrains in otherwise variable components of the simulation to facilitate the simulation processing. The constrain parameters can encompass either only the bonds or bonds and angles, and can be set to affect all (`all-bonds;all-angles`), only bonds with H-atoms (`h-bonds;h-angles`) or none of them (`none`). The `constraint _algorithm` must also be selected between the `LINCS` (Hess, Bekker, Berendsen, & Fraaije, 1997) or `SHAKE` algorithm.

(e) Applying temperature and pressure coupling algorithms

These parameters define the algorithms, using both `tcoupl` and `pcoupl` to temperature and pressure respectively, and other variables for temperature and pressure equilibration during the simulation. It is important to define both `tau-t` and `tau-p` for both couplings to set the time constant for coupling (in picoseconds) and the target temperature (`ref_t` in Kelvin) and pressure (`ref_p` in bar) for the system. For this manuscript focusing on the SARS-CoV-2 SPIKE protein, the algorithms used were the `nose-hoover` (Evans & Holian, 1998) thermostat and the `Parrinello-Rahman` (Parrinello & Rahman, 1980) barostat for temperature and pressure, respectively. The selection and use of compatible ensembles for adequate monitoring of these variables is of high importance to take in account during the simulations (Hollingsworth & Dror, 2018; Hünenberger, 2005) but the usage of these ensembles is already validated and successfully used in past studies (Casarotto et al., 2021; Lupala, Ye, Chen, Su, & Liu, 2022; Mehdipour & Hummer, 2021).

Glycosides linked to proteins are often on the solvent exposed surface and are known to be involved in the formation of water mediated hydrogen

bonds (Stanca-Kaposta et al., 2008; Tachibana et al., 2004). Since the water-box generation often creates a vacuum space between solute (protein) and solvent, it is relevant to allow for additional equilibration with restraints on the glycoprotein structure. This allows the water molecules in the vicinity of the carbohydrate-protein interface to relax and reconstitute the solvation shell, therefore reducing many artificial conformational changes in the glycosides.

2.4 Molecular dynamics simulation run

With the .tpr file generated by the `gmx grommp` command for the simulation, the command `gmx mdrun` can be used to generate the .xtc file.

This command requires:

-deffnm: set the default file name for all file output options
-nt: total number of threads to start (make sure to never start more threads them your hardware as of number of cores)
-s: portable xdr input file outputted by grompp (.tpr)
-pin: tries to generate thread affinities to optimize run time of the simulation
-dlb: optimizes the ratios of threads distributed between long and short distances interactions calculations
-cpi: checkpoint file for appending to an interrupted run (.cpt)
-append: append to a previously available output file when starting from a checkpoint file

The parameters in the .mpd files must be correctly set to perform 4 different steps:

- Minimization: a step performed to minimize the energy of the system, promoting solvent interactions and structure relaxation to represent better a biological context. The critical parameters to be set in this step is that the `integrator` for the simulation must be set to `steep`, the h-bonds can be used as constrains and no temperature and pressure coupling must be implemented in this step.
- Temperature equilibration: `integrator` set to `md` and apply the `tcoupl` using the `nose-hoover` ensemble to simulate the system in a specific temperature and modulate the forces in effect in the simulation taking in account the temperature effect on them. `Tau_t` was set to update every 1 ps.

- Pressure equilibration: same parameters set for the temperature equilibration, with addition of `pcoupl=Parrinello-Rahman` to encompass the pressure effect in the system. `tau_p` was set to update every 2 ps.
- Production: upon completing the equilibration step, the system may be released of all constrains and allowed to run for a defined simulation time (in this manuscript 100 ns were used for this step), using a longer time step for integration of 2 fs.

Using this segmented relaxation of the system before production is of high relevance, as adding multiple variables in the step could lead to artifactual trajectories of protein dynamics (Walton & VanVliet, 2006).

2.5 Processing of trajectories and molecular dynamics outputs

The production step generates the .xtc and .trr files. These files must be preprocessed before the analysis. Indeed, during the simulation, the protein could present atoms jumping across the periodic boundary conditions (PBC) and that could present itself as artifacts to calculate solvent accessibility and simulation RMSD/RMSF, among other analyses. To correct this artifact, the PBC must be recentered in the simulation cell.

To perform this correction, it is needed to generate an index using it to envelop all the sugars and protein chains in the same group, so the common center of mass can be centered in the PBC later on. That requires the extraction of the first frame of the trajectory, which is extracted by using the `gmx trjconv` command:

```
-f: trajectory file (.xtc)
-s: structure file (.gro)
-o: output file name
-dump 0 To extract the first frame of the simulation
-sep to isolate the initial frame of the production step
```

Then, to build the index for all the glycans in the protein, a group was created using `gmx make_ndx -f firstframe.gro` and selecting all the glycans added to the structure during the model building.

To recenter the glycoprotein in the trajectory file in the simulation box, the `gmx trjconv` command is used twice during this step.

1st: removing all jumps of bonded atoms across the PBC, making the glycoprotein as a whole but allowing atoms to diffuse out of the PBC.

```
-pdc nojump
-n: generated .ndx file with the protein and sugars grouped together
```

2nd: set the whole glycoprotein center of mass to the center of the PBC using the nojump.tpr file and the group of glycans and protein described in the step before.

```
-pbc mol
-center
-n: generated .ndx file with the protein and sugars grouped together
```

The trajectory file with the PBC recentered in the glycoprotein can be used to check if the simulation variables agree with the input parameters. To perform this, the command `gmx energy` was used together with the .edr file as input to plot:

- Potential energy: in all steps of the simulation to check if the energy minimization was successful and if the energy in the system kept itself regular through the equilibrations and productions.
- Temperature: in the equilibrations and production steps to check if the system was correctly kept at the inputted temperature of `ref_t`.
- Pressure and density: in the pressure equilibration and production steps to check it the system was accurately equilibrated at the inputted pressure of `ref_t` and if the density of the system followed that trend.

The output plots for the energy, temperature, pressure and density for the SPIKE glycoproteins analyzed in this manuscript are reported in Fig. 5.

3. Trajectory file analysis (RMSD/RMSF/H-bridges)

The recentered trajectory file (.xtc or .trr) was analyzed using the VMD v.1.9.4. (Humphrey, Dalke, & Schulten, 1996) to perform trajectory visualization and other analyses to evaluate the behavior and interactions of the different segments of the glycoproteins along the simulation.

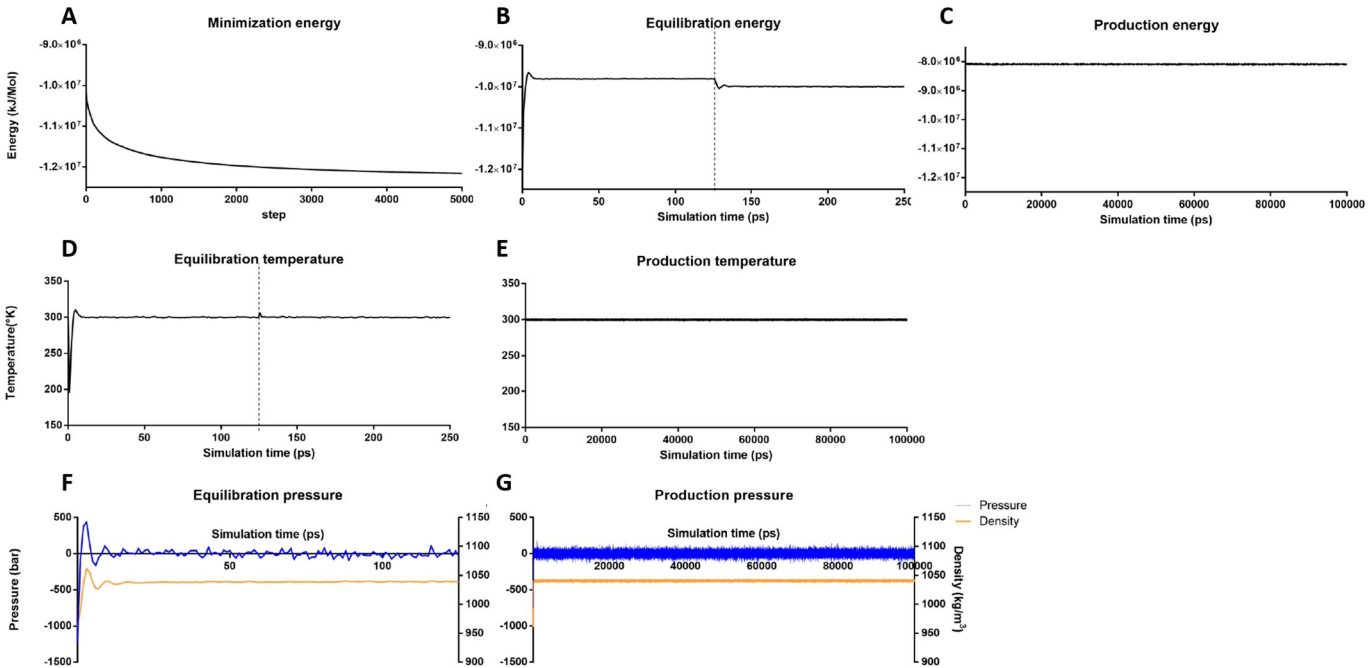

Fig. 5 output of the `gmx energy` for energy minimization (A), constant number volume and temperature (NVT) and constant number pressure and temperature (NPT) equilibration (first and second part of B, respectively), and production steps (C), for temperature during the NVT and NPT equilibration (D) and production steps (E), and pressure and density for the NPT (F) and production steps (G).

To load the trajectory file into the VMD software, a "new molecule" must be created and add the trajectory file. The molecule file browser (File > new molecule) is used to open the firstframe.gro and then loading the .xtc file into the molecule (Fig. 6A and B). The VMD will start loading frame by frame into the molecule. To improve the observation of the trajectory (representation > graphics…), we chose to hide the waters and ions of the representation, change the depiction of the protein to NewCartoon, and apply a 5 frames trajectory smoothing window to observe the trajectory (Fig. 6C–J).

To perform the calculations of RMSF/RMSD and search for H-bonds we used some command lines in the Tk console of VMD (Extensions">"Tk console) as described below.

```
##RMSD/RMSF setup
set start [atomselect top "protein and backbone" frame 0]
set current [atomselect top "protein and backbone"]
set num_steps [molinfo top get numframes]
set carbA [atomselect top "name CA"]
##RMSD
set outfile [open rmsd.dat w]
for {set frame 0} {$frame < $num_steps} {incr frame} {
    $current frame $frame
    set trans_mat [measure fit $current $start]
    $current move $trans_mat
    set rmsd [measure rmsd $current $start ]
  puts $outfile "$frame $rmsd"
}
close $outfile
##RMSF
set outfile [open rmsf.dat w]
set sel [atomselect top "name CA"]
set rmsf [measure rmsf $sel first 0 last [expr
{$num_steps - 1}] step 1]
for {set i 0} {$i < [$sel [expr {$num_steps - 1}]]} {incr i} {
  puts $outfile "[expr {$i+1}] [lindex $rmsf $i]"
}
close $outfile
## H-bonds
package require hbonds
hbonds -sel1 [atomselect top protein] -sel2 [atomselect top
"all not water and not ion and not protein"] -writefile
yes -plot yes
```

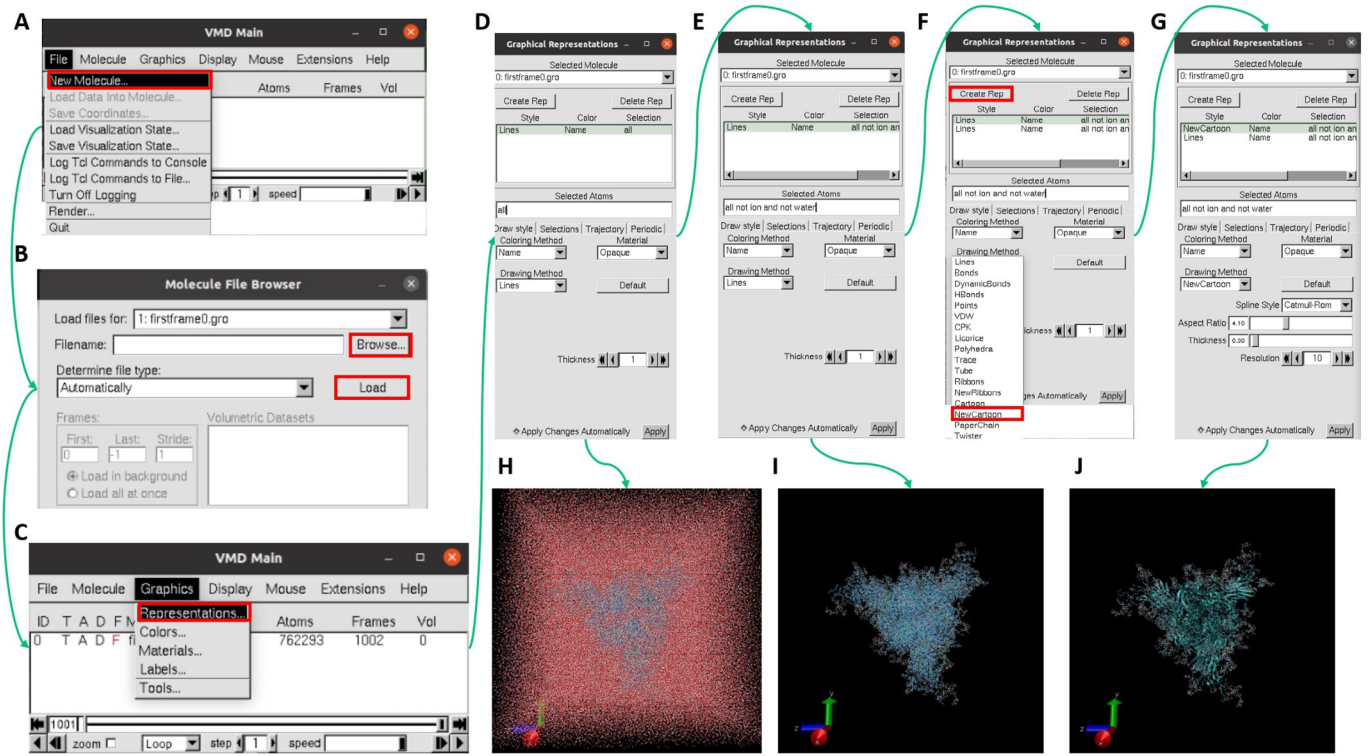

Fig. 6 Analysis and visualization workflow using the VMD software. (A and B) Load the molecule and its trajectory into VMD. In VMD main, go to File - > New molecule (A). Load the first frame of the production simulation, browsing for the .gro file, then load the .xtc file into the molecule (B). (C–G) Define the representation of the molecule. Open the Graphical Representation panel by selecting Graphics → Representations… (C). In this panel (D), change the "selected atoms" panel to represent only the glycoprotein, by excluding the water atoms and the ions from the representation (E). Create a new representation, select only the protein atoms and change the "Drawing method" to "New Cartoon" (F). The final setup in the Graphical Representation panel is showed in (G). (H) All atom representation. (I) glycoprotein representation. (J) Glycoprotein representation using NewCartoon drawing method for the protein.

This will generate 3 output files:

hbonds-details.dat, referring to each residue that interacted with a glycan during the simulation and that interaction occupancy, Fig. 7.

RMSD.txt, which outputs a comparison of a frame of the simulation with the first frame, Fig. 8A;

RMSF.txt, which represents the average fluctuations of each aminoacid in the model during the simulations, Fig. 8B.

4. SASA/AbASA calculations and analysis

To perform the SASA and AbASA measurements, the `gmx sasa` command was used within the GROMACS software (Eisenhaber, Lijnzaad, Argos, Sander, & Scharf, 1995):

```
-f: the recentered trajectory file
-s: the protein coordinates file
-n: the generated index with the whole glycoprotein group (.ndx)
-o: the output file name for total exposed area (.xvg)
-or: the output file name for exposed area for each aminoacid residue (.xvg)
-probe: radius of the solvent in nm (0.14 for water; 0.7 for Ab)
```

The selected probe radius for Ab was based on the estimated size of the hypervariable loop performed by Grant, Montgomery, Ito, & Woods, 2020. The `gmx sasa` command performs a double cubic lattice method for surface area calculation and uses as input the group of interest to have its area measured and which group to be in the results. For the first two iterations of this command, the surface selection should be the whole glycoprotein group defined in the index file, and for the output the group selection that implies only the protein. This will generate a resarea.xvg file with the average GlySASA (for the probe radius of 0.14) and GlyAbASA (for the probe radius of 0.7) for each amino acid residue considering the presence of the oligosaccharides. Following this, a second calculation with the same inputs should be performed, using only the protein for both surface calculation and output writing SASA (for the probe radius of 0.14) and AbASA (for the probe radius of 0.7). This step is performed to evaluate the effect of site-specific glycosylation on solvent and antibody accessibility. Based on the

A

Donor	Acceptor	Occupancy
BGLCN1-Main	ILE1132-Main	93.11%
AFUC13-Side	GLU132-Side	66.17%
BGLCN1-Main	PRO579-Main	52.79%
BGLCN1-Side	ILE235-Main	51.30%
LYS129-Side	BGLCN2-Main	49.10%
BGLCN1-Main	ASN125-Side	48.50%
BGLCN1-Main	ASN710-Side	46.71%
BGLCN2-Main	TYR160-Side	42.22%
BGAL12-Side	GLU484-Side	42.12%
AMAN5-Side	GLU702-Side	40.12%
TYR351-Side	BGLCN2-Main	39.32%
BGLCN1-Main	THR1100-Side	38.22%
AMAN6-Side	GLU169-Side	37.03%
TYR449-Main	ANE5A7-Side	34.53%
GLN895-Side	BGLCN1-Main	34.33%
ALA163-Main	AMAN4-Side	34.23%
AMAN4-Side	ASN164-Side	33.73%
BGLCN5-Side	LEU492-Main	32.24%
TYR1110-Side	AFUC10-Side	29.54%
AMAN5-Side	SER161-Main	29.34%
BMAN3-Side	GLU583-Side	26.65%
THR236-Side	BGLCN1-Side	26.55%
SER494-Main	BGLCN5-Side	25.15%
BGLCN5-Side	GLU918-Side	24.35%
AMAN6-Side	ASP88-Side	22.75%
BGLCN1-Main	GLU281-Side	22.36%
AMAN5-Side	ASP88-Side	22.26%
AFUC9-Side	ASN556-Main	21.66%
BGLCN1-Main	GLN580-Main	21.36%
AMAN4-Side	GLU702-Main	21.06%
TYR160-Side	BGLCN1-Side	20.86%
ASN17-Main	AFUC10-Side	20.56%
BGLCN1-Main	HSD655-Main	20.36%

B

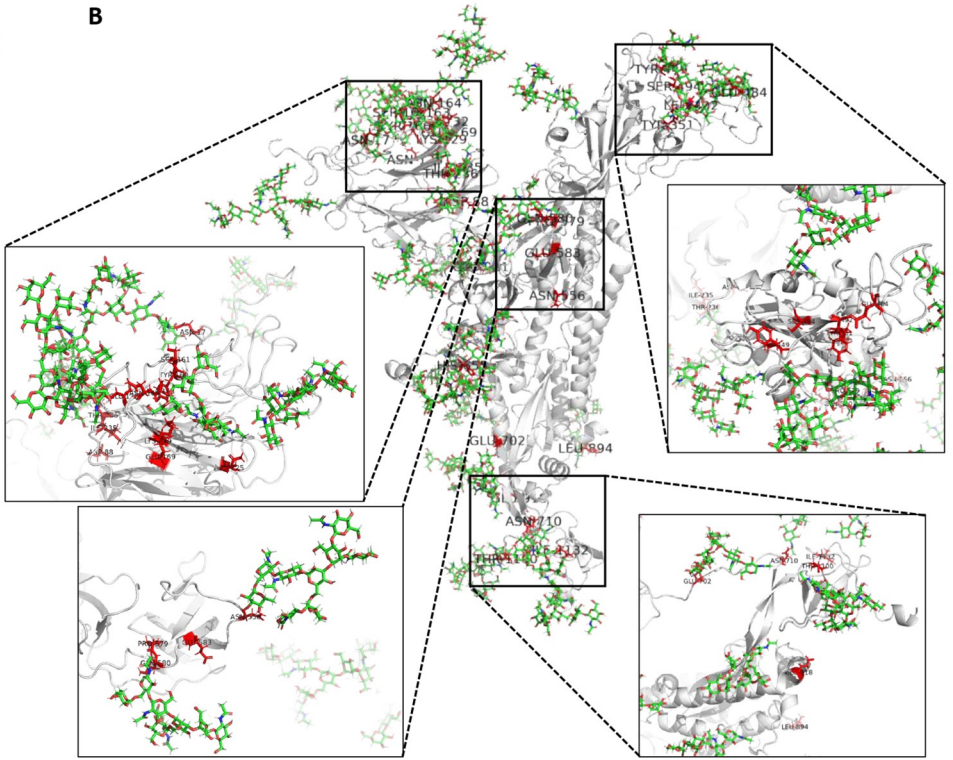

Fig. 7 Analysis of H-bridges content. (A) Relation of hydrogen bonds with above 20% occupancy and involving a glycan chain with a protein chain found in the simulation. BGLCN: *N*-acetyl-D-glucosamine; AFUC: L-fucose; A/BMAN: D-Mannose. (B) Graphical representation of a monomer of the SPIKE protein with its glycans. Aminoacid residues that presented a high occupancy H-bridge with glycans where highlighted in red and labeled using Pymol.

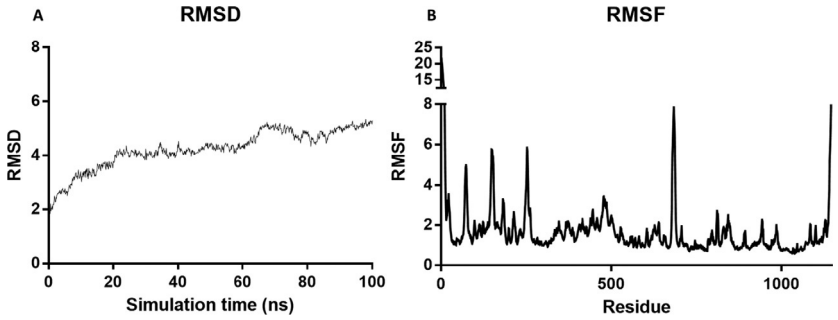

Fig. 8 Calculation of the root mean square deviation (A) during the simulation and root mean square fluctuation (B) per residue.

ratios of GlyAbASA/AbASA for each residue, the glycan-dependent occlusion for the whole protein can be determined for each residue. Also, using the ratios of SASA/Max.SASA for each residue, being Max.SASA defined as the maximum available area per residue calculated by Tien, Meyer, Sydykova, Spielman, and Wilke (2013), it is possible to discern if a residue is buried into the structure (SASA/Max.SASA < 0.15). Comparing the GlyAbASA/AbASA in a specific region against the whole glycoprotein and removing from the analysis residues that are buried based on the SASA/Max.SASA < 0.15, it is possible to calculate which residues are statistically occluded (Fig. 9). Other analysis can be derived from the output trajectories using this methodology, such as salt bridges, that could indicate protein stabilization, or radius of gyration, that is related to protein compactness, but they were not the scope of this project.

There is extensive literature discussing the relevance of the simulation length and sampling, which must be decided based on the hypothesis (Henzler-Wildman & Kern, 2007). Convergence in the carbohydrate structure conformations has been discussed as one of the main issues and, phenomena such as ring puckering were shown to require microsecond length simulations and cannot be addressed by nanoscecond length simulations (Sattelle, Hansen, Gardiner, & Almond, 2010). That being said, there are papers in the literature that successfully reported interesting findings by studying glycoprotein molecular dynamics trajectory files with the nanosecond timescale (Yokoyama et al., 2017; Bernardi, Kirschner, & Faller, 2017). In the latter timescale, it is not possible to explore all the glycan conformations that could be populated in the system; however, glycan-protein interactions can still be assessed by this methodology.

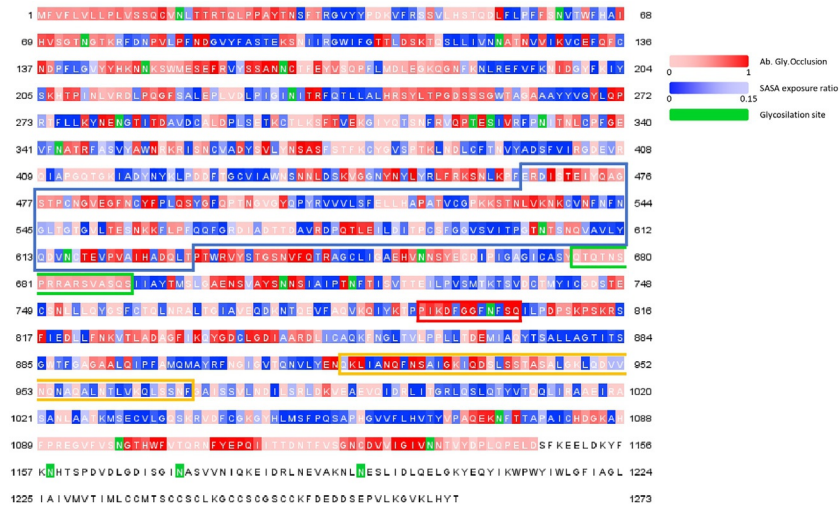

Fig. 9 Representation of Ab accessibility for each aminoacid residue. Residues in red indicate antibody occlusion dependent on the glycan chains. Residues in blue indicate solvent exposure in the protein without considering the glycan chains. Residues in green indicate asparagine with N-linked glycosylation. Blue box: RBD (465–630). Green box: Furin cleavage loop (675–692). Red box: Modulated region (792–804) with a significant shift toward higher glycan occlusion ($P = 0.0003$). Orange box: Heptad repeat 1 (920–970).

5. Results and discussion

The applied methodology was already performed in other MD simulations (Barros et al., 2021; Liguori, Croce, Marrink, & Thallmair, 2020; Mehdipour & Hummer, 2021; Turoňová et al., 2020). The analysis of the simulation energy files (Fig. 4) showed that the simulation correctly followed the defined parameters and the plot in JalView 2.11.1.7 (Waterhouse, Procter, Martin, Clamp, & Barton, 2009) allowed to evaluate and quantify the differential glycan shield protection in solvent-exposed residues (Fig. 9), such as the modulated regions in 792–804. We were also able to identify a large array of protein-glycan hydrogen bonds and assess their interactions inside the structure (Figs. 7 and 9). This methodology can be applied to other glycoproteins to evaluate a large array of effects that the glycosylation might impose to the structure and accessibility of the protein in question, such as evaluating glycan promoted structure stability, glycan occlusion from solvent or antibody and ligand interaction.

6. Conclusions

The growing interest in protein glycosylation and structure to function relationship calls for reliable analytical and computational tools. MD simulation of glycoproteins can help in elucidating the role of oligosaccharides on protein structure, antibody binding, receptor interaction and function. In infectious diseases, glycoproteins can serve as chemotherapeutic targets and biomarkers giving a deeper understanding on the biological mechanisms of host-pathogen interaction. In this step-by-step guide, we provide a detailed view of the different computational platforms that can be used to perform MD simulation of glycoproteins focusing on the SARS-CoV-2 SPIKE glycoprotein. We believe that this guide will serve as a platform to start MD simulations of glycoproteins in several infectious diseases and help in improving the tools currently available.

Acknowledgments

We are grateful for the financial support provided by the São Paulo Research Foundation (FAPESP, grants processes no 2018/18257-1 (GP), 2018/15549-1 and 2020/04923-0 (GP), 2021/00140-3 (JMDS), 2015/26722-8 (CW), 2021/14751-4 (SNM)); by the Conselho Nacional de Desenvolvimento Científico e Tecnológico ("Bolsa de Produtividade" (307854/2018-3 to GP); by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES bolsa PNPD 88887.372048/2019-00 to LRF and to JVPC).

References

Aebi, M. (2013). N-linked protein glycosylation in the ER. *Biochimica et Biophysica Acta*, *1833*, 2430–2437.

Alocci, D., Mariethoz, J., Gastaldello, A., Gasteiger, E., Karlsson, N. G., Kolarich, D., et al. (2019). GlyConnect: Glycoproteomics goes visual, interactive, and analytical. *Journal of Proteome Research*, *18*, 664–677. https://doi.org/10.1021/acs.jproteome.8b00766.

Baloch, S., Baloch, M. A., Zheng, T., & Pei, X. (2020). The coronavirus disease 2019 (COVID-19) pandemic. *The Tohoku Journal of Experimental Medicine*, *250*, 271–278. https://doi.org/10.1620/tjem.250.271.

Banerjee, N., & Mukhopadhyay, S. (2016). Viral glycoproteins: biological role and application in diagnosis. *Virusdisease*, *27*, 1–11. https://doi.org/10.1007/s13337-015-0293-5.

Barros, E. P., Casalino, L., Gaieb, Z., Dommer, A. C., Wang, Y., Fallon, L., et al. (2021). The flexibility of ACE2 in the context of SARS-CoV-2 infection. *Biophysical Journal*, *120*, 1072–1084. https://doi.org/10.1016/J.BPJ.2020.10.036.

Bauer, P., Hess, B., & Lindahl, E. (2022a). *GROMACS 2022 source code*. https://doi.org/10.5281/ZENODO.6103835.

Bauer, P., Hess, B., & Lindahl, E. (2022b). *GROMACS 2022 manual*. https://doi.org/10.5281/ZENODO.6103568.

Bernardi, A., Kirschner, K. N., & Faller, R. (2017). Structural analysis of human glycoprotein butyrylcholinesterase using atomistic molecular dynamics: The importance

of glycosylation site ASN241. *PLoS One*, *12*, e0187994. https://doi.org/10.1371/JOURNAL.PONE.0187994.

Campbell, M. P. (2017). A review of software applications and databases for the interpretation of glycopeptide data. *Trends in Glycoscience and Glycotechnology*, *29*, E51–E62. https://doi.org/10.4052/tigg.1601.1E.

Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., et al. (2020). Beyond shielding: The roles of glycans in the SARS-CoV-2 spike protein. *ACS Central Science*, *6*, 1722–1734. https://doi.org/10.1021/acscentsci.0c01056.

Casarotto, P. C., Girych, M., Fred, S. M., Kovaleva, V., Moliner, R., Enkavi, G., et al. (2021). Antidepressant drugs act by directly binding to TRKB neurotrophin receptors. *Cell*, *184*, 1299–1313. e19 https://doi.org/10.1016/J.CELL.2021.01.034.

Chang, V. T., Crispin, M., Aricescu, A. R., Harvey, D. J., Nettleship, J. E., Fennelly, J. A., et al. (2007). Glycoprotein structural genomics: Solving the glycosylation problem. *Structure*, *15*, 267–273. https://doi.org/10.1016/j.str.2007.01.011.

Cheong, Y., Shim, G., Kang, D., & Kim, Y. (1999). Carbohydrate binding specificity of pea lectin studied by NMR spectroscopy and molecular dynamics simulations. *Journal of Molecular Structure*, *475*, 219–232. https://doi.org/10.1016/S0022-2860(98)00511-0.

Chernykh, A., Kawahara, R., & Thaysen-Andersen, M. (2021). Towards structure-focused glycoproteomics. *Biochemical Society Transactions*, *49*, 161–186. https://doi.org/10.1042/BST20200222.

Cook, J. D., & Lee, J. E. (2013). The secret life of viral entry glycoproteins: Moonlighting in immune evasion. *PLoS Pathogens*, *9*, e1003258. https://doi.org/10.1371/journal.ppat.1003258.

Dalziel, M., Crispin, M., Scanlan, C. N., Zitzmann, N., & Dwek, R. A. (2014). Emerging principles for the therapeutic exploitation of glycosylation. *Science (New York, N.Y.)*, *343*. https://doi.org/10.1126/SCIENCE.1235681.

Davis, S. J., & Crispin, M. (2010). Solutions to the glycosylation problem for low- and high-throughput structural glycoproteomics. In R. Owens, & J. Nettleship (Eds.), *Functional and structural proteomics of glycoproteins* (pp. 127–158). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-9355-4_6.

Duan, L., Zheng, Q., Zhang, H., Niu, Y., Lou, Y., & Wang, H. (2020). The SARS-CoV-2 spike glycoprotein biosynthesis, structure, function, and antigenicity: Implications for the design of spike-based vaccine immunogens. *Frontiers in Immunology*, *11*, 2593. https://doi.org/10.3389/FIMMU.2020.576622/BIBTEX.

Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C., & Scharf, M. (1995). The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry*, *16*, 273–284. https://doi.org/10.1002/JCC.540160303.

Evans, D. J., & Holian, B. L. (1998). The Nose–Hoover thermostat. *The Journal of Chemical Physics*, *83*, 4069. https://doi.org/10.1063/1.449071.

Grant, O. C., Montgomery, D., Ito, K., & Woods, R. J. (2020). Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Scientific Reports*, *2020*, 1–11. https://doi.org/10.1038/s41598-020-71748-7. 10:1 10.

Hacisuleyman, E., Hale, C., Saito, Y., Blachere, N. E., Bergh, M., Conlon, E. G., et al. (2021). Vaccine breakthrough infections with SARS-CoV-2 variants. *The New England Journal of Medicine*, *384*, 2212–2218. https://doi.org/10.1056/NEJMoa2105000.

Hanisch, F. G. (2001). O-glycosylation of the mucin type. *Biological Chemistry*, *382*, 143–149. https://doi.org/10.1515/BC.2001.022/MACHINEREADABLECITATION/RIS.

Hargett, A. A., Marcella, A. M., Yu, H., Li, C., Orwenyo, J., Battistel, M. D., et al. (2021). Glycosylation states on intact proteins determined by NMR spectroscopy. *Molecules*, *26*, 4308. https://doi.org/10.3390/molecules26144308.

Hastie, K. M., Zandonatti, M. A., Kleinfelter, L. M., Heinrich, M. L., Rowland, M. M., Chandran, K., et al. (2017). Structural basis for antibody-mediated neutralization of Lassa virus. *Science*, *356*, 923–928. https://doi.org/10.1126/SCIENCE.AAM7260.

Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. *Nature*, *450*(2007), 964–972. 4507172 https://doi.org/10.1038/nature06522.

Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, *18*, 14631472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12.

Hollingsworth, S. A., & Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron*, *99*, 1129–1143. https://doi.org/10.1016/J.NEURON.2018.08.011.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., et al. (2016). CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods*, *14*(1), 71–73. https://doi.org/10.1038/nmeth.4067.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*, 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

Hünenberger, P. H. (2005). Thermostat algorithms for molecular dynamics simulations. *Advances in Polymer Science*, *173*, 105–147. https://doi.org/10.1007/B99427.

Jo, S., Cheng, X., Lee, J., Kim, S., Park, S. J., Patel, D. S., et al. (2017). CHARMM-GUI 10 years for biomolecular modeling and simulation. *Journal of Computational Chemistry*, *38*, 1114–1124. https://doi.org/10.1002/JCC.24660.

Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, *29*, 1859–1865.

Jo, S., Song, K. C., Desaire, H., MacKerell, A. D., & Im, W. (2011). Glycan reader: Automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *Journal of Computational Chemistry*, *32*, 3135–3141. https://doi.org/10.1002/JCC.21886.

Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., et al. (2003). Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nature Biotechnology*, *21*, 667–672. https://doi.org/10.1038/nbt829.

Kang, P., Mechref, Y., & Novotny, M. V. (2008). High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid Communications in Mass Spectrometry*, *22*, 721–734. https://doi.org/10.1002/rcm.3395.

Kawahara, R., Chernykh, A., Alagesan, K., Bern, M., Cao, W., Chalkley, R. J., et al. (2021). Community evaluation of glycoproteomics informatics solutions reveals high-performance search strategies for serum glycopeptide analysis. *Nature Methods*, *18*, 1304–1316. https://doi.org/10.1038/s41592-021-01309-x.

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., et al. (2008). GLYCAM06: A generalizable biomolecular force field. *Carbohydrates*. *Journal of Computational Chemistry*, *29*, 622–655. https://doi.org/10.1002/JCC.20820.

Kwon, Y. D., Pancera, M., Priyamvada, A., Georgiev, I. S., Crooks, E. T., et al., & Kwong, P. D. (2015). Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1. *Nature Structural & Molecular Biology*, *22*, 522–531.

Larsen, M. R., Jensen, S. S., Jakobsen, L. A., & Heegaard, N. H. H. (2007). Exploring the Sialiome using titanium dioxide chromatography and mass spectrometry. *Molecular & Cellular Proteomics*, *6*, 1778–1787. https://doi.org/10.1074/mcp.M700086-MCP200.

Lasky, L. A. (1991). Lectin cell adhesion molecules (LEC-CAMs): A new family of cell adhesion proteins involved with inflammation. *Journal of Cellular Biochemistry*, *45*, 139–146. https://doi.org/10.1002/JCB.240450204.

Lee, P., Ohshima, N., Stanfield, R., Yu, W., Okuno, Y., Kurosawa, Y., et al. (2014). Receptor mimicry by antibody F045–092 facilitates universal binding to the H3 subtype of influenza virus. *Nature Communications*, *5*, 3614.

Lemieux, R. U., & Koto, S. (1974). The conformational properties of glycosidic linkages. *Tetrahedron*, *30*, 1933–1944. https://doi.org/10.1016/S0040-4020(01)97324-7.

Liguori, N., Croce, R., Marrink, S. J., & Thallmair, S. (2020). Molecular dynamics simulations in photosynthesis. *Photosynthesis Research*, *144*(2), 273–295. https://doi.org/10.1007/S11120-020-00741-Y.

Lippi, G., Mattiuzzi, C., & Henry, B. M. (2022). Updated picture of SARS-CoV-2 variants and mutations. *Diagnosis*, *9*, 11–17. https://doi.org/10.1515/DX-2021-0149/PDF.

Lis, H., & Sharon, N. (1993). Protein glycosylation. Structural and functional aspects. *European Journal of Biochemistry*, *218*, 1–27. https://doi.org/10.1111/j.1432-1033.1993.tb18347.x.

Liu, H., Wei, P., Zhang, Q., Chen, Z., Aviszus, K., Downing, W., et al. (2021). 501Y.V2 and 501Y.V3 variants of SARS-CoV-2 lose binding to bamlanivimab in vitro. *bioRxiv*, *13*. https://doi.org/10.1080/19420862.2021.1919285.

Liu, H., Zhang, Q., Wei, P., Chen, Z., Aviszus, K., Yang, J., & Zhang, G. (2021). The basis of a more contagious 501Y.V1 variant of SARS-COV-2. *Cell Res*, *31*, 720–722. https://doi.org/10.1101/2021.02.02.428884.

Lupala, C. S., Ye, Y., Chen, H., Su, X. D., & Liu, H. (2022). Mutations on RBD of SARS-CoV-2 omicron variant result in stronger binding to human ACE2 receptor. *Biochemical and Biophysical Research Communications*, *590*, 34–41. https://doi.org/10.1016/J.BBRC.2021.12.079.

Macedo-da-Silva, J., Santiago, V. F., Rosa-Fernandes, L., Marinho, C. R. F., & Palmisano, G. (2021). Protein glycosylation in extracellular vesicles: Structural characterization and biological functions. *Molecular Immunology*, *135*, 226–246. https://doi.org/10.1016/j.molimm.2021.04.017.

Machiels, B., Lété, C., Guillaume, A., Mast, J., Stevenson, P. G., Vanderplasschen, A., et al. (2011). Antibody evasion by a Gammaherpesvirus O-glycan shield. *PLoS Pathogens*, *7*, e1002387. https://doi.org/10.1371/JOURNAL.PPAT.1002387.

Mehdipour, A. R., & Hummer, G. (2021). Dual nature of human ACE2 glycosylation in binding to SARS-CoV-2 spike. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, 2022.

Meng, B., Kemp, S. A., Papa, G., Datir, R., Ferreira, I. A. T. M., Marelli, S., et al. (2021). Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Reports*, *35*, 109292. https://doi.org/10.1016/J.CELREP.2021.109292.

Mohan, G. S., Li, W., Ye, L., Compans, R. W., & Yang, C. (2012). Antigenic subversion: A novel mechanism of host immune evasion by Ebola virus. *PLoS Pathogens*, *8*, e1003065. https://doi.org/10.1371/JOURNAL.PPAT.1003065.

Morelle, W., & Michalski, J.-C. (2007). Analysis of protein glycosylation by mass spectrometry. *Nature Protocols*, *2*, 1585–1602. https://doi.org/10.1038/nprot.2007.227.

Mule, S. N., Rosa-Fernandes, L., Coutinho, J. V. P., Gomes, V. D. M., Macedo-da-Silva, J., Santiago, V. F., et al. (2021). Systems-wide analysis of glycoprotein conformational changes by limited deglycosylation assay. *Journal of Proteomics*, *248*, 104355. https://doi.org/10.1016/j.jprot.2021.104355.

Nagae, M., & Yamaguchi, Y. (2012). Function and 3D structure of the N-glycans on glycoproteins. *International Journal of Molecular Sciences*, *13*, 8398–8429. https://doi.org/10.3390/ijms13078398.

Naveca, F. G., Nascimento, V., de Souza, V. C., de Corado, A. L., Nascimento, F., Silva, G., et al. (2021). COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. *Nature Medicine*, *27*, 1230–1238. https://doi.org/10.1038/s41591-021-01378-7.

Oliveira, T., Thaysen-Andersen, M., Packer, N. H., & Kolarich, D. (2021). The Hitchhiker's guide to glycoproteomics. *Biochemical Society Transactions*, *49*, 1643–1662. https://doi.org/10.1042/BST20200879.

Ongay, S., Boichenko, A., Govorukhina, N., & Bischoff, R. (2012). Glycopeptide enrichment and separation for protein glycosylation analysis: Sample preparation. *Journal of Separation Science*, *35*, 2341–2372. https://doi.org/10.1002/jssc.201200434.

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., et al. (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature Communications*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-15562-9.

Palmisano, G., Larsen, M. R., Packer, N. H., & Thaysen-Andersen, M. (2013). Structural analysis of glycoprotein sialylation—Part II: LC-MS based detection. *RSC Advances*, *3*, 22706. https://doi.org/10.1039/c3ra42969e.

Palmisano, G., Lendal, S. E., Engholm-Keller, K., Leth-Larsen, R., Parker, B. L., & Larsen, M. R. (2010). Selective enrichment of sialic acid–containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nature Protocols*, *5*, 1974–1982. https://doi.org/10.1038/nprot.2010.167.

Palmisano, G., Melo-Braga, M. N., Engholm-Keller, K., Parker, B. L., & Larsen, M. R. (2012). Chemical deamidation: A common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *Journal of Proteome Research*, *11*, 1949–1957. https://doi.org/10.1021/pr2011268.

Park, S. J., Lee, J., Patel, D. S., Ma, H., Lee, H. S., Jo, S., et al. (2017). Glycan reader is improved to recognize most sugar types and chemical modifications in the protein data bank. *Bioinformatics*, *33*, 3051–3057. https://doi.org/10.1093/BIOINFORMATICS/BTX358.

Park, S.-J., Lee, J., Qi, Y., Kern, N. R., Lee, H. S., Jo, S., et al. (2019). CHARMM-GUI *glycan modeler* for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology*, *29*, 320–331. https://doi.org/10.1093/glycob/cwz003.

Parker, B. L., Gupta, P., Cordwell, S. J., Larsen, M. R., & Palmisano, G. (2011). Purification and identification of *O*-Glc *N* Ac-modified peptides using phosphate-based alkyne CLICK chemistry in combination with titanium dioxide chromatography and mass spectrometry. *Journal of Proteome Research*, *10*, 1449–1458. https://doi.org/10.1021/pr100565j.

Parrinello, M., & Rahman, A. (1980). Crystal structure and pair potentials: A molecular-dynamics study. *Physical Review Letters*, *45*, 1196. https://doi.org/10.1103/PhysRevLett.45.1196.

Pascarella, S., Ciccozzi, M., Zella, D., Bianchi, M., Benedetti, F., Benvenuto, D., et al. (2021). SARS-CoV-2 B.1.617 Indian variants: Are electrostatic potential changes responsible for a higher transmission rate? *Journal of Medical Virology*, *93*, 6551–6556. https://doi.org/10.1002/JMV.27210.

Pasing, Y., Sickmann, A., & Lewandrowski, U. (2012). N-glycoproteomics: Mass spectrometry-based glycosylation site annotation. *Biological Chemistry*, *393*, 249–258. https://doi.org/10.1515/HSZ-2011-0245/MACHINEREADABLECITATION/RIS.

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., et al. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, *91*, 1–41. https://doi.org/10.1016/0010-4655(95)00041-D.

Peters, T., Meyer, B., Stuike-Prill, R., Somorjai, R., & Brisson, J. R. (1993). A Monte Carlo method for conformational analysis of saccharides. *Carbohydrate Research*, *238*, 49–73. https://doi.org/10.1016/0008-6215(93)87005-D.

Planas, D., Veyer, D., Baidaliuk, A., Staropoli, I., Guivel-Benhassine, F., Rajah, M. M., et al. (2021). Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature*, *596*, 276–280. https://doi.org/10.1038/S41586-021-03777-9.

Reiding, K. R., Blank, D., Kuijper, D. M., Deelder, A. M., & Wuhrer, M. (2014). High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Analytical Chemistry*, *86*, 5784–5793. https://doi.org/10.1021/ac500335t.

Reily, C., Stewart, T. J., Renfrow, M. B., & Novak, J. (2019). Glycosylation in health and disease. *Nature Reviews. Nephrology*, *15*, 346–366. https://doi.org/10.1038/s41581-019-0129-4.

Rodriguez, R., Chinea, G., Lopez, N., Pons, T., & Vriend, G. (1998). Homology modeling, model and software evaluation: Three related resources. *Bioinformatics*, *14*, 523–528. https://doi.org/10.1093/BIOINFORMATICS/14.6.523.

Sattelle, B. M., Hansen, S. U., Gardiner, J., & Almond, A. (2010). Free energy landscapes of iduronic acid and related monosaccharides. *Journal of the American Chemical Society*, *132*, 13132–13134. https://doi.org/10.1021/JA1054143/SUPPL_FILE/JA1054143_SI_001.PDF.

Schindler, B., Barnes, L., Renois, G., Gray, C., Chambert, S., Fort, S., et al. (2017). Anomeric memory of the glycosidic bond upon fragmentation and its consequences for carbohydrate sequencing. *Nature Communications*, *8*, 973. https://doi.org/10.1038/s41467-017-01179-y.

Schjoldager, K. T., Narimatsu, Y., Joshi, H. J., & Clausen, H. (2020). Global view of human protein glycosylation pathways and functions. *Nature Reviews. Molecular Cell Biology*, *21*, 729–749. https://doi.org/10.1038/s41580-020-00294-x.

Stanca-Kaposta, E. C., Gamblin, D. P., Cocinero, E. J., Frey, J., Kroemer, R. T., Fairbanks, A. J., et al. (2008). Solvent interactions and conformational choice in a core N-glycan segment: Gas phase conformation of the central, branching trimannose unit and its singly hydrated complex. *Journal of the American Chemical Society*, *130*, 10691–10696. https://doi.org/10.1021/JA801892H/SUPPL_FILE/JA801892H-FILE010.PDF.

Struwe, W. B., & Robinson, C. V. (2019). Relating glycoprotein structural heterogeneity to function—Insights from native mass spectrometry. *Current Opinion in Structural Biology*, *58*, 241–248. https://doi.org/10.1016/j.sbi.2019.05.019.

Sugrue, R. J. (2007). Viruses and glycosylation. In R. J. Sugrue (Ed.), *Glycovirology protocols. Methods in molecular biology* (pp. 1–13). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-59745-393-6_1.

Tachibana, Y., Fletcher, G. L., Fujitani, N., Tsuda, S., Monde, K., & Nishimura, S. I. (2004). Antifreeze glycoproteins: Elucidation of the structural motifs that are essential for antifreeze activity. *Angewandte Chemie, International Edition*, *43*, 856–862. https://doi.org/10.1002/ANIE.200353110.

Thaysen-Andersen, M., & Packer, N. H. (2014). Advances in LC–MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1844*, 1437–1452. https://doi.org/10.1016/j.bbapap.2014.05.002.

Thaysen-Andersen, M., Packer, N. H., & Schulz, B. L. (2016). Maturing Glycoproteomics technologies provide unique structural insights into the N-glycoproteome and its regulation in health and disease. *Molecular & Cellular Proteomics*, *15*, 1773–1790. https://doi.org/10.1074/mcp.O115.057638.

Tian, D., Sun, Y. H., Zhou, J. M., & Ye, Q. (2022). The global epidemic of SARS-CoV-2 variants and their mutational immune escape. *Journal of Medical Virology*, *94*, 847–857. https://doi.org/10.1002/JMV.27376.

Tien, M. Z., Meyer, A. G., Sydykova, D. K., Spielman, S. J., & Wilke, C. O. (2013). Maximum allowed solvent accessibilites of residues in proteins. *PLoS One*, *8*, e80635. https://doi.org/10.1371/JOURNAL.PONE.0080635.

Turoňová, B., Sikora, M., Schürmann, C., Hagen, W. J. H., Welsch, S., Blanc, F. E. C., et al. (2020). In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by

three hinges. *Science*, *370*, 203–208. https://doi.org/10.1126/SCIENCE.ABD5223/SUPPL_FILE/ABD5223S1.MOV.

Unione, L., Ardá, A., Jiménez-Barbero, J., & Millet, O. (2021). NMR of glycoproteins: Profiling, structure, conformation and interactions. *Current Opinion in Structural Biology*, *68*, 9–17. https://doi.org/10.1016/j.sbi.2020.09.009.

Valverde, P., Quintana, J. I., Santos, J. I., Ardá, A., & Jiménez-Barbero, J. (2019). Novel NMR avenues to explore the conformation and interactions of glycans. *ACS Omega*, *4*, 13618–13630. https://doi.org/10.1021/acsomega.9b01901.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, *26*, 1701–1718. https://doi.org/10.1002/JCC.20291.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., et al. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, *31*, 671–690. https://doi.org/10.1002/JCC.21367.

Vigerust, D. J., & Shepherd, V. L. (2007). Virus glycosylation: Role in virulence and immune interactions. *Trends in Microbiology*, *15*, 211–218. https://doi.org/10.1016/j.tim.2007.03.003.

Volz, E., Hill, V., McCrone, J. T., Price, A., Jorgensen, D., O'Toole, Á., et al. (2021). Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell*, *184*, 64–75. e11 https://doi.org/10.1016/J.CELL.2020.11.020.

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, *181*, 281–292. e6 https://doi.org/10.1016/j.cell.2020.02.058.

Walls, A. C., Tortorici, M. A., Frenz, B., Snijder, J., Li, W., Rey, F. A., et al. (2016). Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nature Structural and Molecular Biology*, *23*, 899–905. https://doi.org/10.1038/nsmb.3293.

Walton, E. B., & VanVliet, K. J. (2006). Equilibration of experimentally determined protein structures for molecular dynamics simulation. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *74*. https://doi.org/10.1103/PHYSREVE.74.061901.

Wang, J., Fan, Q., Satoh, T., Arii, J., Lanier, L. L., Spear, P. G., et al. (2009). Binding of herpes simplex virus glycoprotein B (gB) to paired immunoglobulin-like type 2 receptor α depends on specific sialylated O-linked glycans on gB. *Journal of Virology*, *83*, 13042–13045. https://doi.org/10.1128/JVI.00792-09/ASSET/82927C0C-8656-48E9-8BDF-0E30EA5E71B7/ASSETS/GRAPHIC/ZJV0240926790004.JPEG.

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., & Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*, eabb9983. https://doi.org/10.1126/science.abb9983.

Watanabe, Y., Berndsen, Z. T., Raghwani, J., Seabright, G. E., Allen, J. D., Pybus, O. G., et al. (2020). Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nature Communications*, *11*, 1–10. https://doi.org/10.1038/s41467-020-16567-0.

Watanabe, Y., Bowden, T. A., Wilson, I. A., & Crispin, M. (2019). Exploitation of glycosylation in enveloped virus pathobiology. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1863*, 1480–1497. https://doi.org/10.1016/J.BBAGEN.2019.05.012.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*, 1189–1191. https://doi.org/10.1093/BIOINFORMATICS/BTP033.

WHO Coronavirus (COVID-19) Dashboard. (n.d.). WHO Coronavirus (COVID-19) Dashboard with Vaccination Data. [WWW Document]. https://covid19.who.int/. (Accessed 28 February 2022).

Wilkinson, H., & Saldova, R. (2020). Current methods for the characterization of *O*-glycans. *Journal of Proteome Research*, *19*, 3890–3905. https://doi.org/10.1021/acs.jproteome.0c00435.

Wolter, N., Jassat, W., Walaza, S., Welch, R., Moultrie, H., Groome, M., et al. (2022). Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in South Africa: A data linkage study. *The Lancet*, *399*, 437–446. https://doi.org/10.1016/S0140-6736(22)00017-4.

Woo, H., Park, S. J., Choi, Y. K., Park, T., Tanveer, M., Cao, Y., et al. (2020). Developing a fully glycosylated full-length SARS-COV-2 spike protein model in a viral membrane. *Journal of Physical Chemistry B*, *124*, 7128–7137. https://doi.org/10.1021/acs.jpcb.0c04553.

Woods, R. J. (2018). Predicting the structures of glycans, glycoproteins, and their complexes. *Chemical Reviews*, *118*, 8005–8024. https://doi.org/10.1021/ACS.CHEMREV.8B00032/ASSET/IMAGES/ACS.CHEMREV.8B00032.SOCIAL.JPEG_V03.

Wormald, M. R., & Dwek, R. A. (1999). Glycoproteins: Glycan presentation and protein-fold stability. *Structure*, 7, R155–R160. https://doi.org/10.1016/S0969-2126(99)80095-1.

Wu, D., & Robinson, C. V. (2022). Understanding glycoprotein structural heterogeneity and interactions: Insights from native mass spectrometry. *Current Opinion in Structural Biology*, 74, 102351. https://doi.org/10.1016/j.sbi.2022.102351.

Yang, Z., & Hancock, W. S. (2004). Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *Journal of Chromatography. A*, *1053*, 79–88.

Yokoyama, M., Fujisaki, S., Shirakura, M., Watanabe, S., Odagiri, T., Ito, K., et al. (2017). Molecular Dynamics Simulation of the Influenza A (H3N2) hemagglutinin trimer reveals the structural basis for adaptive evolution of the recent epidemic clade 3c.2a. *Frontiers in Microbiology*, *8*, 584. https://doi.org/10.3389/fmicb.2017.00584.

Zhao, Y., Ren, J., Harlos, K., Jones, D., Zeltina, A., et al., & Stuart, D. I. (2016). Toremifene interacts with and destabilizes the Ebola virus glycoprotein. *Nature*, *535*, 169–172.