# Computational modeling of protein-carbohydrate interactions: Current trends and future challenges

Serge Pèrez*
Centre de Recherches sur les Macromolècules Vègètales, CNRS, Universitè Grenoble Alpes, Grenoble, France
*e-mail address: serge.perez@cermav.cnrs.fr

## Contents



## Abbreviations

AA all-atom
Aβ amyloid
CD44 a cell-surface glycoprotein
CG coarse-grained
DAG diacylglycerol
DFT density functional theory
MGD1monogalactosyldiacylglycerol synthase
DGDG digalactosyldiacylglycerol
ECM extracellular matrix
GAG glycosaminoglycan
HPC high-performance computing
IDP intrinsically disordered protein
iEM inner envelope membrane
MD molecular dynamics
MGDG monogalactosyldiacylglycerol
MM molecular mechanics
PG phosphatidylglycerol
QM quantum mechanics
UDP uridine diphosphate
VR virtual reality

## 1. Introduction

This chapter is a perspective piece that aims to continue the trajectory of an ongoing body of work that explores simulation and theoretical approaches to characterizing and understanding glycans, or carbohydrates, and their interactions with proteins.[1-17] These contributions and related articles form the basis of the current state of computational aspects of protein–carbohydrate interactions, as summarized in the following paragraphs.

Computational techniques now cover a more comprehensive range of carbohydrate systems and assemblies in terms of types and sizes. They have contributed to the elucidation of the dynamics, interactions, and structures of complex carbohydrates. Understanding how the specificity and the strength of protein–carbohydrate interactions differ, depending on the kind of proteins

involved, is essential. For several reasons, interactions between proteins and carbohydrates are crucial.

They are obligatory in maintaining life, holistic tissue, and homeostasis. They play a role in inflammation, cell proliferation, differentiation, aggregation, signal transduction, host–pathogen recognition, and protein structure stabilization. Additionally, they have numerous uses in the design of pharmaceuticals, including creating antibodies, vaccinations, and inhibitors. They also have broad applications in drug design, such as developing antibodies, vaccines, and inhibitors.

Some examples of protein-carbohydrate interactions are as follows: Carbohydrate-active enzymes that catalyze biochemical reactions involving glycosylation, as well as the synthesis and hydrolysis of carbohydrates, lectins, antibodies, sugar transporters, glycosaminoglycans, and lipopolysaccharides.

The characterizations of protein–carbohydrate interactions are challenging from both a theoretical and an experimental point of view. Several theoretical models and their underlying approximations delineate the scope of applying computational methods to elucidate the structural and dynamical features. These range from ab initio to coarse-grained (CG) methods and from deterministic to heuristic approaches. They are as follows: (1) DFT-based ab initio simulations, (2) quantum mechanics/ molecular mechanics (QM/MM) and QM/QM hybrid methods, (3) semi-empirical methods, (4) molecular mechanics (MM) and molecular dynamics (MD) simulations, (5) heuristic methods (Monte Carlo and genetic algorithms), (6) coarse-grained methods, and (7) docking calculations. Molecular dynamics has become the method of choice in its all-atom (AA) and coarse-grained (CG) representations.

Theoretical and technological advances often accompany diffraction methods, high-resolution spectroscopy, and other spectroscopic methods.[18] They provide a way to reconcile the experimentally available data and to predict structural and dynamical features that may not yet be accessible (Fig. 1).

Carbohydrates are influenced by stereo-electronic effects, which can be understood using quantum chemical methods. These effects are also incorporated into force fields that allow many users to perform computational studies of their systems in conjunction with experimental work. Molecular simulation methods have become powerful and sophisticated enough to provide structural information that can explain or support experiments and lead to discoveries. High-performance computing has also enhanced the role of molecular simulation methods in guiding experimental design and exploring new phenomena. Molecular simulation methods can access previously unreachable scales of space and time. In particular, atomistic MD simulations can capture the "true" 3D structure and dynamics of molecules as they occur in real-time. This can help to establish structure−function and structure−property relationships in some cases, as well as to characterize physicochemical and mechanical properties.

Coarse-grained simulations can be applied to various glycoscience systems, such as polysaccharides (from different sources, including animals, plants, bacteria, and marine organisms), N-linked and O-linked glycans, and glycolipids. Coarse-grained simulations can cover different length scales and model complex carbohydrate-based materials. Coarse-grained simulations can also overcome the limitations of experiments and all-atom simulations for highly complex systems. They are beneficial for studying the dynamic formation of glycolipid nanostructures, where carbohydrate−carbohydrate interactions are essential.

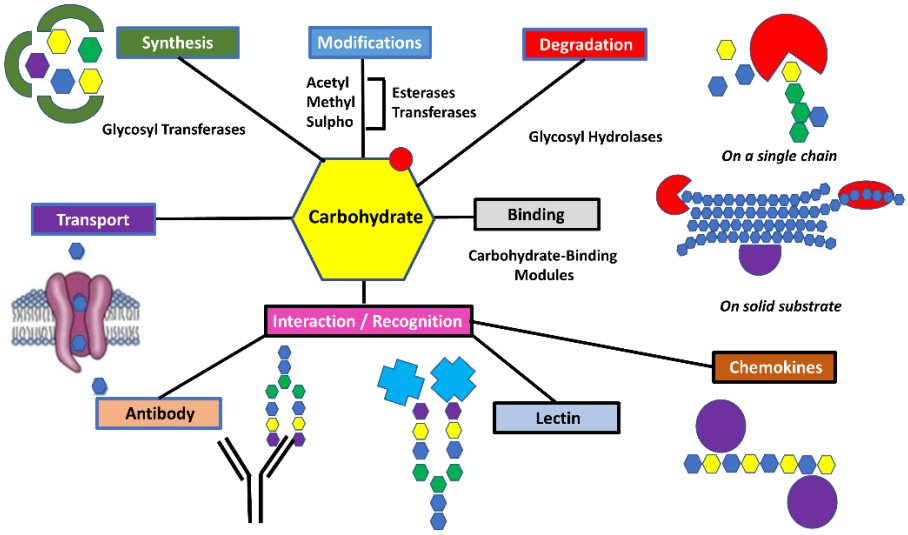


Fig. 1 Synopsis of the families of proteins interacting with carbohydrates along with their functions: transport, synthesis (glycosyl transferases), modification (auxiliaries, enzymes), degradations (glycosyl hydrolases, on a single glycan and semi-crystalline and crystalline glycans), carbohydrate-binding modules, antibodies, lectins, and chemokines. Adapted *From Pérez, S.; Fadda, E.; Makshakova, O. Computational Modeling in Glycoscience. In* Comprehensive Glycoscience*, 2nd ed.; Barchi, J. J., Ed.; Elsevier: Amsterdam, 2021; pp 374—404.*

A key challenge for computational methods is to capture the multivalent effect that governs protein–carbohydrate assembly. Many carbohydrate-binding proteins, such as adhesins and lectins, have low affinity and narrow carbohydrate recognition domains, but they achieve specificity by binding multiple identical glycoside units in different arrangements. This requires understanding the physicochemical principles that underlie such associations, such as the formation of glycolipid and glyco-surface patches that create a "glyco landscape" or "glycotope." Computational tools should be able to model the glyco-surface resulting from the spatial distribution of glycolipids and their interaction with the glyco-canopy (analogous to the top layer of a forest formed by tree crowns).

Computational methods can also help to study the enzymatic degradation of polysaccharides in the solid state, which is a challenging problem. Computer simulations have already contributed to revealing the chemical mechanisms of glycosyl hydrolases. They have helped to identify catalytic residues, complex conformational changes, and mechanistic details that are not accessible by experiments. This research area benefits from the availability of many crystal structures of proteins and their carbohydrate complexes.

Despite the many advances reported, there are several interrelated challenges to the further development of the field. For this article, the presentation of challenges is divided into those that depend on general "methodological" developments and those that address specific "glyco-oriented" issues.

## 2. The Impetuous form methodological advance

### 2.1 Lessons learned from SARS-Cov2

The imminent threat to global human health and socioeconomic stability posed by the novel coronavirus SARS-CoV-2 spurred an unprecedented effort by many communities, including a broad coalition ranging from virologists to computational scientists, to meet the challenge of rapidly building models of the

viral glycoproteins and complex they form. Over a million citizen-scientists collaborated through the Folding@home distributed computing project to create the first exascale computer and simulate 0.1 s of the viral proteome. In its ability to capture the entire ensemble of structures adopted by a glycoprotein, molecular dynamics simulation predicted the dramatic opening of the apo spike complex, far beyond that seen experimentally, explaining and predicting the existence of "cryptic" epitopes and helping to characterize crucial stages of infection.[19] From a computational glycoscience perspective, the results highlight the incredible utility of the community-built computer to enable rapid understanding of health and disease, providing a rich source of structural data to accelerate the design of therapeutics. It established the soundness of the principles underlying molecular dynamics simulations. It brought to the attention of the largest community the essential role of the extensive N-glycosylation coat on viral fusion proteins.[20] It exemplifies the endeavor which can and will be addressed through high-performance computing (HPC) based molecular simulations.

### 2.2. A repository for MD-generated glycan and protein−glycan structures

Despite its massive use of computational resources, the field of molecular modeling does not have a place where all deposited and documented results can be stored and made publicly available. It is, therefore, tempting to propose creating and organizing such a repository of 3D data. Ideally, the data would correspond to the most populated conformers identified by simulation analysis, with detailed information on relative populations and energetics. Along these lines, an ongoing development called GlycoShape3D aims to provide complete and consistent data on free (unbound) glycans from equilibrium MD simulations in a format accessible to expert and non-expert glycobiologists.[21] In addition to providing the ability to reproduce published results, when unlocked, such a database would contribute to the wealth of publicly available resources in glycoscience.

However, such a recent initiative does not cover the computationally intensive simulations performed to unravel specific molecular phenomena, essentially used for a single publication. When these MD simulation files are made available, due to the rise of open science, they accumulate in generalist data repositories that are neither indexed, curated, nor easily searchable. No central repository hosts all MD simulation files to date, creating a so-called "dark matter of MD."[22] Initiatives are being taken by scientists who have inferred metadata to propose a prototype search engine to explore the collected MD data. To pursue this direction, they call on the community to continue sharing MD data and to increase the filling and standardization of metadata to reuse this valuable matter. Their integration into meta-databases would complement the experimental and computational data set that, with machine learning-based applications, will enable the rapid advancement of glycoscience and its contribution to understanding the many processes and architectures of these complex biomolecules.

Computational glycoscientists would benefit from joining such an initiative by creating and organizing a 3D data repository, where the stored data would correspond to the most populated conformers derived from the simulation with annotated details on energetics and relative populations and transition states,… The creation of such a structural database would allow the reproducibility of data, a feature that is currently lacking. It would provide users with the unique ability to monitor the actual volume of space-specific glycosylation patterns occupied on a membrane or a protein surface, allowing or preventing recognition from other receptors.

## 2.3 Deep learning methods and data management

Due to the evolution of hardware, algorithms, and software, computational methods are receiving increasing attention. In particular, innovative approaches based on deep learning algorithms offer new ways to explore protein-carbohydrate interactions[23] by analyzing the structural and functional features of proteins and carbohydrates and learning the patterns and rules that govern their binding, therefore extending the capacity of computational tools such as MD simulation, hybrid (QM/MM) methods, or molecular docking. Recent impressive progress in using deep learning methods such as AlfaFold,[24] RoseTTAFold,[25] RaptorX[26] and others to predict 3D protein structure illustrates the value of collecting well-characterized data over a long period. Ultimately, standardized, structured, and well-annotated data can accumulate and provide opportunities to train models and improve the prediction of complex carbohydrates in specific contexts or environments. This is not quite the case in computational glycan research, often hampered by glycan-related data complexity, sparsity, and diversity. Published studies report only a limited fraction of the data leading to the results.

However, in a few cases, experimental data derived from X-ray resolution of crystalline protein–carbohydrate complexes are organized in curated and annotated databases.[27] Their construction follows the generation of findable, accessible, interoperable and reusable (FAIR) biological data. This organization is essential to feed and train machine learning-based applications to predict different structural organization levels and characterize the unique features of recognition and binding of carbohydrate structures by specific proteins. Typically, such databases cross-reference other databases that rely on different strategies to visualize the interaction between carbohydrate ligands and their protein environments.

Beyond simply predicting protein–carbohydrate interactions, structural glycobioinformatics opens up a much broader field of exploration and prediction. Thousands of 3D structures of lectin–glycan complexes are available from X-ray crystallography and NMR studies. They are stored in a searchable database,[28,29] providing binding information to protein sequences and structures. The resulting mapping to a sequence-based lectin prediction application uniquely explores candidate lectins identified in available proteomes from all kingdoms and for all available lectin classes. Using machine learning algorithms to predict protein–carbohydrate interactions is under active development.[30]

## 2.4 Immersive molecular visualization and data analysis

Structural biology and bioinformatics advances significantly increase data throughput, generation and complexity. In parallel to these developments, several innovative applications are being developed for immersive environments that promote direct interactions on semantically linked 2D and 3D heterogeneous data displayed in a shared workspace.[31] Applying such generic tools to the fields of structural biology is likely to be relevant to glycobiology, which is firmly rooted in multidisciplinary approaches, often implemented orthogonally to compensate for the lack of genetic determinants for glycosylation. These approaches all produce rich catalogs of data, each of which is not sufficient to determine glycan types of glycosylation sites/populations but is necessary when put in context with additional complementary sets.

Interactive molecular simulations create, manipulate and visualize molecules in 3D and can be merged with virtual reality to create immersive and

interactive environments for exploring and manipulating molecular systems in 3D.[31] The "interactive molecular simulations" offer the possibility to visualize a running simulation in interactive time, i.e., a timeframe that is compatible with human perception and the possibility to manipulate and make changes during the simulation (Fig. 2).

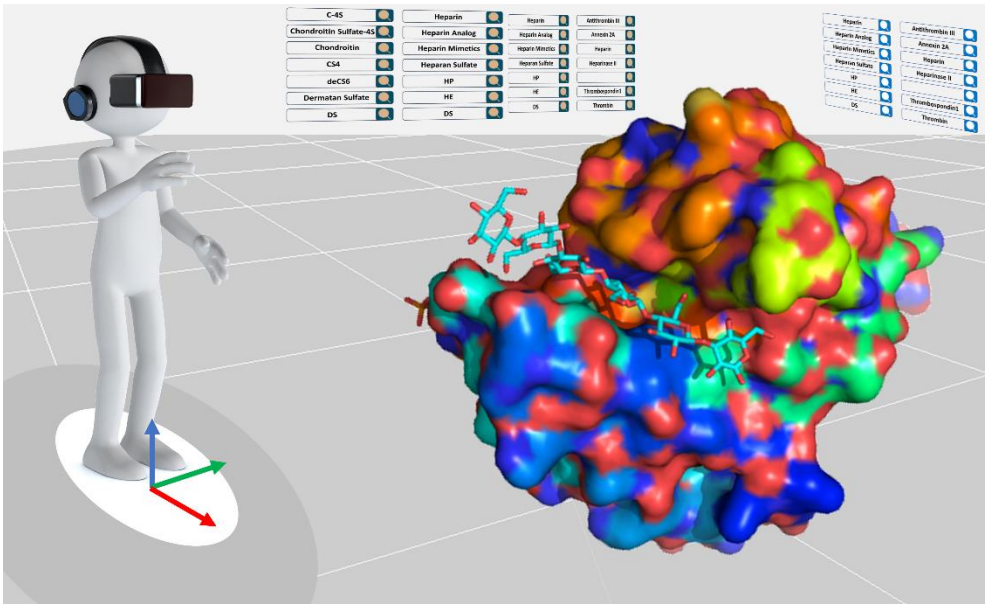


Fig. 2 Immersive visualization of the molecular shape of a complex carbohydrate actively docked in a protein extracted from a specialized database of 3D structures.

A mixed-reality device superimposes holograms in the real world. Users can interact with the molecules using hand gestures, voice commands, or gaze tracking and explore various molecular properties such as electrostatic potential, hydrogen bonding, or molecular dynamics.[32] Another method uses narrative techniques such as annotations, transitions, and interactivity to guide users through exploring scientific data in an immersive environment. Users can view and manipulate volumetric data using a VRheadset and a handheld device, accessing a rich crosslinked database of biological information. The tool also supports multi-scale and multi-modal exploration of the data. Semantics for an integrative and immersive pipeline combining visualization and analysis of molecular data was proposed to create an intelligent system.[34]

## 3. GLYCO-LANDSCAPE COMPLEXITY CHALLENGE

### 3.1 The glycocalyx: An overview

The dense layer of glycoconjugates (glycolipids, glycoproteins and proteo-glycans) attached to the surface of various cell types, termed the glycocalyx, is an information-rich barrier that mediates many molecular interactions in cell–cell communication, recognition, adhesion, signal transduction, and host–pathogen interactions. As such, the glyocalyx influences the physicochemical properties of cell membranes. The study of glycocalyx components and their dynamics provides a playground where deciphering protein–carbohydrate interactions reaches a higher level of complexity through the interplay between membrane proteins,

their lipid environment, and the glycoconjugate. It provides some reference background for macromolecular interactions in highly crowded media.

### 3.2 Glycosyl transferases at work

Subject to their association with the lipid bilayer, membrane-bound proteins can be either "peripheral" or "integral." Peripheral membrane proteins transiently bind to one side of the membrane (monotopic interaction) or are bound to other proteins by weak, noncovalent interactions at the membrane interface. Integral membrane proteins tightly bind to the membrane, where they usually perform structural functions such as that of transporters, linkers, channels, and cell-adhesion proteins. Depending on how integral membrane proteins are embedded in the membrane, they are called monotopic, bitopic, or polytopic proteins. Monotopic integral membrane proteins are permanently associated with only one side of the membrane. In contrast, bitopic and polytopic integral membrane proteins have one or more transmembrane segments that cross the membrane at different levels. A limited number of computational explorations of such complex cases open the road to further investigations and shed light on some of the stunning features discovered.[35,36]

Chloroplasts offer one of the first examples as molecular machines that convert the harvested photons into chemical energy. This conversion occurs throughout a unique spatial architecture resulting from the occurrence of and spatial organization of two monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG) galactoglycerolipids, which are the main lipids. Most of the MGDG is synthesized in the inner envelope membrane (iEM) of the chloroplast by a glycosyltransferase, monogalactosyldiacylglycerol synthase, MGD1, a member of the GT-B family that possesses two distinct domains and an acceptor substrate that are typically bound in the cleft between these domains.[37]

Catalysis requires two substrates, hydrophilic sugar-bearing UDP-galactose and hydrophobic fatty acid tails bearing diacylglycerol (DAG); at least one activator, anionic lipid molecule, phosphatidylglycerol (PG), to be bound by MGD1. Molecular dynamics simulations [coarse-grained (CG) and all-atom (AA)] revealed the compelling interactions between MGD1 and the lipid bilayers and the lipid capture by self-assembly. Without protein, rafts of PG and DAG molecules spontaneously form. When MGD1 is embedded in the membrane, the protein interacts with the PG/DAG rafts and accumulates DAG by lateral and transverse diffusion. A concomitant change in membrane curvature induces intrinsic dynamics of the protein, which is essential for catalytic activity.[38]

A similar case was reported for the glycosyltransferase synthesizing glycoglycerolipids in *Mycoplasma genitalium* membranes.[39] The authors invoke the role played by an amphiphilic peptide undergoing helix formation and its subsequent influence on catalysis.

### 3.3 The extracellular matrix

Glycosaminoglycans (GAGs) are complex carbohydrates ubiquitously and abundantly expressed on the cell surface and in the extracellular matrix (ECM). Their extraordinary structural diversity allows them to interact with various biological molecules. Through macromolecular interactions, GAGs modulate various biological processes, particularly as signaling molecules, and regulate the presentation of protein ligands to their respective receptors. GAGs, however, do not self-associate on their own under physiological conditions, and the self-organization of GAGs is intimately related to the proteins they bind and the core proteins to which

they are attached.[40]

Despite their functional importance, little is known about the structure and dynamics of the crosslinking nodes in all the many GAG-containing matrices. For example, how is the degree of crosslinking regulated to achieve matrices with the desired morphology and biophysical properties? Intercellular signaling molecules (e.g., morphogens/growth factors for tissue development/repair and chemokines for immune cell trafficking) rely on GAGs for their precise distribution throughout the extracellular space. It has long been thought that the primary function of GAGs is to control the presentation of intercellular signaling proteins to their cognate cell-surface receptors (and the downstream intracellular signaling process). However, recent evidence suggests that many signaling proteins can also crosslink GAGs.[41,42] Thus, these proteins may also exert their functions independently of cognate receptors by dynamically reorganizing the GAG-rich extracellular matrix and modulating matrix morphology and biophysical properties.[43] This promising new research field requires new tools and mindsets.[44]

Understanding the molecular and physical mechanisms underlying the functions of GAGs under the influence of proteins is still at an early stage of development. More convergent approaches are needed, ranging from biophysics to high computational simulations, including appropriate analytical tools and methods to reconstitute multi-partner GAG-protein interactions and protein-mediated GAG self-assembly in vitro. Such molecularly defined environments will enable new studies of structure–property functional relationships that are impossible with the more complex, less defined, and less tunable matrices produced by cells and tissues.[45]

### 3.4 Intrinsically disordered protein-carbohydrate interactions

For glycoscience, it is essential to consider the macromolecular interactions that occur in highly crowded media, as partially illustrated in the previous sections, where various macromolecules evolve to function with minimal available space and limited free water. Among them, a class of intrinsically disordered proteins (IDPs) lack regular secondary or well-defined tertiary structures; they can be highly flexible under native physiological conditions. While largely or partially unstructured, they are susceptible to change and adopt a conformation upon binding. to an interacting partner that undergoes a coupled-folding-and-binding mechanism.[46-48] IDPs can interact with different partners and modulate their functions through various mechanisms such as electrostatic interactions, hydrogen bonding, hydrophobic interactions, and conformational changes. Some examples of interactions between carbohydrates and IDPs involve glycogenin amyloid beta and mucins.

Several systematic bioinformatics studies have shown that IDPs constitute a significant fraction of every known proteome, with the number of IDPs per proteome increasing with the organism's complexity.

Glycosylation patterns can influence the conformational preferences of the IDP interactions, as reflected by their stability and activity. Examples include the following:

- Glycosylation patterns can modulate the conformational ensemble of IDPs by introducing steric hindrance, electrostatic repulsion, or hydrogen bonding between the sugar chains and the protein backbone or side chains. For example, glycosylation of the N-terminal region of amyloid beta (Aβ), a peptide implicated in Alzheimer's disease, reduces its propensity to form β-sheets and aggregates by increasing its conformational heterogeneity and disorder.

- Glycosylation patterns can affect the interactions of IDPs with other molecules, such as proteins, lipids, carbohydrates, or metals, by altering their binding affinity, specificity, or selectivity. For example, glycosylation of mucins, glycoproteins that form a protective layer on the surface of various tissues, affects their interactions with pathogens and immune cells by modulating their recognition and adhesion properties.
- Glycosylation patterns can influence the stability of IDPs by protecting them from degradation, denaturation, or aggregation. For example, the glycosylation of glycogenin, a self-glycosylating enzyme that initiates glycogen synthesis, enhances its stability by preventing its proteolysis and oxidation.
- Glycosylation patterns can affect the solubility of IDPs by increasing their hydrophilicity, hydration, or viscosity. For example, glycosylation of prothrombin, a blood coagulation factor that undergoes a disorder-to-order transition upon activation, increases its solubility by reducing its aggregation tendency and facilitating its folding.
- Glycosylation patterns can regulate the activity of IDPs by modulating their catalytic efficiency, substrate specificity, or allosteric regulation.

In the context of crucial interactions involving some GAGs and proteoglycans, the case of syndecan[49] and the case of CD44 provide an example of a protein possessing both ordered domains and functionally important intrinsically disordered protein regions. The ordered region consists of the extra-cellular domain, which binds to hyaluronic acid (hyaluronan) and other ligands, and the transmembrane domain, which anchors the protein to the cell membrane. The disordered region of the cytoplasmic tail is considered an IDP because it does not have a well-defined structure and can adopt different conformations depending on its environment and interactions. It plays a significant role in the function of the receptor.

The cytoplasmic tail of CD44 is involved in regulating the activity, stability, solubility, and interactions of the receptor with other molecules, such as the ezrin-radixin-moesin protein family, which connects the actin cytoskeleton with the membrane receptor. Palmitoylation can modulate the affinity of CD44 for lipid rafts and affects the activity, stability and trafficking of CD-44.[50] Part of the cytoplasmic tail undergoes various regulations and modifications, interacting with various signaling molecules and experiencing post-translational modifications. Several ligands of CD44 can affect its structure and function in various physiological and pathological processes. Examples include (1) osteopontin, a glycoprotein involved in inflammation, wound healing, and bone remodeling; (2) collagens; (3) matrix metalloproteinases; and (4) hyaluronic acid (hyaluronan), which interacts through the N-terminal domain and modulates various cellular functions, such as adhesion, migration, proliferation and survival. The delineation of those interactions and their structural features that occur either in the ordered region of CD44 or in the disordered region of its cytoplasmic tail represents a complementary line of research in the study of macromolecular crowding in structural glycoscience. Recent experimental [51] and computational approaches are already setting up the basis for deeper investigations.

## Conclusions

In this short perspective article, I have shared my current opinion and vision of the prevailing status of some issues concerning protein–carbohydrate interactions and beyond and how the HPC technology revolution and machine learning algorithms, which lead to integrative immersion, would provide a step change in the global fields of chemistry and biology. The wave will impact the glycoscience community in general, addressing the role of carbohydrates in health disease and environmental issues to a complexity never before attained or imagined.

Among the few examples described, properly considering intrinsically disordered proteins will require the development of new concepts and tools to deal with their spatiotemporal heterogeneity and high conformational flexibility.

**Acknowledgements**

The author thanks Dragana Nikitovic for suggestions and careful reading and editing of this manuscript. The following funding supported this work: The Cross-Disciplinary Program Glyco@Alps within the framework "Investissement d'Avenir" program [ANR-15IDEX-02].

**Competing interest**. The author declares that he has no known competing financial interests or personal relationships that influence the work reported in this chapter.

## References

1. Fadda, E.; Woods, R. J. Molecular Simulations of Carbohydrates and Protein– Carbohydrate Interactions: Motivation, Issues and Prospects. Drug Discovery Today 2010, 15, 596–609.
2. P'erez, S.; Tvaroˇska, I. Carbohydrate–Protein Interactions: Molecular Modeling Insights.
3. Adv. Carbohydr. Chem. Biochem. 2014, 71, 9–136.
4. Grant, O. C.; Woods, R. J. Recent Advances in Employing Molecular Modelling to Determine the Specificity of Glycan-Binding Proteins. Curr. Opin. Struct. Biol. 2014, 28, 47–55.
5. Frank, M. Conformational Analysis of Oligosaccharides and Polysaccharides Using Molecular Dynamics Simulations. Methods Mol. Biol. 2015, 1273, 359–377.
6. Arde`vol, A.; Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights From Ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. J. Am. Chem. Soc. 2015, 137, 7528–7547.
7. Arde`vol, A.; Iglesias-Ferna´ndez, J.; Rojas-Cervellera, V.; Rovira, C. The Reaction Mechanism of Retaining Glycosyltransferases. Biochem. Soc. Trans. 2016, 44, 51–60.
8. Almond, A. Multiscale Modeling of Glycosaminoglycan Structure and Dynamics: Current Methods and Challenges. Curr. Opin. Struct. Biol. 2018, 50, 58–64.
9. Coines, J.; Raich, L.; Rovira, C. Modeling Catalytic Reaction Mechanisms in Glycoside Hydrolases. Curr. Opin. Chem. Biol. 2019, 53, 183–191.
10. Scherbinina, S. I.; Toukach, P. V. Three-Dimensional Structures of Carbohydrates and Where to Find Them. Int. J. Mol. Sci. 2020, 21, 7702.
11. P'erez, S.; Fadda, E.; Makshakova, O. Computational Modeling in Glycoscience. In Comprehensive Glycoscience; Barchi, J. J., Ed, 2nd ed.; Elsevier: Amsterdam, 2021; pp. 374–404.
12. Fadda, E. Molecular Simulations of Complex Carbohydrates and Glycoconjugates. Curr. Opin. Chem. Biol. 2022, 69, 102175. https://doi.org/10.1016/j.cbpa.2022.102175.
13. Cao, Y.; Park, S. J.; Im, W. A Systematic Analysis of Protein–Carbohydrate Interactions in the Protein Data Bank. Glycobiology 2021, 31, 126–136.
14. P'erez, S.; Makshakova, O. Multifaceted Computational Modeling in Glycoscience.
15. Chem. Rev. 2022, 122, 15914–15970.
16. Nagarajan, B.; Holmes, S. G.; Sankaranarayanan, N. V.; Desai, U. R. Molecular Dynamics Simulations to Understand Glycosaminoglycan Interactions in the Free- and Protein-Bound States. *Curr. Opin. Struct. Biol.* 2022, *74*, 102356.
17. Marcisz, M.; Zacharias, M.; Samsonov, S. A. Modeling Protein–Glycosaminoglycan
18. Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zˇˇıdek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, 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. Highly Accurate Protein Structure Prediction With AlphaFold. Nature 2021, 596, 583–589.
19. 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 Complexes: Does the Size Matter? J. Chem. Inf. Model. 2021, 61, 4475–4485.
20. Ingolfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Q8 Marrink, S. J. The Power of Coarse Graining in Biomolecular Simulations. WIREs Comput. Mol. Sci. 2014, 4, 225–248.
21. Morais, M. A. B.; Nin-Hill, A.; Rovira, C. Glycosidase Mechanisms: Sugar Conformations and Reactivity in Endo- and Exo-Acting Enzymes. Curr. Opin. Chem. Biol. 2023, 74, 102282.
22. Marchetti, R.; P'erez, S.; Arda, A.; Imberty, A.; Jimenez-Barbero, J.; Silipo, A.; Molinaro, A. "Rules of Engagement" of Protein–Glycoconjugate Interactions: A Molecular View Achievable by Using NMR Spectroscopy and Molecular Modeling. ChemistryOpen 2016, 5, 274–296.
23. Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in SARS-CoV-2 Spike Protein. ACS Cent. Sci. 2020, 28, 1722–1734.
24. Zimmerman, M. I.; Porter, J. R.; Ward, M. D.; Singh, S.; Vithani, N.; Meller, A.; Mallimadugula, U. L.; Kuhn, C. E.; Borowsky, J. H.; Wiewiora, R. P.; Hurley, M. F. D.; Harbison, A. M.; Fogarty, C. A.; Coffland, J. E.; Fadda, E.; Voelz, V. A.; Chodera, J. D.; Bowman, G. R. SARS-CoV-2 Simulations Go Exascale to Predict Dramatic Spike Opening and Cryptic Pockets Across the Proteome. Nat. Chem. 2021, 13, 651–659.
25. Ives, C.; Singh, O.; Fogarty, C.; Harbison, A.; Aoife, H.; Satheesan, A.; D'Andrea, S.; Tropea, B.; Fadda, E. GlycoShape3D: A Database and Toolbox for Structural Glycomics. In 21st European Carbohydrate Symposium; 2023. Paris 9-1, July 2023.
26. Tiemann, J. K. S.; Szczuka, M.; Bouarroudj, L.; Oussaren, M.; Garcia, S.; Howard, R. J.; Q9 Delemotte, L.; Lindahl, E.; Baaden, M.; Lindorff-Larsen, K.; Chavent, M.; Poulain, P. MDverse: Shedding Light on the Dark Matter of Molecular Dynamics Simulations. bioRxiv 2023, 2023.05.02.538537 (Preprint) https://doi.org/10. 1101/2023.05.02.538537.
27. Bojar, D.; Lisacek, F. Glycoinformatics in the Artificial Intelligence Era. Chem. Rev.
28. 2022, 26, 15971–15988.
29. Jumper, J.;, 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. M.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. Science 2021, 373, 871–876.
30. Xu, J. Distance-Based Protein Folding Powered by Deep Learning. Proc. Natl. Acad. Sci. U. S. A. 2019, 116, 16856–16865.
31. Pèrez, S.; Sarkar, A.; Rivet, A.; Breton, C.; Imberty, A. Glyco3D: A Portal for Structural Glycosciences. *Methods Mol. Biol.* 2015, *1273*, 241−258.
32. Bonnardel, F.; Mariethoz, J.; Salentin, S.; Robin, X.; Schroeder, M.; Pèrez, S.; Lisacek, F.; Imberty, A. UniLectin3D, a Database of Carbohydrate Binding Proteins With Curated Information on 3D Structures and Interacting Ligands. *Nucleic Acids Res.* 2019, *47*, D1236−D1244.
33. Bonnardel, F.; Mariethoz, J.; Pèrez, S.; Imberty, A.; Lisacek, F. LectomeXplore, an Update of UniLectin for the Discovery of Carbohydrate-Binding Proteins Based on a New Lectin Classification. *Nucleic Acids Res.* 2021, *49*, D1548−D1554.
34. Canner, S. W.; Shanker, S.; Gray, J. J. Structure-Based Neural Network Protein−Carbohydrate Interaction Predictions at the Residue Level. *Front. Bioinform.* 2023, *3*, 1186531.
35. Lanrezac, A.; Fèrey, N.; Baaden, M. Wielding the Power of Interactive Molecular Simulations. *WIREs Comput. Mol. Sci.* 2021, *12*, e1594.
36. Liu, R.; Wang, H.; Zhang, C.; Chen, X.; Wang, L.; Ji, G.; Zhao, B.; Mao, Z.; Yang, D. Narrative Scientific Data Visualization in an Immersive Environment. *Bioinformatics* 2021, btab052.
37. Sommer, B.; Baaden, M.; Krone, M.; Woods, A. From Virtual Reality to Immersive Analytics in Bioinformatics. *J. Integr. Bioinform.* 2018, *15*, 20180043.
38. Trellet, M.; Fèrey, N.; Flotyński, J.; Baaden, M.; Bourdot, P. Semantics for an Integrative and Immersive Pipeline Combining Visualization and Analysis of Molecular Data. *J. Integr. Bioinform.* 2018, *15*, 20180004.
39. Albesa-Jovè, D.; Guerin, M. E. The Conformational Plasticity of Glycosyltransferases. *Curr. Opin. Struct. Biol.* 2016, *40*, 23−32.
40. Albesa-Jovè, D.; Giganti, D.; Jackson, M.; Alzari, P. M.; Guerin, M. E. Structure-Function Relationships of Membrane-Associated GT-B Glycosyltransferases. *Glycobiology* 2014, *24*, 108−124.
41. Nitenberg, M.; Makshakova, O.; Rocha, J.; Pèrez, S.; Marèchal, E.; Block, M. A.; Girard-Egrot, A.; Breton, C. Mechanism of Activation of Plant MonoGalactosylDiacylglycerol Synthase 1 (MGD1) by Phosphatidylglycerol. *Glycobiology* 2020, *30*, 396−406.
42. Makshakova, O.; Breton, C.; Pèrez, S. Unraveling the Complex Enzymatic Machinery Making a Key Galactolipid in Chloroplast Membrane: A Multiscale Computer Simulation. *Sci. Rep.*

2020, *10*, 13514.
43. Romero-García, J.; Biarnès, X.; Planas, A. Essential Mycoplasma Glycolipid Synthase Adheres to the Cell Membrane by Means of an Amphipathic Helix. *Sci. Rep.* 2019, *9*, 7085.
44. Pèrez, 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. P.; Samsonov, S. A.; Vivès, R. R.; Nikitovic, D.; Ricard-Blum, S. Glycosaminoglycans: What Remains To Be Deciphered? *JACS Au* 2023, *3*, 628−656.
45. Dyer, D. P.; Migliorini, E.; Salanga, C. L.; Thakar, D.; Handel, T. M.; Richter, R. P. Differential Structural Remodelling of Heparan Sulfate by Chemokines: The Role of Chemokine Oligomerization. Open Biol. 2017, 7, 160286.
46. Migliorini, E.; Thakar, D.; Kuẗhnle, J.; Sadir, R.; Dyer, D. P.; Li, Y.; Sun, C.; Volkman, B. F.; Handel, T. M.; Coche-Guerente, L.; Fernig, D. G.; Lortat-Jacob, H.; Richter, R. P. Cytokines and Growth Factors Crosslink Heparan Sulfate. Open Biol. 2015, 5, 150046.
47. Seror, J.; Merkher, Y.; Kampf, N.; Collinson, L.; Day, A. J.; Maroudasn, A.; Klein, J. Articular Cartilage Proteoglycans as Boundary Lubricants: Structure and Frictional Interaction of Surface-Attached Hyaluronan and Hyaluronan–Aggrecan Complexes. Biomacromolecules 2011, 12, 3432–3443.
48. Richter, R. P.; Baranova, N. S.; Day, A.; Kwok, J. C. Glycosaminoglycans in Extracellular Matrix Organisation: Are Concepts From Soft Matter Physics Key to Understanding the Formation of Perineuronal Nets? Curr. Opin. Struct. Biol. 2018, 5, 65–74.
49. Burla, F.; Tauber, J.; Dussi, S.; van der Gucht, J.; Koenderink, G. H. Stress Management in Composite Biopolymer Networks. Nat. Phys. 2019, 15, 549–553.
50. Chiti, F.; Sablina, A. Editorial Overview: Folding and Binding. Curr. Opin. Struct. Biol.
51. 2022, 74, 102359.
52. Uversky, V. N. Protein Intrinsic Disorder and Structure–Function Continuum. Prog. Mol. Biol. Transl. Sci. 2019, 166, 1–17.
53. Uversky, V. N. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics.
54. Front. Phys. 2019, 7, e00010. https://doi.org/10.3389/fphy.2019.00010.
55. Ricard-Blum, S.; Couchman, J. R. Conformations, Interactions and Functions of Intrinsically Disordered Syndecans. Biochem. Soc. Trans. 2023, 51, 1083–1096.
56. Sun, F.; Schroer, C. F. E.; Palacios, C. R.; Xu, L.; Luo, S. Z.; Marrink, S. J. Molecular Mechanism for Bidirectional Regulation of CD44 for Lipid Raft Affiliation by Palmitoylations and PIP2. PLoS Comput. Biol. 2020, 16, e1007777.
57. Fruẗhbauer, B.; Mateos, B.; Konrat, R. 1H, 15N, 13C Resonance Assignment of the Human CD44 Cytoplasmic Tail (669–742). Biomol. NMR Assignments 2019, 13, 109–113.
58. Camponeschi, C.; Righino, B.; Pirolli, D.; Semeraro, A.; Ria, F.; De Rosa, M. C. Prediction of CD44 Structure by Deep Learning-Based Protein Modeling. Biomolecules 2023, 13, 1047.