



Computational toolbox for the analysis of protein–glycan interactions

Ferran Nieto-Fabregat, Maria Pia Lenza, Angela Marseglia, Cristina Di Carluccio, Antonio Molinaro, Alba Silipo and Roberta Marchetti*

Review

[Open Access](#)

Address:

Department of Chemical Sciences, University of Naples Federico II,
Via Cinthia 4, 80126, Italy

Email:

Roberta Marchetti* - roberta.marchetti@unina.it

* Corresponding author

Keywords:

computational tools; glycan–protein interactions; MD; molecular recognition

Beilstein J. Org. Chem. **2024**, *20*, 2084–2107.

<https://doi.org/10.3762/bjoc.20.180>

Received: 20 December 2023

Accepted: 01 August 2024

Published: 22 August 2024

This article is part of the thematic issue "Chemical glycobiology".

Guest Editor: E. Fadda



© 2024 Nieto-Fabregat et al.; licensee

Beilstein-Institut.

License and terms: see end of document.

Abstract

Protein–glycan interactions play pivotal roles in numerous biological processes, ranging from cellular recognition to immune response modulation. Understanding the intricate details of these interactions is crucial for deciphering the molecular mechanisms underlying various physiological and pathological conditions. Computational techniques have emerged as powerful tools that can help in drawing, building and visualising complex biomolecules and provide insights into their dynamic behaviour at atomic and molecular levels. This review provides an overview of the main computational tools useful for studying biomolecular systems, particularly glycans, both in free state and in complex with proteins, also with reference to the principles, methodologies, and applications of all-atom molecular dynamics simulations. Herein, we focused on the programs that are generally employed for preparing protein and glycan input files to execute molecular dynamics simulations and analyse the corresponding results. The presented computational toolbox represents a valuable resource for researchers studying protein–glycan interactions and incorporates advanced computational methods for building, visualising and predicting protein/glycan structures, modelling protein–ligand complexes, and analyse MD outcomes. Moreover, selected case studies have been reported to highlight the importance of computational tools in studying protein–glycan systems, revealing the capability of these tools to provide valuable insights into the binding kinetics, energetics, and structural determinants that govern specific molecular interactions.

Review

Introduction

Carbohydrates also referred to as saccharides, sugars, or glycans, constitute one of the main building blocks of biomolecules, alongside lipids, proteins, and nucleic acids. In humans

and animals, they form the so-called glycocalyx, a protecting sugar coat decorating the cell surface and modulating a myriad of cell–cell interactions [1]. It is composed of branched or elon-

gated glycan chains covalently linked to proteins or lipids, hereby constituting glycoproteins or glycolipids, respectively. Recently, glycan structures exposed on the cellular membrane have also been found to be associated with tRNA [2]. In other species, such as prokaryotes, plants or fungi, glycoconjugates comprise the cell wall, playing critical metabolic, structural and physical functions [3].

Glycoscience encompasses the comprehensive study of glycans focusing on their structural, biosynthetic, biological and evolutionary aspects [4], thus playing a central role in the identification and characterisation of the glycome' structure and function, and in unveiling its interaction with host proteins [5,6]. Notably, the complexity of the glycome far surpasses that of the genome, transcriptome, and proteome, not only due to the structural and conformational diversity of glycans, whose synthesis is not template driven, but also due to their dynamic nature [5,6]. Although mammalian glycans rely on a group of "only" 10 monosaccharide units, they can be assembled, in linear or branched chains, through different glycosidic linkages and diverse spatial orientations, which can also undergo modifications, such as

methylation, sulfation, and phosphorylation, resulting in a plethora of different and particular structures [7,8]. Additionally, glycans can adopt a wide variety of different shapes; five-membered ring sugars can exhibit envelope and twist conformations usually represented on a pseudo-rotational wheel; while six-membered ring structures can adopt chair (C), boat (B), skew (S), and half-chair (H) conformations (Figure 1). Among them, chair' shapes typically have the lowest energy and are thus preferred, except few cases in which different conformations can exist in a dynamic equilibrium, as for the iduronic acid that can adopt three low-energy solution conformations (Figure 1): 1C_4 , 4C_1 (chair forms) and an additional skew-boat shape (2S_0) [9]. The glycosidic torsion angles Φ ($H_1-C_1-O_x-C_x$) and Ψ ($C_1-O_x-C_x-H_x$) describe the relative orientation of two connected monosaccharide units; moreover, when dealing with monosaccharides containing an exocyclic hydroxymethyl group, such as in the case of 1-6 linked sugars, an additional torsion, namely ω ($O_6-C_6-C_5-O_5$), must be defined and three staggered conformers, denoted as gg/tg/gt (ω angles of $-60^\circ/180^\circ/60^\circ$, respectively), should be considered (Figure 1).

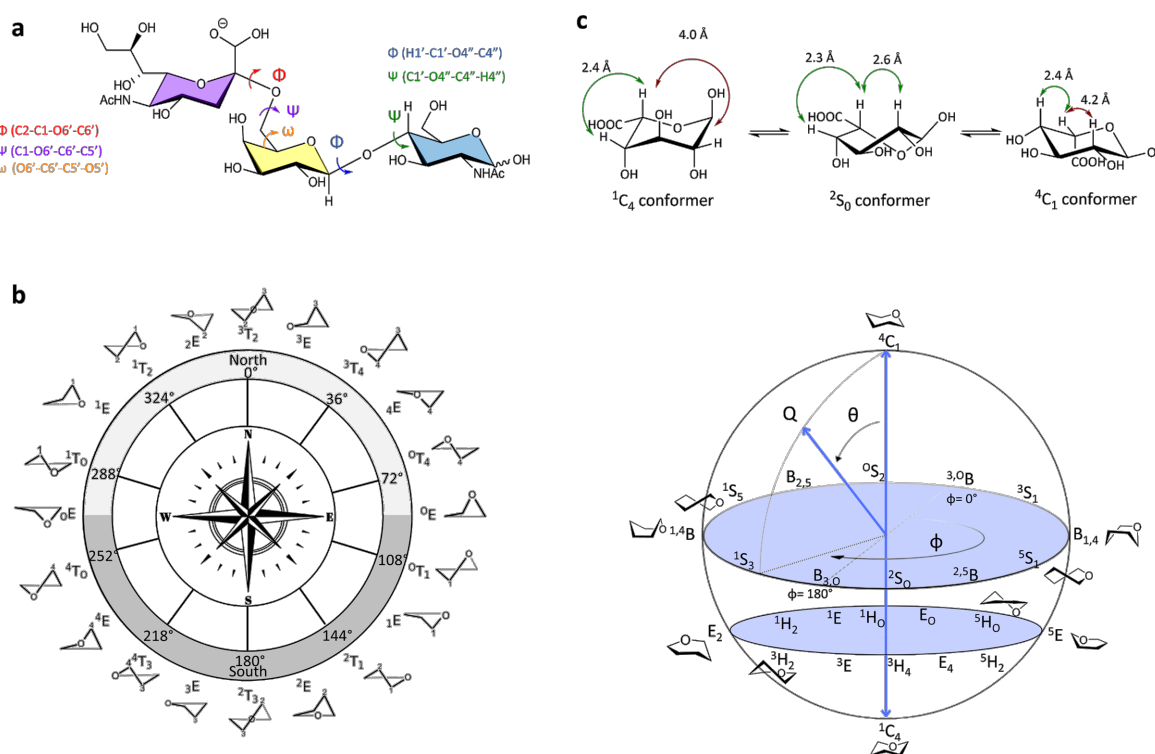
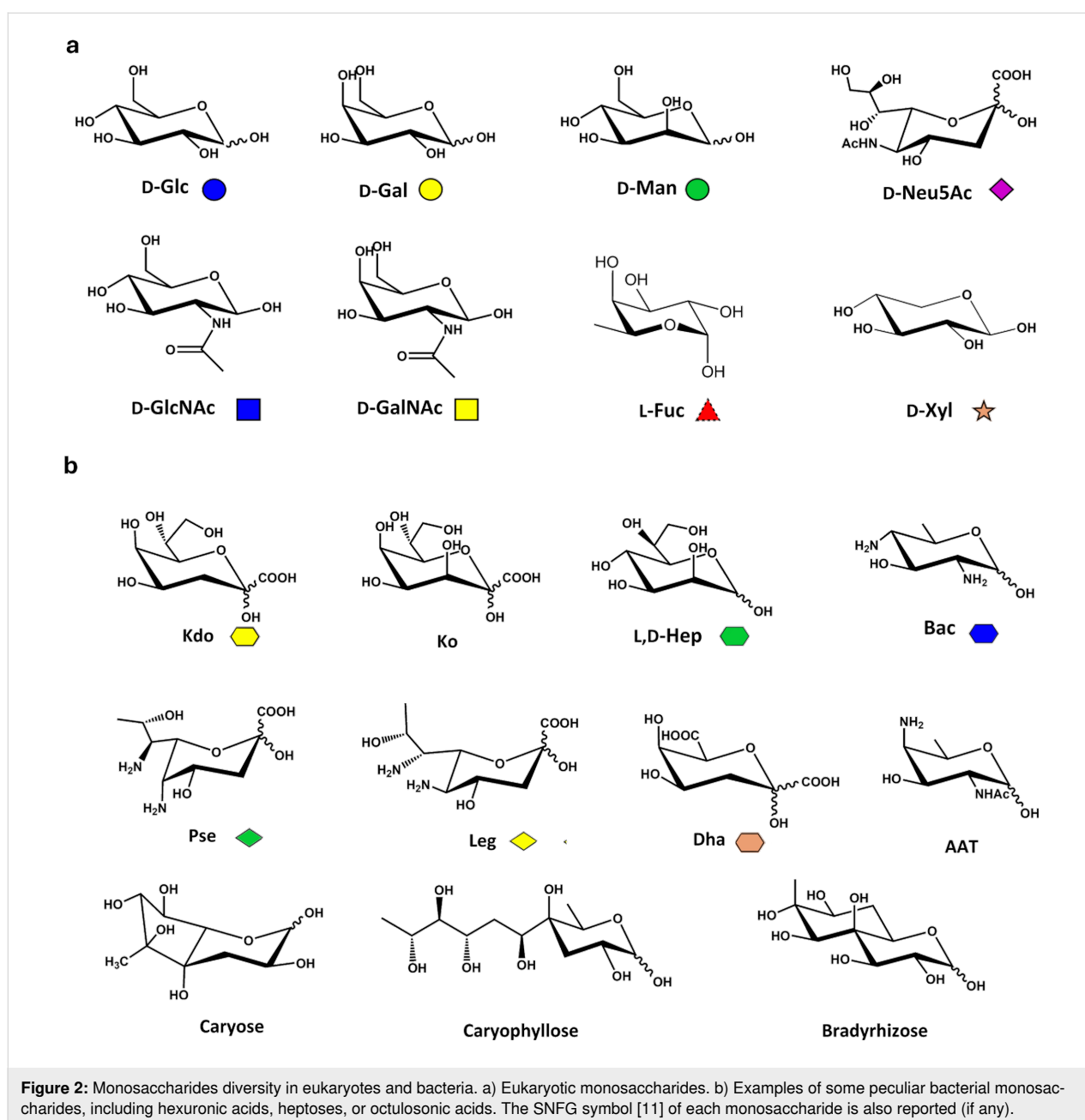


Figure 1: Carbohydrate conformational variability. a) Illustration of Φ , Ψ and ω dihedral angles for a representative trisaccharide coloured according to the symbol nomenclature for glycans (SNFG). b) On the left: Pseudo-rotational wheel depiction of five-membered ring structures showcasing envelope (E) and twist (T) conformations. On the right: Glove representation illustrating the puckering of six-membered pyranoside ring conformations. c) Equilibrium between the low-energy solution conformations of the iduronic acid. The exclusive (Nuclear Overhauser Effect) 1H - 1H NOE contacts characteristic of each conformation, 1C_4 , 4C_1 and 2S_0 , are also depicted.

Longer and branched glycans exhibit heightened structural dynamics, depending on the values adopted by the torsional angles around the glycosidic linkages [10].

The high variability of linkages type, branching, stoichiometry, anomeric configuration (alpha and beta), and conformation contributes to the intricate nature of glycans. The complexity of the glycome is even higher in bacteria, which are able to use most of the mammalian sugar units to construct their glycoconjugates but, in addition, can also use a wide variety of particular, and potentially endless, monosaccharides that are instead not present in eukaryotes (Figure 2).

This huge diversity and complexity, especially in bacterial glycans, makes the structural and conformational analysis of glycans extremely difficult, posing a considerable challenge when employing conventional structural biology methods for glycan analysis [10,12]. Nevertheless, understanding the three-dimensional structure of glycans is crucial for comprehending their roles and biological activities and for correlating their structural features with their activity [3,13]. Given the plethora of remarkable biological roles played by complex glycans, this knowledge is essential for their potential applications in promoting health benefits for humans, animals and plants, including drug design [14,15], vaccine development [15,16] and



numerous other possibilities in the field of carbohydrate chemistry and biology.

Notably, the regulation of the host immune response is often mediated by glycans, particularly through their recognition by a wide array of glycan-binding proteins (GBP) [17], which have a unique capability to specifically interact with endogenous and/or exogenous glycans [18,19]. Thus, disclosing the molecular basis of protein–glycan interactions has a unique potential to help modulate a myriad of complex biological events affecting the health and well-being of living organisms and the natural environment. Being key participants in the molecular dialogue, glycan binding proteins emerge as fascinating and critical components of molecular events that regulate life at its core. Their functions span from the catalysis of chemical processes [20,21], transporting and storing molecules [22], transducing and integrating information [23] providing structural and mechanical support [24], and generating movement [25], among other functions [26]. To fold and carry out their function properly, proteins often need post-translational modifications, including glycosylation, in which a carbohydrate chain is directly attached to a specific amino acid to generate glycoproteins and proteoglycans [27]. Based on the amino acid involved in the link with the carbohydrates chain, it is possible to classify different types of glycosylation: i) *N*-glycosylation, where a *N*-acetylglucosamine (GlcNAc) is linked to the nitrogen atom of an asparagine side chain [28]; ii) *O*-glycosylation, where a GlcNAc or *N*-acetylgalactosamine (GalNAc) is linked to the hydroxy group of a serine or threonine residue [29]; iii) *C*-glycosylation, where a mannose (Man) directly binds a tryptophan residue [30]; iv) the covalent attachment to core protein of glycosaminoglycans (GAGs), anchored to a Ser, or at lesser extent to Thr or Asn, forms proteoglycans. GAGs are complex negatively charged polysaccharides composed by disaccharide repeats of GlcNAc or GalNAc combined with uronic acid (glucuronic or iduronic acid) or galactose residues, forming chains which can also be partially sulfated. GAGs family includes heparan sulphate (HS), dermatan sulphate (DS), chondroitin sulphate (CS), keratan sulphate (KS), and hyaluronic acid (HA) [31]. The extraordinary proteins versatility places them at the core of almost every biological event, including cell–cell communication and regulation of immune responses. In the majority of cases, these mechanisms are significantly influenced by the molecular interactions occurring between glycans and receptor proteins.

A well-known family of GBPs is constituted by the lectins, ubiquitous receptors that exhibit the ability to specifically recognise different carbohydrates through their well-defined binding pocket and they conserved three-dimensional structure similarities [32]. On the other hand, GAG-binding proteins,

which are able to recognise carboxylic acid and sulphate groups along glycosaminoglycan chains using clusters of positively charged amino acids [33], also mediate a wide variety of cell–cell and cell–pathogen communication, controlling immune cell functions, and overseeing cellular trafficking [34]. Another class of GBP is represented by anti-carbohydrate antibodies, that are generally produced by the host organism for example against bacterial, fungal, and viral carbohydrates [35].

Given the wide variety of biological processes influenced by the protein–glycan interplay, an increasing attention has been focused in the last decades on the development of new techniques and technologies for the systematic analysis of complex glycans and the study of their interactions with proteins. A multidisciplinary approach, spanning from wet laboratory experiments and biophysical techniques to bioinformatics methods is needed to deeply investigate the multifaceted aspects of protein–glycan interactions. To date, advanced and versatile NMR, X-ray crystallography, and MS methods [36–38] above all, have been developed to reach extensive information on the structural and conformational features of glycans and proteins. The experimental techniques employed for the analysis of these complex biomolecules are not discussed here; for a more in-depth understanding on this topic, the reader is referred to some comprehensive reviews [7,36,39–41]. Here, we focus instead on different computational and bioinformatic tools, designed to guide the structural and conformational elucidation process, and on the application of molecular dynamic simulations to the study of proteins and glycans in free and bound states. Detailed protocols and methods for protein and glycan modelling are extensively described and links to web servers and downloadable software, which can help researchers in designing the workflow to study a glycan–protein system, are also reported.

Computational tools to study glycans in the free state

Since the first molecular dynamics simulations performed in the late 1980s on oligomannose type glycans [42] and in the early 1990s on complex type glycans [43], great steps forward have been made in the computational analysis of complex carbohydrates. The advancement of computing power, the emergence of GPUs, and specialised processors accelerated MD simulations making it a key scientific tool to explore complex systems, including glycans, with ever-increasing accuracy and efficiency [44].

Tools for building structural models of carbohydrates

Before going into details of the computational tools that can be used to dissect the 3D conformational features of glycans, an

overview of the most useful web services and software to build 2D and 3D models of carbohydrate structures is reported here.

Notably, despite the existence of several encoding formats for glycans (Figure 3), significant efforts have been made in the years to enable a simple and standardised glycan representation, which would simplify the transmission and efficiency of the communication within the scientific community. This led to the extensive use of the symbol nomenclature for glycans (SNFG) representation that is used in all the tools described below (Figure 3) [11,45].

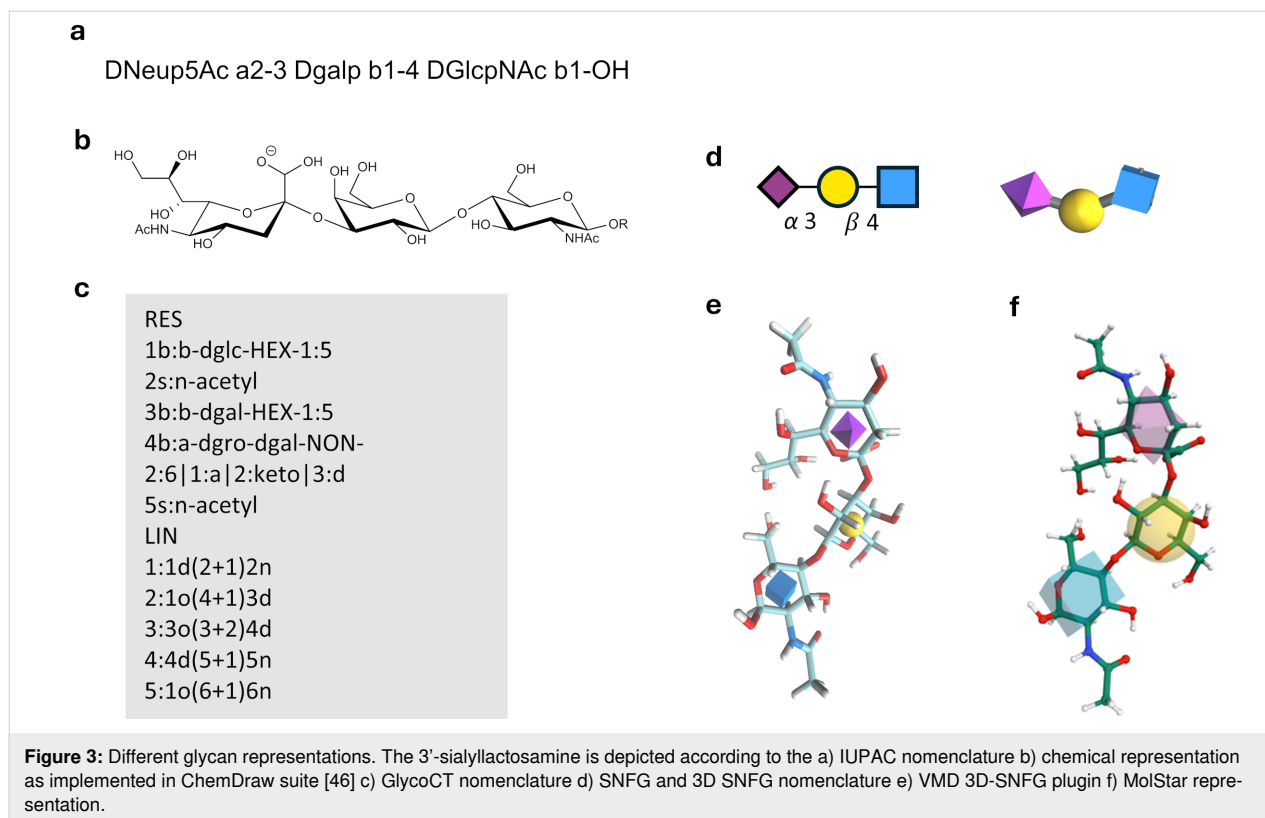
Numerous computer applications have been developed to allow manual drawing and sketching of carbohydrates, as reviewed by Lal et al. [47]. Here, we list free tools useful not only for sketching and drawing but also for building the glycan structure of interest (last accessed date: May 2024) [47,48].

1. doGlycans [49]: Free desktop software package in the python framework that allows users to prepare carbohydrate structures for atomistic simulations of complex glycoproteins, glycolipids and carbohydrate polymers in the GROMACS force field format (see below). Polysaccharides can be prepared by using the *prepreader.py* tool, glycoproteins and glycolipids by using

the *doglycans.py* tool. (<https://bitbucket.org/biophys-uh/doglycans/>).

2. Glycam-Web carbohydrate builder [50]: Free online web-service that gives the possibility to model the 3D structures of molecules and complexes containing carbohydrates starting from monosaccharide building blocks, being also able to add branching points and some sugar derivatisation, including methylation and acetylation. The user has also the possibility to choose the ring type and the anomeric configuration of each monosaccharide. Once the structure is complete, it is possible to download not only the generated .pdb files of the minimised resulting structures but also files for input to an AMBER simulation (<https://glycam.org/>). Notably, among the currently available interfaces for modelling oligosaccharide conformations on glycam website, one is dedicated to GAG modelling [51] (<https://glycam.org/gag/>).

3. CHARMM-GUI [52]: Online free web-service that offers a great variety of possibilities for reading and modelling .pdb files. It is a versatile program for atomic-level simulations, which can be run directly in the webserver. It has a special focus on macromolecules of biological interest; indeed, this platform contains a number of different modules designed to construct complex glycans, glycoconjugates as lipopolysaccharides (LPS), or even building a membrane system or solvating a



protein. To use all these features, grouped in the *Input Generator* tab (<https://www.charmm-gui.org/>), it is necessary to be registered to the web site.

4. Azahar [53]: freely available python-based plugin that permits to visualise, analyse and model glycans and glycoconjugates (<https://pymolwiki.org/index.php/Azahar>).

5. 3D-SNFG: It is a script integrated in the visual molecular dynamics (VMD) program [54] (see below) that allows a cartoon representation of glycans according to the symbol nomenclature for glycans (see Figure 3) (<https://glycam.org/docs/othertoolsservice/downloads/downloads-software/index.html>).

Choosing the most appropriate simulation software package

With the aim to accurately prepare glycans for MD simulations, it is fundamental not only to build their 3D structure but also to choose the appropriate force field, that is a set of empirical energy functions and parameters used to calculate the potential energy of a system as a function of the molecular coordinates. The collection of equations and associated constants designed to reproduce the molecular geometry and selected properties of a system, as well as the naming and labelling of the system atoms, vary from one force field to another; therefore, it is important to ensure compatibility between the input file of the tested structure and the chosen force field. The Automated Topology Builder (ATB) and repository [55] (<https://atb.uq.edu.au/>) is a free web server providing topologies and parameters for a wide range of molecules. It provides access to classical force fields in formats compatible with different simulation packages, including GROMACS (see below), even offering a GROMOS to AMBER topology file converter. In the years, different MD simulation software packages have been developed and designed to simulate the movements and interactions of atoms and molecules over time; the three described below are currently widely used in the field of computational chemistry and biochemistry:

1. AMBER [56]: AMBER (<https://ambermd.org/>) is the acronym for "Assisted Model Building with Energy Refinement", and it is an open-source software widely employed for molecular modelling and simulation. It is known for its stability, user-friendly interface, and a wide range of analysis tools for studying complex biomolecular systems. AMBER provides various force fields, specifically optimised for simulating biological molecules, as lipids (lipids21) [57], proteins (ff14SB) [58], water molecules (TIP3P) [59], general organic molecules (gaff2) [60], and sugars (GLYCAM_06j) [61]. Notably, as mentioned above, on the GLYCAM-web site, it is

possible to easily construct glycans with GLYCAM force field nomenclature; however, it is worth to note that only a few bacterial monosaccharides are available in the GLYCAM-web carbohydrate builder. For most bacterial sugars, the parametrisation of each building block is needed and requires the use of ab initio methodologies, including several steps of charges and electron density calculations, optimization and minimization, making the computational study of bacterial glycans difficult and time-consuming.

2. CHARMM [62]: CHARMM, acronym of "Chemistry at HARvard Molecular Mechanics", is a free extensively utilised molecular modelling and simulation software package. Its force field is at the core of CHARMM's capabilities, which serves as a comprehensive set of parameters and mathematical functions to describe the potential energy and interatomic interactions within a molecular system. The CHARMM force field includes parameters for various types of atoms, bonds, angles, dihedrals and non-bonded interactions, encompassing van der Waals forces and electrostatic interactions. CHARMM19 (united atom), CHARMM22, CHARMM27, and CHARMM36 (all atom) are some of the popular force fields available in the program. (<https://www.academiccharmm.org/>)

3. GROMACS [63]: GROMACS is the acronym for "Groningen Machine for Chemical Simulations"; it is a powerful open-source software package for molecular dynamics simulations in the field of computational chemistry and bioinformatics. It is extensively used to model and simulate the dynamic behaviour of various molecular systems, including proteins, nucleic acids, lipids, and small molecules. GROMACS provides various tools for system preparation, simulation setup, and post-simulation analysis. It is possible to use different force fields that include GROMOS96, GROMOS53A6, and GROMOS54A7, which are suitable for simulations of biomolecules, organic compounds, and a wide range of solvents (<https://www.gromacs.org/index.html>).

Despite several carbohydrate-specific force fields have been developed over the years [64–68], to date, the most widely used force field for carbohydrates is GLYCAM, which is continuously updated and improved to accurately describe their peculiar and complex set of conformational and energetic properties [61,69]. For specific studies involving unusual ligands, useful tools can be employed to provide the parameters needed for running MD simulations. Charge calculations and electron density computations for glycan units can be performed using tools like the online RED Server [70]. Although information on force fields is usually available, modifications can sometimes be required and can be achieved through ab initio calculations or programs like for example VFFDT [71].

Tools for the conformational analysis of glycans in the free state

The structure and biological functions of glycans are closely intertwined; the roles they play are influenced not only by their chemical composition but also by their conformation. As mentioned above, glycans are characterised by a huge conformational diversity (see Figure 1): even individual furanoid or pyranoid monosaccharides can assume various shapes and in longer glycans the relative orientation of the different monosaccharide building blocks is dictated by the values of different glycosidic torsion angles.

MM calculations: Investigating the energetically favourable conformations of carbohydrate disaccharide units composing the molecule of interest represents a pivotal step for generating reliable 3D glycan structure. A first analysis of glycan conformational features can be done by means of molecular mechanic calculations that allow to build the adiabatic energy maps, represented as a function of Φ and Ψ torsion angles, in which the energetic minima that can be populated by a specific disaccharide are reported [72–74]. Currently, different databases, which are described below, collect adiabatic energy maps facilitating the construction of glycan 3D models by enabling the selection only of permitted low energy conformations:

1. CSDB [75–77]: Carbohydrate Structure Database is a publicly accessible platform for multiple glycoinformatic studies and web tools, which among the other services allows the users to locate adiabatic maps for specific glycosidic linkages. Generally, CSDB offers a wealth of valuable features; it provides structural, bibliographic, taxonomic, NMR spectroscopic and other information on glycan and glycoconjugate structures of prokaryotic, plant and fungal origin. The retrospective literature analysis is the main source of structural data, which are then manually curated and approved. Besides structures, the database includes bibliography, abstracts, keywords, biological source data up to strains, methods used to elucidate structures, NMR signal assignment and other information (<http://csdb.glycoscience.ru/database/>).

2. Disac3DB: free annotated database that contains the 3D structural information of about 120 entries of disaccharides. For each disaccharide, an exhaustive search was performed using the MM3 molecular mechanics force field [66], giving a complete sampling of the conformational space and yielding the construction of relaxed adiabatic energy maps (<https://glyco3d.cermav.cnrs.fr/disac3db/>). It is worth to note that the presence of additional residues in the neighborhood of the studied glycosidic linkage may cause shifts in the values of the favored torsional angles. Thus, to evaluate if the presence of further residues results in limitations of the possible conforma-

tions of an individual glycosidic linkage, and/or if the adiabatic map of interest is not present in the aforementioned databases, the Schrodinger Suite of programs through the Maestro graphical interface can be exploited to generate the maps by using the MM3 force field [66]. The Schrodinger platform (<https://www.schrodinger.com/>) offers several services for molecular design and discovery providing access to physics-based molecular modelling tools and machine learning technologies from a single modelling environment, however, it is not free-accessible.

Further valuable insights into the structure and conformation of saccharides, determined by experiment and simulation, are available on the Stenutz's website (<https://www.stenutz.eu>). In particular, information on the preferred conformation of glycosidic linkages and the favoured dihedral angles for the OH group at position 6 in hexoses are reported. This website also provides a compilation of standardised procedures, providing practical guidelines for carbohydrate structural analysis, spanning from the purification to the structural analysis of polysaccharides.

MD simulations: Once the 3D glycan structure is built, taking into account the energetically favourable conformations of each constituent disaccharide unit, and the appropriate force field/simulation package is chosen, molecular dynamics simulations can be performed to gain insights into the glycan conformational behaviour. MD simulation generates an ensemble of conformations by applying the laws of motion to the atoms of the molecule [48], allowing to: i) sample the glycan conformational space; ii) investigate how the glycan behaves in a solution (if the MD is performed in explicit solvent), describing carbohydrate–water interactions; iii) monitoring the intramolecular interactions. Usually, this information has to be further validated by performing experimental studies (primarily nuclear Overhauser effect (NOE) and residual dipolar coupling-based experiments) to get accurate information on the glycan conformational behaviour and eventually apply some experimentally derived constraints.

The calculation and analysis of MD simulations of glycans in the free state can be performed with the same tools described below in the protein–ligand interactions section.

Computational tools to study proteins in the free state

Knowledge on the three-dimensional structure of a protein is essential for understanding the functions and the dynamics of protein interactions. Several experimental techniques, including NMR, X-ray crystallography and Cryo-EM, can provide critical information for the characterisation of protein structure and

conformation. Their widespread and wise use permitted to experimentally determine the structures of around 200,000 proteins [78], all organized in the Protein Data Bank (PDB), that is a freely and publicly available central archive of macromolecular structural data, established in 1971. However, the three-dimensional shape of billions of known protein sequences is not available yet. In this scenario, bioinformatic tools can come to the aid of predicting protein three-dimensional structure with high accuracy, as outlined below [79–85]. Notably, generated models have also occasionally helped solve protein structures [86], further highlighting the great potential of the integrated use of bioinformatic tools and experimental data.

Tools for protein structure prediction

Due to the vast conformational space and a complex energy function, protein structure prediction (PSP) is a computationally challenging task. Homology modelling is a template-based PSP that may be used to predict the 3D structure of a protein based on its amino acid sequence and the structure of a related protein that is already known. However, also template-free PSP has obtained significant progress recently via machine learning and search-based optimisation approaches [87]. There are several software programs and tools available for homology modelling, and some of the most popular include:

1. AlphaFold2 [88]: It is an open-access protein structure prediction system based on artificial intelligence and machine learning. It is based on a neural network that can predict the 3D protein structure at a high accuracy level. The AlphaFold solution is composed of two steps. First, given a protein sequence, it generates multiple alignments with sequences from all the species, including evolutionary profiles from different sources. In the second step, a model refinement is generated based on structural refinement (where the network optimises the torsion angles, bond length and bond angles), distance constraints (according to laws of physics) and gives an output with the structure with the minimised energy (<https://alphafold.ebi.ac.uk/>).

2. I-TASSER [89]: Iterative Threading ASSEmbly Refinement is a free online server that combines ab initio protein structure prediction with template-based modelling. It is known for its ability to predict both the structure and function of a protein. It is based on identifying structural templates from the PDB by several threading methodologies with full-length atomic models (<https://zhanggroup.org/I-TASSER/>).

3. Modeller [90]: It is an open-access program used for homology or comparative modelling of proteins. The user inputs an alignment of a sequence to be modelled with known related structures, and the computer generates a model of all

non-hydrogen atoms. It can do de-novo modelling of protein loops and apply spatial constraints (<https://salilab.org/modeller/>).

4. Rosetta [91]: It was developed for de novo protein structure prediction in a free version. Homology modelling is also applied in this instance by using several protein templates that hybridise the most homologous sections of various templates into a single model while modelling missing residues de novo. Advances in the scoring function, which is a mix of physics-based and knowledge-based potentials fitted against known structures and thermodynamic observables, have increased the accuracy of predictions. Incorporating experimental data into models has been made more accessible. The same research group also developed RoseTTAFold, which uses deep learning to quickly and accurately predict protein structures based on limited information [92]. However, very accurate structures for complex proteins are yet to be achieved at a level suitable for effective drug design. Moreover, ab initio prediction of a protein's structure only from its amino acid sequence remains unsolved. Accessing Rosetta molecular modelling software tools (<https://www.rosettacommons.org/software>) has traditionally required expertise in the Unix command line environment, limiting their use. A web server called ROSIE [93] was created to provide a more accessible environment for selected Rosetta protocols. Academic users can access ROSIE freely (<https://rosie.rosettacommons.org/>).

5. SWISS-MODEL [94]: It is a web-based integrated free service dedicated to protein structure homology modelling. It guides the user in building protein homology models at different levels of complexity. This program builds a homology model by employing four main steps: (i) identification of structural template(s), (ii) alignment of target sequence and template structure(s), (iii) model-building, (iv) model quality evaluation. Each of the above processes may be repeated iteratively until a satisfactory model is produced (<https://swiss-model.expasy.org/>).

6. UniLectin [95]: It is an interactive, publicly accessible platform that provides curated and predicted lectin data, not only including structural information on lectins and their interactions with carbohydrate ligands, but also predicting the occurrence of lectins in genomes. UniLectin3d is one of the modules integrated in UniLectin, which provides curated information on 3D structures of lectins [94–96] a classification system based on both taxonomic origin and structural fold (<https://unilectin.unige.ch/>).

7. GlycoShape3D [97]: It is a freely available database for academic user that enriches the landscape of glycobiology

resources. It offers structural insights into glycoproteins, addressing challenges posed by glycan complexity, flexibility, and heterogeneity. In particular, the Re-glyco tool allows the user to restore the missing glycosylation on glycoproteins deposited in the RCSB PDB or in the EBI-EMBL AlphaFold protein structure database (<https://glycoshape.org/>).

The quality of the generated protein model is contingent on elements such as the chosen template structure (if any), the sequence alignment, and the choice of the modelling algorithm. To ensure a high accuracy of the predicted model, which should be at least comparable to that of experimental structures, several programs can be employed for model validation and refinement. Among them, PROCHEK [98] is an open-source program that permits to check the quality of the protein structure by analysing the Ramachandran plots, the planarity of peptide bonds, the bad non-bonded interactions, the distortions of the geometry around the C α atoms, the energies of hydrogen bonds, and the departure of the side chain χ torsion angles from expected values. Improvement and/or validation of modelled or experimentally solved structures can be also obtained by using CASP (critical assessment of protein structure prediction) [99], which consists of a free platform established in 1994 to help advance the methods of identifying protein structure from sequence. The Protein Structure Prediction Center (<https://www.predictioncenter.org/>) has been organized to allow researchers to objectively test their structure prediction methods. Some of the best performing methods (including among the others AlphaFold, RosettaFold and I-TASSER) are implemented as fully automated servers, which can be used by public for protein structure modelling.

MD simulations: Generating an accurate protein model or choosing the appropriate published 3D structure of a protein is essential to obtain reliable and precise results from MD simulation. It is also worth to note that, to generate the input files for MD, some modifications on the .pdb file of the protein are required. For instance, capping the protein termini with non-charged groups and replacing original hydrogens to guarantee compatibility with the selected force field is required before running MD simulations. Other modifications can include adding a disulfide bond between specific cysteines and filling the missing side chains and missing loops (if any) to restore the integrity of the protein. Here, we list a series of software tools and packages which are commonly employed to generate the protein input files for MD:

1. Molprobit [100]: It is a widely used web-based software suite for evaluating and enhancing the quality of protein structures, especially those intended for molecular dynamics simulations, available in a free version. Specifically, it is possible to

check the H atoms, the quality of the structure, evaluate some steric clashes and visualise in a friendly manner the full structure (<http://molprobit.biochem.duke.edu/>).

2. PDBtools [101]: It is a freely accessible software that allows the manipulation and modification of a PDB file. Different tools are available, such as deleting atoms, renaming the polypeptide chain, calculating disulphide bonds, adding missing atoms and mutating residues (<https://wenmr.science.uu.nl/pdbtools/submit>).

3. ProteinPrepare [102]: This application enables users to modify PDB files and create input files for molecular dynamics by adding missing atoms, removing H atoms, and analysing the proton state of amino acids. The registration of the user to the web-site is required to access these tools (<https://playmolecule.com/proteinPrepare/>).

Several MD simulation packages, including CHARMM-GUI [52], AMBER [56], and GROMACS [63], offer built-in utilities for preparing input files. These tools also provide extensive documentation and tutorials to help users effectively create MD input files for proteins. Once the protein input files are generated, MD simulations can be run.

Computational tools to study protein–ligand complexes

Detailed investigations of protein–ligand interactions, combining experimental and computational methods, provide an indispensable basis to depict holistic pictures of molecular complexes allowing to modulate them at will. The computational approach involves i) predicting/building the protein and the ligand in their optimal conformation (as discussed above), ii) predicting the protein binding site; iii) modelling the ligand into the protein binding site, iv) assessing binding affinity through sampling and scoring, as discussed in the following paragraphs [103].

Prediction of the protein binding site

Over the years, structure-, sequence-, and homology/template-based methods have been employed to identify and predict carbohydrate-binding sites starting from the protein structure [104]. Recently, thanks to the fast development of machine learning techniques, new computational tools have been developed to facilitate the prediction of protein binding sites.

We report here only the applications related to the protein interaction with glycans:

1. PeSTo-Carbs [105]: it is an extension of Protein Structure Transformer (PeSTo) [106], a deep learning method to

predict protein interaction interfaces with other proteins, nucleic acids, lipids, small molecules, and ions, starting from a protein structure. PeSTo-Carbs is specifically trained to predict carbohydrate and cyclodextrin binding interfaces on proteins. Two different modules are available: a general model PS-G for a wide range of carbohydrates, their derivatives and cyclodextrins, and a specific model PS-S for important carbohydrate monomers. All of these features are available for free without registration as online tools (<https://pesto.epfl.ch/>).

2. GlyNet [107]: it is a free deep learning algorithm, based on neural networks (NN), that allows the user to predict protein-glycan binding. Taking a glycan structure as input, this model is able to predict the strength of the interaction based on the relative fluorescence units (RFUs) measured in the Consortium for Functional Glycomics glycan arrays and extrapolating these to RFUs from untested glycans (<https://github.com/shauneth/glynet>).

3. LectinOracle [108]: it is a freely available deep learning-based model that combines transformer-based representations for proteins and graph convolutional neural networks for glycans to predict their interaction (<https://github.com/BojarLab/LectinOracle>).

4. CAPSIF [109]: CARbohydrate-Protein Site Identifier is a convolutional neural network able to predict protein-carbohydrate binding interface from a protein structure. In contrast to other DN algorithms, as GlyNet and LectinOracle, which predict lectin-carbohydrate binding on a protein level, it provides residue-level information for non-covalently bound carbohydrates either from an experimental or generated-model protein structure. It includes two modules: CAPSIF-Voxel that predicts the protein binding residues and CAPCIF-Graph that predicts which residues bind sugars. It is freely available for use, and the code for CAPSIF can be accessed on GitHub (<https://github.com/Graylab/CAPSIF>).

To identify potential binding sites on the protein's surface, docking calculations can also be performed (see below).

Docking calculation tools for interaction studies

Molecular docking plays a crucial role in computer-aided drug development, allowing systematic evaluation of compound libraries to identify high-affinity lead compounds for specific targets. Bio-algorithms enable modelling protein tertiary structures, predicting ligand binding pockets, and supporting drug discovery through molecular docking [110]. Advances in information technology and improved computational efficiency have made computational methods integral to modern biological

research, and large-scale structure-based docking screens have become common, facilitating the exploration of vast chemical spaces and identifying potential target hits from extensive compound libraries [103]. While docking programs and servers may exhibit variations in their operational methods, they generally adhere to a common workflow comprising two primary phases. The first phase involves a conformational search aimed at predicting potential ligand conformations. This is followed by the second phase, which focuses on scoring the binding poses obtained during the conformational search. In this phase, the generated ligand-receptor complexes are assessed and ranked based on their binding energy thanks to the use of scoring functions [111].

Docking calculations can be conducted in two distinct ways: blind dockings, which explore the entire protein surface [112], and directed docking, typically employed when prior knowledge of the binding pocket exists and performed within a predefined box. Blind dockings are performed using cavity detection programs and online servers, as follows:

1. CB-Dock2 [113]: Cavity-detection guided Blind Docking 2 (<https://cadd.labshare.cn/cb-dock2/index.php>) is an online protein-ligand docking program designed to perform blind docking at predicted sites instead of the entire surface of a protein. Thus CB-Dock automatically recognises putative binding sites to determine their centre and size, with the aim to adjust the docking box to suit specific query ligands. Finally, molecular docking calculations are performed with Autodock Vina (see below).

2. Fpocket [114]: It is an open-source pocket detection package based on Voronoi tessellation and alpha spheres. It consists of three main programs: Fpocket for pocket identification, Tpocket for benchmarking pocket detection, and Dpocket for collecting pocket descriptor values. Written in C, Fpocket is well-suited for developing new scoring functions and extracting various pocket descriptors on a large scale. Fpocket 1.0 outperforms industry standards by detecting a high percentage of pockets within the best-ranked ones and offers a fast, open-source solution for protein pocket detection (<https://github.com/Discngine/fpocket>).

3. GRID [115]: It is a computational tool used to identify energetically favourable binding sites, known as molecular interaction fields (MIFs), on molecules with known structures. GRID has various applications, including ADME prediction, site of metabolism prediction, ligand-based and structure-based design, pharmacophore elucidation, water network prediction, and 3D-QSAR. GRID 2021 introduced a new interface aimed at structure-based design. It enables users to explore binding sites

using classic GRID MIFs, encompassing 74 different chemical types. Additionally, it offers a new molecular probe for generating MIFs specific to fragments of interest. GRID 2021 includes a 3D sketcher for visualising ligand modifications, and its Designer mode assists in finding optimal chemical moieties for specific sites (<https://www.moldiscovery.com/software/grid/#:~:text=GRID%202021%20is%20a%20new,MIFs%20for%20fragments%20of%20interest>).

When there is prior knowledge of the protein binding pocket, it is time saving to define an optimal docking search space or box and study specific binding pockets, improving docking accuracy and efficiency. Customising the box size for individual ligands, based on their size and the relationship with the search space, can be done by comparing the target protein to related proteins or those co-crystallised with ligands [103] and manually superimposing the new ligand to the reference structure. Some payment software like Glide [116,117], GOLD [118] and Molecular Operating Environment (MOE) dock [119] can be used for this purpose, although here are listed free docking tools:

1. Autodock [120,121]: It is a suite of advanced docking tools used for predicting how small molecules, such as drug candidates or substrates, interact with known 3D protein structures. It offers two generations of software, namely AutoDock 4 and AutoDock Vina, and a user-friendly graphical interface called AutoDockTools (ADT) to assist in configuring ligand rotatable bonds and analysing docking results. Additionally, the accelerated AutoDock-GPU is designed for faster performance, surpassing the original single-CPU docking code by hundreds of times. AutoDock 4 comprises two main programs: *autodock* handles ligand docking by aligning it with precomputed protein grids, while *autogrid* generates these grids. The grids can also assist organic chemists in designing better binding molecules (<https://autodock.scripps.edu/>).

2. Autodock Vina [122]: It is the open-source improved successor of Autodock. Vina is improved in terms of accuracy and performance as simplifies the process by instantly calculating grids internally, eliminating the need for manual grid map selection and atom type assignments (<https://vina.scripps.edu/#>).

3. FlexAID [123]: It is a molecular docking software capable of setting small molecules and peptides as ligands, and proteins and nucleic acids serve as docking targets. Notably, FlexAID shows support for full ligand flexibility and the flexibility of side chains in the target. It achieves this by employing a soft scoring function that assesses the complementarity between the surfaces of the ligand and the target. Thus, FlexAID has demon-

strated superior performance compared to well-established software like AutoDock Vina, particularly when target flexibility plays a pivotal role, as is often the case when working with homology models (<http://biophys.umontreal.ca/nrg/resources.html>).

4. HADDOCK [124]: High Ambiguity Driven protein–protein DOCKing is an advanced computational approach used for modelling interactions in biomolecular complexes. Noteworthy, HADDOCK incorporates information from known or predicted protein interfaces into the docking process through ambiguous interaction restraints and allows the specification of precise distance restraints (e.g., based on MS cross-links). It also supports a range of experimental data, including NMR residual dipolar couplings, pseudo contact shifts, and cryo-EM maps, positioning HADDOCK as a versatile tool capable of handling various modelling scenarios, such as protein–protein, protein–nucleic acids, and protein–ligand interactions.

The majority of existing docking software was originally designed for small, rigid, drug-like molecules, therefore, limiting their effectiveness in studying protein–carbohydrate interactions [125,126]. The development of specialized programs has been crucial in enhancing the accuracy of docking calculations [125,126]. We here listed a series of programs for running docking calculations with a special focus on those specifically designed to address the unique challenges posed by glycans [127,128].

1. Vina-Carb [129,130]: Vina-Carb is a module of AutoDock Vina (downloadable with a free version at <https://glycam.org/docs/othertoolsservice/downloads/downloads-software/index.html>), proven to be a valuable tool for studying carbohydrates. It incorporates carbohydrate intrinsic (CHI) energy functions and explicit water to better handle glycosidic linkages and improve docking accuracy. When Vina-Carb was applied to antibodies, lectins, and carbohydrate binding modules (CBM), the success rates in predicting accurate binding modes reached 86%, 50%, and 42%, respectively, compared to 70%, 50%, and 0% for AutoDock Vina. Although Vina-Carb generally performed slightly better over AutoDock Vina when docking glycans to proteins, it does not always rank the best docking pose as the top scoring pose.

2. BALLDock/SLICK [131]: It is a molecular docking method specifically designed to accommodate carbohydrate-like compounds, employing a genetic algorithm that allows for ligand and receptor side-chain flexibility. Designed specifically for protein–carbohydrate interactions, SLICK includes terms that consider CH– π interactions, hydrogen bonds, smoothed van der Waals interactions, and electrostatic interactions. The SLICK

scoring function, tailored for carbohydrates, enhances the accuracy of predicting binding modes and free binding energies. Compared to other programs such as AutoDock and FlexX (see below), BALLDock/SLICK demonstrates superior performance in structural and energetic precision. This method is particularly valuable in drug design involving protein–carbohydrate interactions, addressing weaknesses such as the CH– π interactions that are challenging for other programs like Vina-Carb.

3. FlexX [132,133]: It is a molecular docking software (unfortunately not free) designed to predict the conformations of small molecules in protein binding sites, thus facilitating the discovery of new drugs. Within SeeSAR, its functionality allows ligands to be placed in binding sites using an incremental construction algorithm that splits ligands into fragments, places, and scores them quickly in the binding site. The best fragments are then assembled to form the complete ligand, optimizing the generated conformations. FlexX's strengths include rapid and efficient exploration of the conformational space, handling ligand flexibility, a precise scoring function, and smooth integration with other molecular modelling programs. Additionally, FlexX excels in processing large libraries at high speed and is user-friendly, requiring no prior receptor preparation.

4. ROSETTA [134]: The development of GlycanDock [134], a protein–glycoligand docking refinement algorithm integrated in the RosettaCarbohydrate framework [135], allowed the use of Rosetta macromolecular modelling and design software suite to perform docking calculations on glycans bound to proteins with a higher accuracy with respect to previous Rosetta's protein–small molecule docking algorithms. Unlike other docking programs such as AutoDock, AutoDock Vina, DOCK, FlexX, Glide, and GOLD, which are primarily designed for small, rigid ligands, GlycanDock is specifically optimized to address the flexibility and complex structural features of glycans. The carbohydrate chains are treated as flexible oligomers, allowing extensive conformational sampling of the glycoligand while maintaining glycosidic linkages within predetermined, energetically favorable minima to ensure biophysically realistic carbohydrate structures. GlycanDock handles the flexibility and complexity of glycans better than other docking programs and can be downloaded as part of the Rosetta package from the Rosetta Commons (<https://www.rosettacommons.org>).

5. GlycoTorch Vina [136]: GTV is a free molecular docking tool specifically designed for GAGs. Based on Vina-Carb, it enhances the accuracy of modeling these carbohydrates by including parameters for sugars in the 2SO conformation and

glycosidic linkages specific to GAGs. GlycoTorch Vina also allows the integration of experimental data, such as NMR, and considers water-mediated interactions, providing more accurate predictions in the formation of GAG-protein complexes.

6. DOCK [137]: It is a molecular docking program (free for academic research) that predicts the orientation and conformation of ligands within the binding site of proteins or nucleic acids. It uses an incremental construction approach ("anchor-and-grow") to handle ligand flexibility and employs a scoring function based on the AMBER force field to evaluate the stability of the complex. DOCK is particularly useful for GAGs due to its ability to accurately model conformational flexibility and enhance sampling, allowing for more precise predictions of ligand–receptor interactions.

7. ATTRACT [138]: It is a docking (not free) program that models interactions between proteins and other biomolecules such as DNA, RNA, and small ligands. Originally designed for protein docking, it has been successfully adapted for GAGs due to its coarse-grained force field approach, which allows for protein flexibility and the simultaneous handling of multiple protein bodies [139]. This adaptability makes it particularly useful for large and dynamic complexes. Although not initially intended for GAGs, researchers have modified its protocols to account for the unique features of these molecules, such as their high flexibility and electrostatic charge. This has enabled ATTRACT to effectively predict binding poses and rank GAG-protein complexes, demonstrating its utility and versatility in advanced biological interaction studies.

The key distinctions among various docking programs stem from the specific computational search algorithms they employ and the characteristics of the scoring functions utilised to order the docked poses. Over the years a plethora of different scoring functions have been developed, in particular thanks to the evolution of machine learning and collection of high-resolution structural information [140–142]. Recently, some of them have been also optimised for evaluating the binding affinities between proteins and carbohydrates [143,144]. Among them, the CSM (cutoff scanning matrix)-carbohydrate outperforms previous methods and scoring functions, also providing a freely accessible and user-friendly web interface and an application programming interface (API) (http://biosig.unimelb.edu.au/csm_carbohydrate/).

Unravelling complex molecular interactions: tools for molecular dynamics studies

Once the conformational space accessible to the ligand has been studied, the protein binding pocket has been identified, and a

model of protein–ligand complex has been obtained, MD simulations can be performed with the aim to accurately describe the conformational and dynamic properties of the bound state. All-atom MD simulations involve 4 steps: i) energy minimisation, ii) heating, iii) equilibration and iv) production. MD simulations are often run by using explicit solvent box to account for molecular interactions, and if needed, Cl^- or Na^+ ions are added to neutralise the system. Although different programs for running MD simulations, including CP2K [145], DESMOND [146], LAMMPS [147], TINKER [148], YASARA [149], and NAMD [150] are available, the most widely used packages to run all-atom MD simulations on protein glycan complexes are the already described AMBER [56], CHARM-GUI [52] and GROMACS [63].

Tools for the analysis of computational data

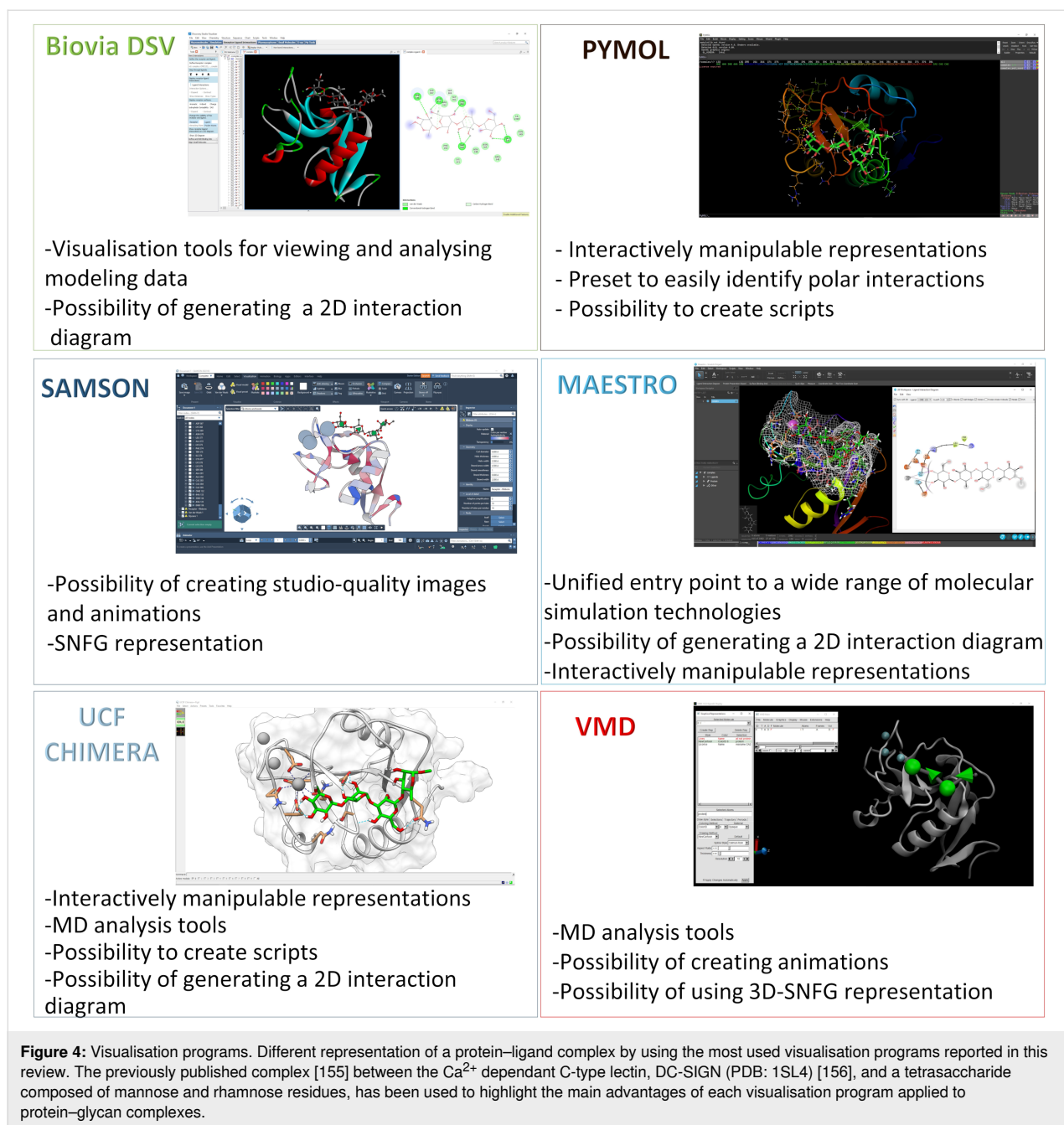
Once the protein in the apo-form has been analysed, the ligand in the free-state has been studied, and the protein–ligand complex has been extensively subjected to MD simulations, in-depth insights into the structural and conformational features governing the molecular interactions can be achieved by analysing and visualising the obtained data. The critical step of post-processing analysis involves examining the binding poses of the ligands, evaluating the stability and dynamics of the complexes, investigating the electronic structure and interactions and calculating binding energies or free energy profiles within the system, thus helping in understanding the energetics and thermodynamics of the interactions [151]. To facilitate these tasks, various software tools have been developed; generally, the simulation packages used to run all-atom MD simulations offer built-in utilities and scripts for post-processing the MD data. For example, AmberTools, released within the AMBER suite of biomolecular simulation programs, includes cptraaj and ptraj codes for analysing structure and dynamics in trajectories [152]. Additionally, other programs as VMD (see below) and PLUMED [153], an open-source library compatible with popular MD engines like Amber and GROMACS. (<https://www.plumed.org/>), can support data analysis for molecular dynamics simulations.

Moreover, several custom scripts have been developed within computational chemistry laboratories (see for example: <https://github.com/roviralab/utills>) enabling the tracking of glycan conformational changes throughout the dynamics, monitoring dihedral angles, distances, and other parameters. Additional programs allow for the combination and comparison of experimental and theoretical data, enhancing the reliability and accuracy of the simulations. As example, the software package MD2NOE [154] permits to properly simulate NOE effects also of flexible molecules sampling multiple conformational states directly from molecular dynamics (MD) trajectories. With the

advent of GPU-based simulation code, indeed, MD simulations have been extended into the microsecond regime, allowing to sample glycan conformational space sufficiently and enabling the computation of key NMR properties [154].

Different visualisation programs, as those described below, play a crucial role in rendering complex 3D structures, visualising molecular interactions, and generating high-quality images for publications or presentations (Figure 4). These user-friendly tools are indispensable for researchers in the fields of structural biology, biochemistry, and computational chemistry, making it easier to comprehend and communicate the results of sophisticated simulations.

1. VMD [54]: Visual Molecular Dynamics is a popular and freely accessible molecular modelling program designed to display, animate, and analyse biomolecular systems using 3D graphics and built-in scripting. It provides tools for simulation preparation, visualisation, and analysis of molecular dynamics. (<https://www.ks.uiuc.edu/Research/vmd/>).
2. PyMOL [157]: It is an open-source molecular visualisation system developed by Schrödinger. It is one of the most used programs for the visualisation of the 3D structure of the protein alone or in a complex with a ligand. Several tools are available to create, manipulate and visualise the 3D structures. Other tools consent to generate the surface of a protein and highlight the electrostatic potentials or hydrophobicity. PyMOL is used for various applications, such as protein structure analysis, molecular docking studies, and drug design.
3. UCSF Chimera [158]: It is a highly versatile and widely used free molecular visualisation and analysis program developed by the University of California, San Francisco (UCSF). It is a powerful software tool for visualising and analysing the 3D structures of biological macromolecules, such as proteins, nucleic acids, and other complex molecular assemblies. UCSF Chimera provides a user-friendly interface for exploring and manipulating molecular structures, offering a wide range of features for tasks like molecular modelling, molecular dynamics analysis, structural biology, and more. UCSF Chimera is commonly used to gain insights into the structure and function of biomolecules. It supports various file formats, offers diverse visualisation options, and allows for the creation of stunning images and animations of molecular structures, making it an invaluable resource (<https://www.cgl.ucsf.edu/chimera/>).
4. BIOVIA Discovery Studio [159]: It is a freely downloadable suite of science applications designed for life sciences discovery research, which includes addressing multiple optimisation objectives in drug discovery. This comprehensive software



suite, built on BIOVIA Pipeline Pilot, provides a wide range of validated applications. It offers a scalable and collaborative research environment, making it a valuable tool for life sciences discovery research (<https://discover.3ds.com/discovery-studio-visualizer-download>).

5. Schrödinger Maestro [160]: It is a comprehensive molecular modelling and computational chemistry software suite designed for researcher fields like drug discovery, materials science, and structural biology. Schrödinger Maestro provides tools for molecular visualisation, ligand-receptor docking (with

Glide [117]), molecular dynamics simulations, quantum mechanics calculations, and more. It is widely used in the pharmaceutical and biotechnology industries for drug design and discovery, as well as in academic research and other scientific applications that involve the study of molecular structures and interactions (<https://www.schrodinger.com/products/maestro>).

6. SAMSON [161]: Software for Adaptive Modelling and Simulation Of Nanosystems is a computer software platform for molecular design, unfortunately not freely available. Its modular

architecture enables a wide range of tasks, including model creation, calculations, interactive or offline simulations, and result visualisation and interpretation. Notably, SAMSON offers modules related to glycans and glycans visual models, facilitating the use of the SNFG nomenclature for ligand design and visualisation (<https://www.samson-connect.net/>).

All the computational tools here reported, summarised in Figure 5, constitute a unique kit for the analysis of protein–glycan interactions.

Computational tools applied to the study of glycans in the free state

In the years, the architectural and conformational features of different mammalian glycans, including oligomannose [162] and complex-type *N*-glycans, have been unravelled by employing computational approaches. As example, the work conducted by A. M. Harbison et al. [163] reported the molecular dynamics of complex biantennary IgG Fc *N*-glycans and their implications for the structural integrity and functionality of human immunoglobulins G (IgGs).

MD methods have also been employed, in combination with experimental methods, as NMR, to explore the three-dimensional features of bacterial glycoconjugates, as the exopolysaccharides [164], and the rough-type lipopolysaccharide [165] isolated from *Methylobacterium extorquens* or the lipopolysaccharide isolated from *Herbaspirillum* Root189 [164]. In these studies, once built the parameters for non-standard bacterial monosaccharides, which were not included in the GLYCAM-website, the overall conformation and properties of the saccharide chain has been accurately described and compared to the experimental NOE data. The tight integration between computational and experimental results allowed to highlight how modifications of the saccharidic backbone, as example with *O*-methyl and *O*-acetyl groups, can affect the polysaccharide biophysical properties tuning its ability to interact with other polymers and/or receptor proteins.

Another example of the application of simulation methods to the analysis of complex glycans is presented by Makshakova et al., who analysed the three-dimensional structure of the exopolysaccharide isolated from *Alteromonas infernus* GY785 [166]. The main chain of the so-called “Infernán” polysaccharide includes glucose, galacturonic acid and galactose, with branches composed of uronic acids and sulphate groups that contribute to modulate its unique properties. Specifically, the authors used molecular mechanics and dynamics calculations to describe the helical structure of the polysaccharide chain and the role of its side chains in the creation of Ca²⁺ chelating sites in the region of the polysaccharide branching points.

A further application of computational tools to the analysis of complex glycans is the study of the conformational behaviour of the naturally cationic polysaccharide, carboxymethyl chitosan (CMCS), reported by Zhang et al. [167]. Due to its non-toxic, biodegradable, biocompatible, and versatile features, chitosan has been widely used in various fields such as biomedicine, cosmetics, agriculture and food. However, its insolubility in neutral or alkaline pH conditions largely limits chitosan's applications. MD simulations were thus employed to mimic the behaviour of CMCS in water under different pH values and different degrees of deacetylation and substitutions in order to study its aggregation pattern.

Computational tools applied to the study of proteins in the free state

Understanding the dynamics of proteins in their free state is key to investigate how they can interact with other biomolecules. MD simulations can be used to study the conformational changes that occur in proteins as they move from one state to another, which is important for understanding their function; additionally, MD simulations can also be used to study the thermodynamics of protein folding, which is important for understanding how proteins fold into their native state.

Recently, extensive MD studies have been performed to investigate the structural and functional features of the SARS-CoV-2 spike glycoprotein, allowing to reveal the critical role of glycans attached to the viral protein in the infection. In particular, the full spike receptor consists of a trimer of S protein, and each monomer comprises 22 *N*-glycan sites. Recent studies suggest that the structure and occupancy of the SARS-CoV-2 S glycans affect the structural integrity of the trimer [168]. Specifically, different Spike protein models with varying glycan compositions in positions N234, N165 and N343 were created by homology modelling using SWISS-MODEL. Then *N*-glycans were added by aligning conformationally equilibrated *N*-glycan structures from a Glyco Shape library to the GlcNAc residues resolved in the cryo-EM structure, with adjustments of the torsion angles to resolve steric clashes with the surrounding protein. By using AMBER, the MD simulation was performed and suggested that diminishing the size of *N*-glycans at position N234 results in destabilising the “wide-open” conformation of the receptor-binding domain (RBD). This destabilisation leads to increased RBD dynamics. Furthermore, the composition of *N*-glycans at positions N165 and N343 influenced the stability of the open RBD, where shorter structures exhibited reduced effectiveness in interacting with the disordered loop within the receptor-binding motif.

Moreover, several groups have employed MD simulation to design new proteins. For example, the group of Mayo [169] em-

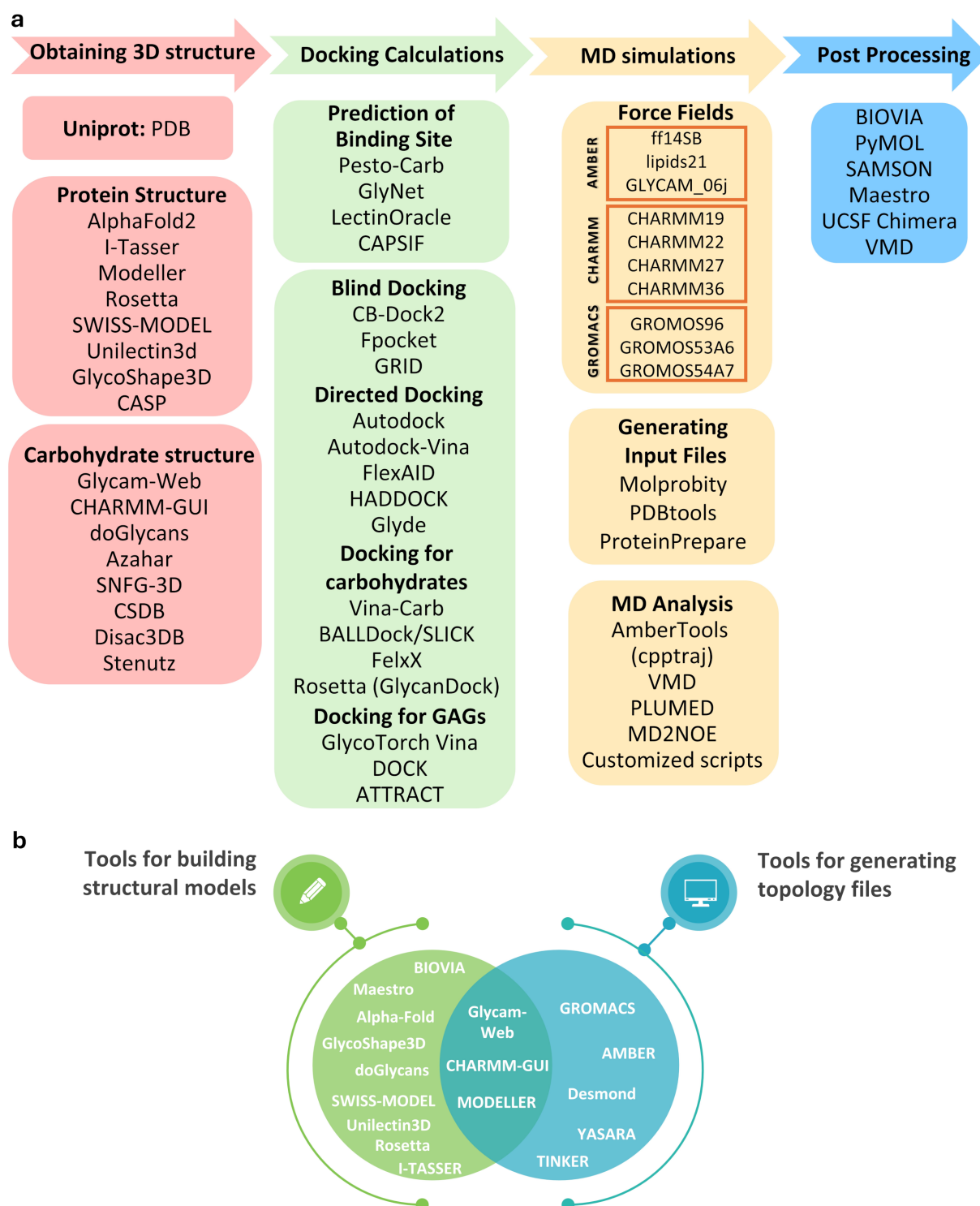


Figure 5: A schematic representation of useful computational methods to study protein–glycan interactions. a) Workflow including different key steps needed to analyse protein–glycan interactions: 1. Building/choosing the appropriate glycan/protein 3D structure; 2. Modelling protein–glycan complex; 3. Running and analysing MD simulations; 4. Processing and visualising the results. b) Summary of the presented tools for building structural model and/or generating topology files of glycan/protein structures.

ployed this approach to engineer a de novo homodimer from a monomeric protein. The integration of computational protein design (CPD) and MD simulation allowed to refine the structural and dynamic aspects of the designed proteins overcoming CPD inherent limitations, including constraints on side chain rotamers, fixed protein backbones, and a lack of consideration for solvent interactions. Thus, the use of MD simulation allowed to provide a more precise depiction of the protein's behaviour, shedding light on its dynamics and stability.

Another example of the relevance of all-atom molecular dynamics simulations is the study performed by X. Cao et al. [170], which reports the structural dynamics of GH33 sialidases. The computational analysis revealed significant conformational rearrangements within the enzyme active sites leading to the formation of a new cleft to accommodate glycosyl acceptors. Furthermore, the simulations shed light on the role of specific residues within the enzyme's active site, such as the arginine triad and other key residues, which adjusted their conformations to interact with sialic acid and facilitate the opening of a new cleft. Computational tools, including GROMACS and AutoDock, were pivotal in uncovering key insights into the catalytic mechanisms of GH33 sialidases offering a promising avenue for the rational design of improved biocatalysts.

Computational tools applied to the study of protein–glycan interactions

Molecular dynamics simulation has proven to be a powerful tool in understanding and elucidating the intricate dynamics of glycan interactions with biomolecules. This computational technique allows researchers to delve deep into the molecular-level details of how glycans bind to their respective target proteins, providing valuable insights into binding mechanisms, thermodynamics, and the overall stability of protein–glycan complexes.

MD methods have been extensively employed by different groups to explore glycan recognition by host receptors, including mammalian and bacterial proteins. For example, several studies have been published on the recognition of sialic acids by different classes of proteins. It is known that sialic acid plays an essential role in the modulation of immune response through the binding with sialic acid-binding immunoglobulin-like lectins (SIGLEC). In the study reported by Martin Frank et al. [125], MD were used to analyse the binding modes of several glycomimetics for Siglec-7 and describe their molecular interactions at atomic level. The conformational characteristics of both natural, unmodified, and synthetic, modified α -sialoside glycerol sidechains of sialic acid were investigated. The applied computational tools allowed to discover a new modification in the sialic acid glycerol chain that binds to Siglec-7,

providing a basis for designing next-generation Siglec-7 ligands.

Sialic acid can also be recognised from some bacterial proteins which exploit this interaction to adhere to host cells during the first stages of infection. An example is given by the sialic acid-binding serine-rich repeat adhesins from *Streptococci*, which contain a sialic acid-binding region (SLBR) and are known as Siglec-like adhesins. Di Carluccio et al. [171] described the interactions between two different siglec-like adhesins with natural glycans by using a combination of NMR and MD simulations. This integrated approach allowed to accurately describe the different selectivity and flexibility of the proteins towards sialoglycans recognition and binding, providing a privileged starting point for the design and development of novel compounds to counteract streptococcal infections by inhibiting bacterial adherence to host tissues.

Another example of the application of MD simulations in studying bacterial proteins in the interaction with glycans comes from Bernardi's group [172], whose focus was on examining the interaction between different glycomimetic antagonists and BC2L-C lectin derived from *B. cenocepacia*. The MD results showed that the binding site at the interface of two BC2L-C-Nt monomers is pre-organised to host the bifunctional ligands. Additionally, the simulation with the water molecules highlights the importance of two of these molecules in the binding site, establishing an interaction network.

Bacterial glycoconjugates, as lipopolysaccharides-related systems, have also been dissected, gaining critical information about the ability of LPS to both stimulate the host immune system, mainly by interacting with TLR-4/MD-2 complex, and interact with several molecules. The Martin-Santamaria group [173] extensively contributed to increase the knowledge on this topic, analysing the conformational changes of the TLR4/MD2 complex when interacting either with small and LPS-like molecules

Notably, the comparison of free and bound state MD results can allow to determine critical differences in the glycan conformational behaviour upon binding with selected proteins, paving the way for the design of tailored synthetic inhibitors and therapeutics. As example, in the study of L. Pirone et al. [174], computational techniques were combined with biophysical and spectroscopic methods to investigate the interaction between a selenoglycoside (SeDG) and galectins Gal-1 or Gal-3CRD. The integration of data from NMR, CD, and ITC provided valuable insights into designing selective inhibitors. The computational studies uncovered two different binding modes: when bound to Gal-1, SeDG adopted a V-shaped conformation driven by van

der Waals interactions; on the contrary, when in complex with Gal-3CRD, it assumed an extended conformation. Comparing these modes identified specific interaction sites, guiding the design of selective inhibitors that can differentiate between the two galectins.

Noteworthy several computational studies have been conducted also for exploring protein-GAG interactions [175,176]. The study conducted by U. Uciechowska-Kaczmarzyk et al. [139] reports an extensive evaluation of protein-GAG complexes using a dataset of 28 complexes where the GAG length exceeded DP3 [139]. Through various statistical analyses to differentiate and highlight the docking programs with superior performance, valuable insights were provided into the most effective tools for studying these biologically relevant systems [177]. The interaction between the chemokine CXCL8/IL-8 and heparin-derived oligosaccharides was investigated by applying these docking procedures together with NMR spectroscopic techniques demonstrating the that higher affinity of the CXCL8 dimer for GAGs compared to the monomer and highlighting the structural plasticity that allows multiple binding modes. The use of HADDOCK in this context underscored its capability to model complex protein-GAG interactions accurately, providing a detailed understanding of the binding mechanisms at play [178].

Conclusion

In structural biology, the investigation of protein-glycan interactions often relies on applying various structural techniques, including NMR, X-ray crystallography, and cryo-EM. Each of these methodologies comes with distinct advantages and limitations. NMR is particularly valuable for its ability to dynamically study molecules at the atomic level while preserving sample integrity. This makes it especially suitable for studying carbohydrates, offering insights into their 3D structures and conformations, but it generates a huge amount of data, which can be challenging to interpret effectively. X-ray crystallography provides high-resolution structural information, but unfortunately, this technique often fails when investigating carbohydrate-protein interactions due to the intrinsic flexibility of sugars, rendering them invisible in the density maps. In recent years, cryo-EM has seen widespread adoption in solving protein structures and glycoconjugates, thanks to significant advancements in instrumentation. Nevertheless, a notable limitation of cryo-EM lies in its capacity to handle large, intricate complexes. In this context, computational approaches can be valuable allies to develop accurate models helping in integrating and rationalizing data obtained from different methods and bridging the gap between the insights obtained from experimental data and the detailed understanding of complex biological systems. As example, models of protein and ligand, both in

the free and bound states, can assist not only the interpretation of NMR spectra but also the building of structures that satisfy experimentally derived distance and angle restraints. Moreover, in X-ray crystallography and cryo-EM, protein models can be used to provide accurate templates for molecular replacement in the crystal cell or for backbone tracing and fitting sequence into a map, respectively.

We provided here an overview of computational tools available for ligand and protein building as well as the analysis of their molecular interactions, with a special focus on carbohydrates (Figure 5). Generally, to allow the prediction of an accurate 3D model of protein-glycan complexes, the combined use of different tools is highly recommended. A typical workflow could include firstly research to investigate the favoured carbohydrates bound to a protein (i.e., by using Glynet or LectinOracle), then other tools (such as CASPIF or PESTO) can be employed to predict the binding location. Subsequently, appropriate docking software (i.e., AutoDock Vina-Carb) can be used to provide a model of protein-glycan complex, which can be further refined (as example thanks to GlycanDock) and explored by molecular dynamic simulations (i.e., by using AMBER). Finally, the detailed analysis of the trajectory (i.e., by using AmberTools) provides unique vision of the 3D structure and real dynamics of glycan motifs in the bound state. Notably, the recent fusion of cutting-edge technologies, such as virtual reality, with interactive molecular simulations also allows to create an immersive environment, offering an opportunity without precedents to explore and manipulate molecular systems in real-time [179].

However, it is worth to note that, despite the continuous improvement of computational techniques and force fields development, there are still some limitations in the application of bioinformatic methods to the analysis of biomolecular interactions, especially in the case of complex carbohydrates, not to speak about bacterial glycans. Step forwards have been done in the improvement of docking programs dedicated to carbohydrates, however, the available software performed better for smaller glycans, while additional glycosidic linkages still remain a big challenge for docking calculations, and there is still room for improvement in ranking the best sugar docking pose. Additionally, over the years, different carbohydrate-specific force fields have been developed, the choice of which varies depending on the preferred simulation conditions, however, only few parameters have been defined for peculiar bacterial monosaccharides hampering a user-friendly and not time-consuming analysis of the system via MD simulations. A step change in this direction will permit the integrated use of valuable bioinformatic tools tailored on carbohydrates and would be of great help in unveiling critical structural and

conformational features, at the atomic level, of complex glycans in the free and bound state, that can serve as essential resources for structural glycomics research to both experts and non-experts in glycobiology.

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement no 851356 to R.M.

The work described in this paper was supported by Italian MUR, PRIN 2020, Project SEA-WAVE 2020BKK3W9. FSE, PON Ricerca e Innovazione Azione I.1 “Dottorati Innovativi con caratterizzazione Industriale” is acknowledged for funding the Ph.D. grant to A.M.

Author Contributions

Ferran Nieto-Fabregat: conceptualization; writing – original draft; writing – review & editing. Maria Pia Lenza: conceptualization; writing – original draft; writing – review & editing. Angela Marseglia: writing – review & editing. Cristina Di Carluccio: writing – review & editing. Antonio Molinaro: writing – review & editing. Alba Silipo: writing – review & editing. Roberta Marchetti: conceptualization; writing – original draft; writing – review & editing.

ORCID® iDs

Ferran Nieto-Fabregat - <https://orcid.org/0000-0001-9847-3030>

Maria Pia Lenza - <https://orcid.org/0000-0002-9733-5020>

Angela Marseglia - <https://orcid.org/0000-0003-1831-6831>

Cristina Di Carluccio - <https://orcid.org/0000-0001-5895-9829>

Antonio Molinaro - <https://orcid.org/0000-0002-3456-7369>

Alba Silipo - <https://orcid.org/0000-0002-5394-6532>

Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed in this study.

References

- Kuo, J. C.-H.; Paszek, M. J. *Annu. Rev. Cell Dev. Biol.* **2021**, *37*, 257–283. doi:10.1146/annurev-cellbio-120219-054401
- Flynn, R. A.; Pedram, K.; Malaker, S. A.; Batista, P. J.; Smith, B. A. H.; Johnson, A. G.; George, B. M.; Majzoub, K.; Villalta, P. W.; Carette, J. E.; Bertozzi, C. R. *Cell* **2021**, *184*, 3109–3124.e22. doi:10.1016/j.cell.2021.04.023
- Varki, A. *Glycobiology* **2017**, *27*, 3–49. doi:10.1093/glycob/cww086
- Varki, A.; Kornfeld, S.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds.; Cold Spring Harbor: New York, NY, USA, 2022; pp 61–91.
- Hart, G. W.; Copeland, R. J. *Cell* **2010**, *143*, 672–676. doi:10.1016/j.cell.2010.11.008
- Rudd, P. M.; Karlsson, N. G.; Khoo, K.-H.; Thaysen-Andersen, M.; Wells, L.; Packer, N. H. *Glycomics and Glycoproteomics. Essentials of Glycobiology*; Cold Spring Harbor: New York, NY, USA, 2022; pp 1188–1213.
- Marchetti, R.; Forgione, R. E.; Fabregat, F. N.; Di Carluccio, C.; Molinaro, A.; Silipo, A. *Curr. Opin. Struct. Biol.* **2021**, *68*, 74–83. doi:10.1016/j.sbi.2020.12.003
- Sarkar, A.; Drouillard, S.; Rivet, A.; Perez, S. *Glycobiology* **2015**, *25*, 1480–1490. doi:10.1093/glycob/cwv054
- Plazinski, W.; Plazinska, A. *Pure Appl. Chem.* **2017**, *89*, 1283–1294. doi:10.1515/pac-2016-0922
- Fadda, E. *Curr. Opin. Chem. Biol.* **2022**, *69*, 102175. doi:10.1016/j.cbpa.2022.102175
- Neelamegham, S.; Aoki-Kinoshita, K.; Bolton, E.; Frank, M.; Lisacek, F.; Lütke, T.; O'Boyle, N.; Packer, N. H.; Stanley, P.; Toukach, P.; Varki, A.; Woods, R. J.; The SNFG Discussion Group. *Glycobiology* **2019**, *29*, 620–624. doi:10.1093/glycob/cwz045
- Turnbull, J. E.; Field, R. A. *Nat. Chem. Biol.* **2007**, *3*, 74–77. doi:10.1038/nchembio0207-74
- Hricovini, M. *Curr. Med. Chem.* **2004**, *11*, 2565–2583. doi:10.2174/0929867043364414
- Frank, M. Computational Docking as a Tool for the Rational Design of Carbohydrate-Based Drugs. In *Carbohydrates as drugs*; Seeberger, P. H.; Rademacher, C., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp 53–72. doi:10.1007/7355_2014_42
- Mallik, B.; Morikis, D. *Curr. Proteomics* **2006**, *3*, 259–270. doi:10.2174/157016406780655568
- Kuttel, M. M.; Ravenscroft, N. The Role of Molecular Modeling in Predicting Carbohydrate Antigen Conformation and Understanding Vaccine Immunogenicity. *Carbohydrate-based vaccines: from concept to clinic*; American Chemical Society: Washington, DC, USA, 2018; pp 139–173. doi:10.1021/bk-2018-1290.ch007
- Baum, L. G.; Cobb, B. A. *Glycobiology* **2017**, *27*, 619–624. doi:10.1093/glycob/cwx036
- Wexler, A. G.; Goodman, A. L. *Nat. Microbiol.* **2017**, *2*, 17026. doi:10.1038/nmicrobiol.2017.26
- Di Lorenzo, F.; Duda, K. A.; Lanzetta, R.; Silipo, A.; De Castro, C.; Molinaro, A. *Chem. Rev.* **2022**, *122*, 15767–15821. doi:10.1021/acs.chemrev.0c01321
- Rauch, J.; Volinsky, N.; Romano, D.; Kolch, W. *Cell Commun. Signaling* **2011**, *9*, 23. doi:10.1186/1478-811x-9-23
- Wang, Z.; Cole, P. A. *Methods Enzymol.* **2014**, *548*, 1–21. doi:10.1016/b978-0-12-397918-6.00001-x
- Yang, N. J.; Hinner, M. J. Getting Across the Cell Membrane: An Overview for Small Molecules, Peptides, and Proteins. In *Site-Specific Protein Labeling. Methods in Molecular Biology*; Gautier, A.; Hinner, M., Eds.; Humana Press: New York, NY, USA, 2015; Vol. 1266, pp 29–53. doi:10.1007/978-1-4939-2272-7_3
- Ford, K. G.; Souberbielle, B. E.; Darling, D.; Farzaneh, F. *Gene Ther.* **2001**, *8*, 1–4. doi:10.1038/sj.gt.3301383
- Litvinov, R. I.; Weisel, J. W. *Matrix Biol.* **2017**, *60–61*, 110–123. doi:10.1016/j.matbio.2016.08.003
- Pollard, T. D. *Cold Spring Harbor Perspect. Biol.* **2016**, *8*, a018226. doi:10.1101/cshperspect.a018226
- Watford, M.; Wu, G. *Adv. Nutr.* **2018**, *9*, 651–653. doi:10.1093/advances/nmy027

27. Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. *Nat. Rev. Nephrol.* **2019**, *15*, 346–366. doi:10.1038/s41581-019-0129-4
28. Schwarz, F.; Aebi, M. *Curr. Opin. Struct. Biol.* **2011**, *21*, 576–582. doi:10.1016/j.sbi.2011.08.005
29. Magalhães, A.; Duarte, H. O.; Reis, C. A. *Mol. Aspects Med.* **2021**, *79*, 100964. doi:10.1016/j.mam.2021.100964
30. Bloch, J. S.; John, A.; Mao, R.; Mukherjee, S.; Boilevin, J.; Irobalieva, R. N.; Darbre, T.; Scott, N. E.; Reymond, J.-L.; Kossiakoff, A. A.; Goddard-Borger, E. D.; Locher, K. P. *Nat. Chem. Biol.* **2023**, *19*, 575–584. doi:10.1038/s41589-022-01219-9
31. Noborn, F.; Nilsson, J.; Larson, G. *Matrix Biol.* **2022**, *111*, 289–306. doi:10.1016/j.matbio.2022.07.002
32. Fujimoto, Z.; Tateno, H.; Hirabayashi, J. Lectin Structures: Classification Based on the 3-D Structures. In *Lectins: Methods and Protocols*; Hirabayashi, J., Ed.; Springer: New York, NY, USA, 2014; pp 579–606. doi:10.1007/978-1-4939-1292-6_46
33. Varki, A.; Etzler, M. E.; Cummings, R. D.; Esko, J. D. *Glycobiology* **2009**, *4*, 28.
34. Napoletano, C.; Zizzari, I. G.; Rugghetti, A.; Rahimi, H.; Irimura, T.; Clausen, H.; Wandall, H. H.; Belleudi, F.; Bellati, F.; Pierelli, L.; Frati, L.; Nuti, M. *Eur. J. Immunol.* **2012**, *42*, 936–945. doi:10.1002/eji.201142086
35. Temme, J. S.; Butler, D. L.; Gildersleeve, J. C. *Biochem. J.* **2021**, *478*, 1485–1509. doi:10.1042/bcj20200610
36. Gimeno, A.; Valverde, P.; Ardá, A.; Jiménez-Barbero, J. *Curr. Opin. Struct. Biol.* **2020**, *62*, 22–30. doi:10.1016/j.sbi.2019.11.004
37. Maldonado-Hernández, R.; Quesada, O.; González-Feliciano, J. A.; Baerga-Ortiz, A.; Lasalde-Dominicci, J. A. *Proteomics* **2024**, *24*, 2300151. doi:10.1002/pmic.202300151
38. Pandey, B.; S., S.; Chatterjee, A.; Mangala Prasad, V. *Proteins: Struct., Funct., Bioinf.* **2024**, in press. doi:10.1002/prot.26636
39. Fontana, C.; Widmalm, G. *Chem. Rev.* **2023**, *123*, 1040–1102. doi:10.1021/acs.chemrev.2c00580
40. Wilkinson, H.; Saldova, R. J. *Proteome Res.* **2020**, *19*, 3890–3905. doi:10.1021/acs.jproteome.0c00435
41. Angulo, J.; Zimmer, J.; Imberty, A.; Prestegard, J. In *Essentials of Glycobiology*; Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds.; Cold Spring Harbor: New York, NY, USA, 2022; pp 61–91.
42. Homans, S. W.; Pastore, A.; Dwek, R. A.; Rademacher, T. W. *Biochemistry* **1987**, *26*, 6649–6655. doi:10.1021/bi00395a014
43. Dauchez, M.; Mazurier, J.; Montreuil, J.; Spik, G.; Vergoten, G. *Biochimie* **1992**, *74*, 63–74. doi:10.1016/0300-9084(92)90185-h
44. Sinha, S.; Tam, B.; Wang, S. M. *Membranes* **2022**, *12*, 844. doi:10.3390/membranes12090844
45. Varki, A.; Cummings, R. D.; Aebi, M.; Packer, N. H.; Seeberger, P. H.; Esko, J. D.; Stanley, P.; Hart, G.; Darvill, A.; Kinoshita, T.; Prestegard, J. J.; Schnaar, R. L.; Freeze, H. H.; Marth, J. D.; Bertozzi, C. R.; Etzler, M. E.; Frank, M.; Vliegthart, J. F.; Lütke, T.; Perez, S.; Bolton, E.; Rudd, P.; Paulson, J.; Kanehisa, M.; Toukach, P.; Aoki-Kinoshita, K. F.; Dell, A.; Narimatsu, H.; York, W.; Taniguchi, N.; Kornfeld, S. *Glycobiology* **2015**, *25*, 1323–1324. doi:10.1093/glycob/cwv091
46. Cousins, K. R. *J. Am. Chem. Soc.* **2005**, *127*, 4115–4116. doi:10.1021/ja0410237
47. Lal, K.; Bermeo, R.; Perez, S. *Beilstein J. Org. Chem.* **2020**, *16*, 2448–2468. doi:10.3762/bjoc.16.199
48. Perez, S.; Fadda, E.; Makshakova, O. Computational Modeling in Glycoscience. *Comprehensive Glycoscience*, 2nd ed.; Elsevier: Amsterdam, Netherlands, 2021; pp 374–404. doi:10.1016/b978-0-12-819475-1.00004-3
49. Danne, R.; Poojar, C.; Martinez-Seara, H.; Rissanen, S.; Lolicato, F.; Róg, T.; Vattulainen, I. *J. Chem. Inf. Model.* **2017**, *57*, 2401–2406. doi:10.1021/acs.jcim.7b00237
50. Woods Group (2005–2024); GLYCAM Web.
51. Singh, A.; Montgomery, D.; Xue, X.; Foley, B. L.; Woods, R. J. *Glycobiology* **2019**, *29*, 515–518. doi:10.1093/glycob/cwz027
52. Jo, S.; Kim, T.; Iyer, V. G.; Im, W. J. *Comput. Chem.* **2008**, *29*, 1859–1865. doi:10.1002/jcc.20945
53. Arroyuelo, A.; Vila, J. A.; Martin, O. A. J. *Comput.-Aided Mol. Des.* **2016**, *30*, 619–624. doi:10.1007/s10822-016-9944-x
54. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38. doi:10.1016/0263-7855(96)00018-5
55. Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. *J. Chem. Theory Comput.* **2011**, *7*, 4026–4037. doi:10.1021/ct200196m
56. AMBER; University of California: San Francisco, CA, USA, 2018.
57. Dickson, C. J.; Walker, R. C.; Gould, I. R. *J. Chem. Theory Comput.* **2022**, *18*, 1726–1736. doi:10.1021/acs.jctc.1c01217
58. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. J. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. doi:10.1021/acs.jctc.5b00255
59. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935. doi:10.1063/1.445869
60. He, X.; Man, V. H.; Yang, W.; Lee, T.-S.; Wang, J. *J. Chem. Phys.* **2020**, *153*, 114502. doi:10.1063/5.0019056
61. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655. doi:10.1002/jcc.20820
62. Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W., II; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2011**, *7*, 3162–3180. doi:10.1021/ct200328p
63. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, *1–2*, 19–25. doi:10.1016/j.softx.2015.06.001
64. Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2009**, *5*, 2353–2370. doi:10.1021/ct900242e
65. Pol-Fachin, L.; Verli, H.; Lins, R. D. *J. Comput. Chem.* **2014**, *35*, 2087–2095. doi:10.1002/jcc.23721
66. Allinger, N. L.; Rahman, M.; Lii, J. H. *J. Am. Chem. Soc.* **1990**, *112*, 8293–8307. doi:10.1021/ja00179a012
67. Dauchez, M.; Derreumaux, P.; Lagant, P.; Vergoten, G. *J. Comput. Chem.* **1995**, *16*, 188–199. doi:10.1002/jcc.540160206
68. Hatcher, E.; Guvench, O.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2009**, *113*, 12466–12476. doi:10.1021/jp905496e
69. Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832–3846. doi:10.1021/j100011a061
70. Vanqualef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y. *Nucleic Acids Res.* **2011**, *39* (Suppl. 2), W511–W517. doi:10.1093/nar/gkr288
71. Zheng, S.; Tang, Q.; He, J.; Du, S.; Xu, S.; Wang, C.; Xu, Y.; Lin, F. *J. Chem. Inf. Model.* **2016**, *56*, 811–818. doi:10.1021/acs.jcim.5b00687

72. Kamerling, J. P. Basics Concepts and Nomenclature Recommendations in Carbohydrate Chemistry. In *Comprehensive Glycoscience*; Kamerling, H., Ed.; Elsevier: Oxford, UK, 2007; Vol. 1, pp 1–38. doi:10.1016/b978-044451967-2/00001-5
73. Yu, Y.; Delbianco, M. *Chem. – Eur. J.* **2020**, *26*, 9814–9825. doi:10.1002/chem.202001370
74. Xu, B.; Unione, L.; Sardinha, J.; Wu, S.; Ethève-Quelejeu, M.; Pilar Rauter, A.; Blériot, Y.; Zhang, Y.; Martín-Santamaría, S.; Díaz, D.; Jiménez-Barbero, J.; Sollogoub, M. *Angew. Chem.* **2014**, *126*, 9751–9756. doi:10.1002/ange.201405008
75. Toukach, P. V.; Egorova, K. S. *Nucleic Acids Res.* **2016**, *44*, D1229–D1236. doi:10.1093/nar/gkv840
76. Toukach, P. V.; Egorova, K. S. Bacterial, Plant, and Fungal Carbohydrate Structure Databases: Daily Usage. In *Glycoinformatics*; Lütke, T.; Frank, M., Eds.; Methods in Molecular Biology, Vol. 1273; Humana Press: New York, NY, USA, 2015; pp 55–85. doi:10.1007/978-1-4939-2343-4_5
77. Egorova, K. S.; Toukach, P. V. Carbohydrate Structure Database (CSDB): Examples of Usage. In *A Practical Guide to Using Glycomics Databases*; Aoki-Kinoshita, K., Ed.; Springer: Tokyo, Japan, 2017; pp 75–113. doi:10.1007/978-4-431-56454-6_5
78. wwPDB consortium. *Nucleic Acids Res.* **2019**, *47*, D520–D528. doi:10.1093/nar/gky949
79. Cheng, H.; Abed, A. M.; Alizadeh, A.; Ghabra, A. A.; Altalbawy, F. M. A.; Sabetvand, R.; Smaism, G. F.; Yadav, A.; Toghraie, D.; Riadi, Y. *J. Mol. Liq.* **2023**, *369*, 120893. doi:10.1016/j.molliq.2022.120893
80. Yu, Y.; Xu, S.; He, R.; Liang, G. *J. Agric. Food Chem.* **2023**, *71*, 2684–2703. doi:10.1021/acs.jafc.2c06789
81. Wang, Y.; Dong, B.; Wang, D.; Jia, X.; Zhang, Q.; Liu, W.; Zhou, H. *Appl. Sci.* **2023**, *13*, 7287. doi:10.3390/app13127287
82. Sabe, V. T.; Ntombela, T.; Jhamba, L. A.; Maguire, G. E. M.; Govender, T.; Naicker, T.; Kruger, H. G. *Eur. J. Med. Chem.* **2021**, *224*, 113705. doi:10.1016/j.ejmech.2021.113705
83. Lokhande, K. B.; Pawar, S. V.; Madkaiker, S.; Nawani, N.; Venkateswara, S. K.; Ghosh, P. J. *Biomol. Struct. Dyn.* **2023**, *41*, 2698–2712. doi:10.1080/07391102.2022.2038271
84. Hisama, K.; Valadez Huerta, G.; Koyama, M. *Comput. Mater. Sci.* **2023**, *218*, 111955. doi:10.1016/j.commatsci.2022.111955
85. Günay, M. G.; Kemerli, U.; Karaman, C.; Karaman, O.; Güngör, A.; Karimi-Maleh, H. *Environ. Res.* **2023**, *217*, 114785. doi:10.1016/j.envres.2022.114785
86. Garcia-Alai, M. M.; Heidemann, J.; Skruzny, M.; Gieras, A.; Mertens, H. D. T.; Svergun, D. I.; Kaksonen, M.; Uetrecht, C.; Meijers, R. *Nat. Commun.* **2018**, *9*, 328. doi:10.1038/s41467-017-02443-x
87. Mufassirin, M. M. M.; Newton, M. A. H.; Sattar, A. *Artif. Intell. Rev.* **2023**, *56*, 7665–7732. doi:10.1007/s10462-022-10350-x
88. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohli, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. *Nature* **2021**, *596*, 583–589. doi:10.1038/s41586-021-03819-2
89. Zhou, X.; Zheng, W.; Li, Y.; Pearce, R.; Zhang, C.; Bell, E. W.; Zhang, G.; Zhang, Y. *Nat. Protoc.* **2022**, *17*, 2326–2353. doi:10.1038/s41596-022-00728-0
90. Webb, B.; Sali, A. *Curr. Protoc. Bioinf.* **2016**, *54*, 5.6.1–5.6.37. doi:10.1002/cpbi.3
91. Leman, J. K.; Weitzner, B. D.; Lewis, S. M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R.; Aprahamian, M.; Baker, D.; Barlow, K. A.; Barth, P.; Basanta, B.; Bender, B. J.; Blacklock, K.; Bonet, J.; Boyken, S. E.; Bradley, P.; Bystroff, C.; Conway, P.; Cooper, S.; Correia, B. E.; Coventry, B.; Das, R.; De Jong, R. M.; DiMaio, F.; Dsilva, L.; Dunbrack, R.; Ford, A. S.; Frenz, B.; Fu, D. Y.; Geniesse, C.; Goldschmidt, L.; Gowthaman, R.; Gray, J. J.; Gront, D.; Guffy, S.; Horowitz, S.; Huang, P.-S.; Huber, T.; Jacobs, T. M.; Jelliazkov, J. R.; Johnson, D. K.; Kappel, K.; Karanicolas, J.; Khakzad, H.; Khar, K. R.; Khare, S. D.; Khatib, F.; Khramushin, A.; King, I. C.; Kleffner, R.; Koepnick, B.; Kortemme, T.; Kuenze, G.; Kuhlman, B.; Kuroda, D.; Labonte, J. W.; Lai, J. K.; Lapidoth, G.; Leaver-Fay, A.; Lindert, S.; Linsky, T.; London, N.; Lubin, J. H.; Lyskov, S.; Maguire, J.; Malmström, L.; Marcos, E.; Marcu, O.; Marze, N. A.; Meiler, J.; Moretti, R.; Mulligan, V. K.; Nerli, S.; Norn, C.; Ó'Conchúir, S.; Ollikainen, N.; Ovchinnikov, S.; Pacella, M. S.; Pan, X.; Park, H.; Pavlovicz, R. E.; Pethe, M.; Pierce, B. G.; Pilla, K. B.; Raveh, B.; Renfrew, P. D.; Burman, S. S. R.; Rubenstein, A.; Sauer, M. F.; Scheck, A.; Schief, W.; Schueler-Furman, O.; Sedan, Y.; Sevy, A. M.; Sgourakis, N. G.; Shi, L.; Siegel, J. B.; Silva, D.-A.; Smith, S.; Song, Y.; Stein, A.; Szegedy, M.; Teets, F. D.; Thyme, S. B.; Wang, Y.-R.; Watkins, A.; Zimmerman, L.; Bonneau, R. *Methods* **2020**, *17*, 665–680. doi:10.1038/s41592-020-0848-2
92. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. *Science* **2021**, *373*, 871–876. doi:10.1126/science.abj8754
93. Moretti, R.; Lyskov, S.; Das, R.; Meiler, J.; Gray, J. J. *Protein Sci.* **2018**, *27*, 259–268. doi:10.1002/pro.3313
94. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. *Nucleic Acids Res.* **2018**, *46*, W296–W303. doi:10.1093/nar/gky427
95. Bonnardel, F.; Mariethoz, J.; Salentin, S.; Robin, X.; Schroeder, M.; Perez, S.; Lisacek, F.; Imbert, A. *Nucleic Acids Res.* **2019**, *47*, D1236–D1244. doi:10.1093/nar/gky832
96. Bonnardel, F.; Perez, S.; Lisacek, F.; Imbert, A. Structural Database for Lectins and the UniLectin Web Platform. In *Lectin Purification and Analysis: Methods and Protocols*; Hirabayashi, J., Ed.; Springer: New York, NY, USA, 2020; pp 1–14. doi:10.1007/978-1-0716-0430-4_1
97. Ives, C. M.; Singh, O.; D'Andrea, S.; Fogarty, C. A.; Harbison, A. M.; Satheesan, A.; Tropea, B.; Fadda, E. *bioRxiv* **2023**. doi:10.1101/2023.12.11.571101
98. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283–291. doi:10.1107/s0021889892009944
99. Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moutl, J. *Proteins: Struct., Funct., Bioinf.* **2023**, *91*, 1539–1549. doi:10.1002/prot.26617

100. Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V.; Keedy, D. A.; Hintze, B. J.; Chen, V. B.; Jain, S.; Lewis, S. M.; Arendall, W. B., III; Snoeyink, J.; Adams, P. D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. *Protein Sci.* **2018**, *27*, 293–315. doi:10.1002/pro.3330
101. Rodrigues, J. P. G. L. M.; Teixeira, J. M. C.; Trellet, M.; Bonvin, A. M. J. *J. F1000Research* **2018**, *7*, 1961. doi:10.12688/f1000research.17456.1
102. Martínez-Rosell, G.; Giorgino, T.; De Fabritiis, G. *J. Chem. Inf. Model.* **2017**, *57*, 1511–1516. doi:10.1021/acs.jcim.7b00190
103. Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. *Curr. Comput.-Aided Drug Des.* **2011**, *7*, 146–157. doi:10.2174/157340911795677602
104. Zhao, H.; Taherzadeh, G.; Zhou, Y.; Yang, Y. *Curr. Protoc. Protein Sci.* **2018**, *94*, e75. doi:10.1002/cpp.75
105. Bibekar, P.; Krapp, L.; Peraro, M. D. *J. Chem. Theory Comput.* **2024**, *20*, 2985–2991. doi:10.1021/acs.jctc.3c01145
106. Krapp, L. F.; Abriata, L. A.; Cortés Rodríguez, F.; Dal Peraro, M. *Nat. Commun.* **2023**, *14*, 2175. doi:10.1038/s41467-023-37701-8
107. Carpenter, E. J.; Seth, S.; Yue, N.; Greiner, R.; Derda, R. *Chem. Sci.* **2022**, *13*, 6669–6686. doi:10.1039/d1sc05681f
108. Lundstrøm, J.; Korhonen, E.; Lisacek, F.; Bojar, D. *Adv. Sci.* **2022**, *9*, 2103807. doi:10.1002/adv.202103807
109. Canner, S. W.; Shanker, S.; Gray, J. J. *Front. Bioinform.* **2023**, *3*, 1186531. doi:10.3389/fbinf.2023.1186531
110. Feinstein, W. P.; Brylinski, M. *J. Cheminf.* **2015**, *7*, 18. doi:10.1186/s13321-015-0067-5
111. Yang, C.; Chen, E. A.; Zhang, Y. *Molecules* **2022**, *27*, 4568. doi:10.3390/molecules27144568
112. Hassan, N. M.; Alhossary, A. A.; Mu, Y.; Kwok, C.-K. *Sci. Rep.* **2017**, *7*, 15451. doi:10.1038/s41598-017-15571-7
113. Liu, Y.; Grimm, M.; Dai, W.-t.; Hou, M.-c.; Xiao, Z.-X.; Cao, Y. *Acta Pharmacol. Sin.* **2020**, *41*, 138–144. doi:10.1038/s41401-019-0228-6
114. Le Guilloux, V.; Schmidtke, P.; Tuffery, P. *BMC Bioinf.* **2009**, *10*, 168. doi:10.1186/1471-2105-10-168
115. Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849–857. doi:10.1021/jm00145a002
116. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. *J. Med. Chem.* **2006**, *49*, 6177–6196. doi:10.1021/jm051256o
117. Schrödinger, Release 2024-2; Schrödinger, LLC: New York, NY, USA, 2024.
118. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727–748. doi:10.1006/jmbi.1996.0897
119. *Molecular Operating Environment (MOE)*, 2022.02; Chemical Computing Group ULC: 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, Canada, 2023.
120. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *J. Comput. Chem.* **2009**, *30*, 2785–2791. doi:10.1002/jcc.21256
121. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639–1662. doi:10.1002/(sici)1096-987x(19981115)19:14<1639::aid-jcc10>3.0.co;2-b
122. Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455–461. doi:10.1002/jcc.21334
123. Gaudreault, F.; Najmanovich, R. *J. Chem. Inf. Model.* **2015**, *55*, 1323–1336. doi:10.1021/acs.jcim.5b00078
124. van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastiris, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. *J. Mol. Biol.* **2016**, *428*, 720–725. doi:10.1016/j.jmb.2015.09.014
125. Frank, M.; Kuhfeldt, E.; Cramer, J.; Watzl, C.; Prescher, H. *J. Med. Chem.* **2023**, *66*, 14315–14334. doi:10.1021/acs.jmedchem.3c01349
126. Bender, B. J.; Gahbauer, S.; Luttens, A.; Lyu, J.; Webb, C. M.; Stein, R. M.; Fink, E. A.; Balias, T. E.; Carlsson, J.; Irwin, J. J.; Shoichet, B. K. *Nat. Protoc.* **2021**, *16*, 4799–4832. doi:10.1038/s41596-021-00597-z
127. Zhang, S.; Chen, K. Y.; Zou, X. *Commun. Inf. Syst.* **2021**, *21*, 147–163. doi:10.4310/cis.2021.v21.n1.a7
128. Agu, P. C.; Afiukwa, C. A.; Orji, O. U.; Ezech, E. M.; Ofoke, I. H.; Ogbu, C. O.; Ugwuja, E. I.; Ajaja, P. M. *Sci. Rep.* **2023**, *13*, 13398. doi:10.1038/s41598-023-40160-2
129. Nivedha, A. K.; Thieker, D. F.; Makeneni, S.; Hu, H.; Woods, R. J. *J. Chem. Theory Comput.* **2016**, *12*, 892–901. doi:10.1021/acs.jctc.5b00834
130. Nivedha, A. K.; Makeneni, S.; Foley, B. L.; Tessier, M. B.; Woods, R. J. *J. Comput. Chem.* **2014**, *35*, 526–539. doi:10.1002/jcc.23517
131. Kerzmann, A.; Fuhrmann, J.; Kohlbacher, O.; Neumann, D. *J. Chem. Inf. Model.* **2008**, *48*, 1616–1625. doi:10.1021/ci800103u
132. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470–489. doi:10.1006/jmbi.1996.0477
133. Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. *J. Med. Chem.* **2006**, *49*, 5912–5931. doi:10.1021/jm050362n
134. Nance, M. L.; Labonte, J. W.; Adolf-Bryfogle, J.; Gray, J. J. *J. Phys. Chem. B* **2021**, *125*, 6807–6820. doi:10.1021/acs.jpcc.1c00910
135. Labonte, J. W.; Adolf-Bryfogle, J.; Schief, W. R.; Gray, J. J. *J. Comput. Chem.* **2017**, *38*, 276–287. doi:10.1002/jcc.24679
136. Boittier, E. D.; Burns, J. M.; Gandhi, N. S.; Ferro, V. *J. Chem. Inf. Model.* **2020**, *60*, 6328–6343. doi:10.1021/acs.jcim.0c00373
137. Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 601–619. doi:10.1007/s10822-006-9060-4
138. de Vries, S. J.; Schindler, C. E. M.; Chauvot de Beauchêne, I.; Zacharias, M. *Biophys. J.* **2015**, *108*, 462–465. doi:10.1016/j.bpj.2014.12.015
139. Uciechowska-Kaczmarzyk, U.; Chauvot de Beauchêne, I.; Samsonov, S. A. *J. Mol. Graphics Mod.* **2019**, *90*, 42–50. doi:10.1016/j.jmgm.2019.04.001
140. Ballester, P. J.; Mitchell, J. B. O. *Bioinformatics* **2010**, *26*, 1169–1175. doi:10.1093/bioinformatics/btq112
141. Pires, D. E. V.; Ascher, D. B. *Nucleic Acids Res.* **2016**, *44*, W557–W561. doi:10.1093/nar/gkw390
142. Li, H.; Sze, K.-H.; Lu, G.; Ballester, P. J. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11*, e1478. doi:10.1002/wcms.1478
143. Chalmers, G.; Glushka, J. N.; Foley, B. L.; Woods, R. J.; Prestegard, J. H. *J. Magn. Reson.* **2016**, *265*, 1–9. doi:10.1016/j.jmr.2016.01.006
144. Nguyen, T. B.; Pires, D. E. V.; Ascher, D. B. *Briefings Bioinf.* **2022**, *23*, bbab512. doi:10.1093/bib/bbab512

145. Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; Golze, D.; Wilhelm, J.; Chulkov, S.; Bani-Hashemian, M. H.; Weber, V.; Borštnik, U.; Taillefumier, M.; Jakobovits, A. S.; Lazzaro, A.; Pabst, H.; Müller, T.; Schade, R.; Guidon, M.; Andermatt, S.; Holmberg, N.; Schenter, G. K.; Hehn, A.; Bussy, A.; Belleflamme, F.; Tabacchi, G.; Glöß, A.; Lass, M.; Bethune, I.; Mundy, C. J.; Plessl, C.; Watkins, M.; VandeVondele, J.; Krack, M.; Hutter, J. J. *Chem. Phys.* **2020**, *152*. doi:10.1063/5.0007045
146. Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, 2006; pp 84–ee. doi:10.1145/1188455.1188544
147. Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. *Comput. Phys. Commun.* **2022**, *271*, 108171. doi:10.1016/j.cpc.2021.108171
148. Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. *Chem. Sci.* **2018**, *9*, 956–972. doi:10.1039/c7sc04531j
149. Bowers, K.; Chow, E.; Xu, H.; Dror, R.; Eastwood, M.; Gregersen, B.; Klepeis, J.; Kolossvary, I.; Moraes, M.; Sacerdoti, F.; Salmon, J.; Shan, Y.; Shaw, D. Scalable algorithms for molecular dynamics simulations on commodity clusters. In *SC '06: Proceedings of the 2006 ACM/IEEE conference on Supercomputing*, Association for Computing Machinery: New York, NY, USA, 2006. doi:10.1145/1188455.1188544
150. Kalé, L. V.; Bhatel, A.; Böhm, E. J.; Phillips, J. C. NAMD (Nanoscale Molecular Dynamics). In *Encyclopedia of Parallel Computing*; Padua, D., Ed.; Springer: Boston, MA, USA, 2011; pp 1249–1254. doi:10.1007/978-0-387-09766-4_505
151. Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. *Int. J. Mol. Sci.* **2016**, *17*, 144. doi:10.3390/ijms17020144
152. Roe, D. R.; Cheatham, T. E., III. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095. doi:10.1021/ct400341p
153. The PLUMED consortium. *Nat. Methods* **2019**, *16*, 670–673. doi:10.1038/s41592-019-0506-8
154. Kuprov, I.; Morris, L. C.; Glushka, J. N.; Prestegard, J. H. *J. Magn. Reson.* **2021**, *323*, 106891. doi:10.1016/j.jmr.2020.106891
155. Nieto-Fabregat, F.; Zhu, Q.; Vivès, C.; Zhang, Y.; Marseglia, A.; Chiodo, F.; Thépaut, M.; Rai, D.; Kulkarni, S. S.; Di Lorenzo, F.; Molinaro, A.; Marchetti, R.; Fieschi, F.; Xiao, G.; Yu, B.; Silipo, A. *JACS Au* **2024**, *4*, 697–712. doi:10.1021/jacsau.3c00748
156. Guo, Y.; Feinberg, H.; Conroy, E.; Mitchell, D. A.; Alvarez, R.; Blixt, O.; Taylor, M. E.; Weis, W. I.; Drickamer, K. *Nat. Struct. Mol. Biol.* **2004**, *11*, 591–598. doi:10.1038/nsmb784
157. *The PyMOL Molecular Graphics System*, Version~1.8; Schrödinger LLC: New York, NY, USA, 2015.
158. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612. doi:10.1002/jcc.20084
159. *Design, L.*; BIOVIA: California, 2014.
160. *Schrödinger Release 2022-3*, 2021; Schrödinger, LLC: New York, NY, USA.
161. *SAMSON: Software for Adaptive Modeling and Simulation of Nanosystems*; Samson: Grenoble, France, 2016.
162. Fogarty, C. A.; Fadda, E. *J. Phys. Chem. B* **2021**, *125*, 2607–2616. doi:10.1021/acs.jpcc.1c00304
163. Harbison, A. M.; Brosnan, L. P.; Fenlon, K.; Fadda, E. *Glycobiology* **2019**, *29*, 94–103. doi:10.1093/glycob/cwy097
164. Vanacore, A.; Vitiello, G.; Wanke, A.; Cavasso, D.; Clifton, L. A.; Mahdi, L.; Campanero-Rhodes, M. A.; Solís, D.; Wuhler, M.; Nicolardi, S.; Molinaro, A.; Marchetti, R.; Zuccaro, A.; Paduano, L.; Silipo, A. *Carbohydr. Polym.* **2022**, *277*, 118839. doi:10.1016/j.carbpol.2021.118839
165. Di Lorenzo, F.; Nicolardi, S.; Marchetti, R.; Vanacore, A.; Gallucci, N.; Duda, K.; Nieto Fabregat, F.; Nguyen, H. N. A.; Gully, D.; Saenz, J.; Giraud, E.; Paduano, L.; Molinaro, A.; D'Errico, G.; Silipo, A. *JACS Au* **2023**, *3*, 929–942. doi:10.1021/jacsau.3c00025
166. Makshakova, O.; Zykwiniska, A.; Cuenot, S.; Collic-Jouault, S.; Perez, S. *Carbohydr. Polym.* **2022**, *276*, 118732. doi:10.1016/j.carbpol.2021.118732
167. Lei, M.; Huang, W.; Jin, Z.; Sun, J.; Zhang, M.; Zhao, S. *Carbohydr. Polym.* **2022**, *297*, 119993. doi:10.1016/j.carbpol.2022.119993
168. Harbison, A. M.; Fogarty, C. A.; Phung, T. K.; Satheesan, A.; Schulz, B. L.; Fadda, E. *Chem. Sci.* **2022**, *13*, 386–395. doi:10.1039/d1sc04832e
169. Mou, Y.; Huang, P.-S.; Thomas, L. M.; Mayo, S. L. *J. Mol. Biol.* **2015**, *427*, 2697–2706. doi:10.1016/j.jmb.2015.06.006
170. Cao, X.; Yang, X.; Xiao, M.; Jiang, X. *Int. J. Mol. Sci.* **2023**, *24*, 6830. doi:10.3390/ijms24076830
171. Di Carluccio, C.; Forgione, R. E.; Bosso, A.; Yokoyama, S.; Manabe, Y.; Pizzo, E.; Molinaro, A.; Fukase, K.; Fragai, M.; Bensing, B. A.; Marchetti, R.; Silipo, A. *RSC Chem. Biol.* **2021**, *2*, 1618–1630. doi:10.1039/d1cb00173f
172. Antonini, G.; Civera, M.; Lal, K.; Mazzotta, S.; Varrot, A.; Bernardi, A.; Belvisi, L. *Front. Mol. Biosci.* **2023**, *10*, 1201630. doi:10.3389/fmolb.2023.1201630
173. Matamoros-Recio, A.; Franco-Gonzalez, J. F.; Perez-Regidor, L.; Billod, J.-M.; Guzman-Caldentey, J.; Martin-Santamaria, S. *Chem. – Eur. J.* **2021**, *27*, 15406–15425. doi:10.1002/chem.202102995
174. Pirone, L.; Nieto-Fabregat, F.; Di Gaetano, S.; Capasso, D.; Russo, R.; Traboni, S.; Molinaro, A.; Iadonisi, A.; Saviano, M.; Marchetti, R.; Silipo, A.; Pedone, E. *Int. J. Mol. Sci.* **2022**, *23*, 8273. doi:10.3390/ijms23158273
175. Zhang, S.; Trinh, C.; Schweitzer-Stenner, R.; Urbanc, B. *Biophys. J.* **2019**, *116*, 61a. doi:10.1016/j.bpj.2018.11.374
176. Köhling, S.; Blaszkiewicz, J.; Ruiz-Gómez, G.; Fernández-Bachiller, M. I.; Lemmnitzer, K.; Panitz, N.; Beck-Sickinger, A. G.; Schiller, J.; Pisabarro, M. T.; Rademann, J. *Chem. Sci.* **2019**, *10*, 866–878. doi:10.1039/c8sc03649g
177. Sankaranarayanan, N. V.; Nagarajan, B.; Desai, U. R. *Curr. Opin. Struct. Biol.* **2018**, *50*, 91–100. doi:10.1016/j.sbi.2017.12.004
178. Joseph, P. R. B.; Mosier, P. D.; Desai, U. R.; Rajarathnam, K. *Biochem. J.* **2015**, *472*, 121–133. doi:10.1042/bj20150059
179. Perez, S.; Makshakova, O.; Angulo, J.; Bedini, E.; Bisio, A.; de Paz, J. L.; Fadda, E.; Guerrini, M.; Hricovini, M.; Hricovini, M.; Lisacek, F.; Nieto, P. M.; Pagel, K.; Paiardi, G.; Richter, R.; Samsonov, S. A.; Vivès, R. R.; Nikitovic, D.; Ricard Blum, S. *JACS Au* **2023**, *3*, 628–656. doi:10.1021/jacsau.2c00569

License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (<https://www.beilstein-journals.org/bjoc/terms>), which is identical to the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0>). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
<https://doi.org/10.3762/bjoc.20.180>