Computational Modeling in Glycoscience

Preprint · November 2020

DOI: 10.13140/RG.2.2.12083.45607

3 authors:

Serge Perez
French National Centre for Scientific Research - Grenoble University Alpes
409 PUBLICATIONS **16,837** CITATIONS
SEE PROFILE

Elisa Fadda
National University of Ireland, Maynooth
52 PUBLICATIONS **2,343** CITATIONS
SEE PROFILE

Olga N. Makshakova
Kazan Institute of Biochemistry and Biophysics
24 PUBLICATIONS **243** CITATIONS
SEE PROFILE

Computational Modeling in Glycoscience

Serge Perez[a], Elisa Fadda[b], and Olga Makshakova[c], [a] Centre de Recherche sur les Macromolecules Vegetales, University of Grenoble Alpes, Centre National de la Recherche Scientifique, Grenoble, France; [b] Department of Chemistry and Hamilton ; Institute, Maynooth University, Maynooth, Ireland and [c] Kazan Institute of Biochemistry and Biophysics, FRC Kazan ; Scientific Center of Russian Academy of Sciences, Kazan, Russia

1. Introduction

Glycoscience is an interdisciplinary field involving biochemistry, polymer chemistry, material science engineering, physiology, developmental biology, microbiology, medicine and ecology. The exploration by glycoscientists of the physicochemical and biological functions of carbohydrates sustains a vibrant field of research worldwide. Within such a large field of disciplines and the expected advances of fundamental knowledge and translational applications, there is a contribution devoted to the determination of the three-dimensional (3-D) structure of complex carbohydrates, carbohydrate polymers, and glycoconjugates. The understanding of the molecular basis underlying their associations and interactions represents the main challenges in structural glycosciences. Elucidation of the 3-D structures and the dynamical properties of oligosaccharides is a prerequisite for a better understanding of the biochemistry of recognition processes and the rational design of carbohydrate-derived drugs. Seemingly, the elucidation and the understanding of the different structural levels of polysaccharides are required to relate structure to properties. Ultimately, some polysaccharides are also carriers of biological information that can only be deciphered if their interactions with other biological macromolecules are understood. Oligosaccharides, either in their free form or as part of glycoconjugates, are inherently difficult to crystallize and structural data from X-ray studies are sparse.[1]

In solution, the flexibility of glycosidic linkages produces multiple conformations which coexist in equilibrium. The use of several spectroscopic methods, with appropriate time resolution, is necessary for the analysis of the conformational behavior of such molecules. As for polysaccharides, they differ from other biological macromolecules as the diffraction data are not sufficient to permit crystal structure determination based on the data alone. Hence, procedures for molecular modeling of carbohydrates and carbohydrate polymers constitute essential tools for structural studies of these compounds. The progress made in algorithms and computational power now allows for the simulation of carbohydrates in their natural environment, that is, solvated in water or organic solvent, in concentrated solution, or the binding site of a protein receptor. Within an affordable time of computation, new dimensions, both spatial and temporal, can be assessed (**Fig. 1**).

In the first edition of Comprehensive Glycoscience, the review chapter devoted to the presentation of the computational procedure covered the significant advances in the elucidation and understanding of the conformation and dynamics of glycans.[2]The different concepts and tools were described thoroughly, including those aimed at characterizing their interactions with other like molecules, solvent or ions. These are now fully established, and regular practitioners commonly use some of them.

Nevertheless, the present highly updated review still keeps some sections, which we consider essential to mention, that appeared in the first edition. These sections describe the Heuristic approach; the Monte Carlo Method, the Genetic Algorithm search. Another paragraph focuses on carbohydrate–water interactions.

Over a 15 year lap time, advancements in high-performance computing (HPC) have allowed molecular simulation methods not only to play a more substantial role in supporting experiments but to transcend such mandate to guide experimental design and to lead autonomously scientific discovery. In this review, we will present the foundations underlying the multiscale molecular simulations methods, from atomistic to coarse-grained, used in complex carbohydrates research. It is also the aim of this chapter to illustrate how these new developments contributed to significant advances in the field of molecular modeling, restricting the examples to well-documented cases where conformational characterizations have been achieved. We also take into consideration their relevance to significant biological functions and properties. Other

Comprehensive Glycoscience, 2nd edition, https://doi.org/10.1016/B978-0-12-819475-1.00004-3

biologically active oligosaccharides and glycoconjugates have been reported, but are not included, thus avoiding the risk of transforming this chapter into a catalogue.

2. Structural diversity and conformational challenges

Monosaccharides are the chemical units from which all members of the significant family of natural products, the carbohydrates, are built. They are the individual carbohydrate building blocks, i.e. the monomeric constituents of more complex architectures are referred to as glycans, an assembly of sugars either in free forms or attached to another molecule or macromolecule. The presentation will use the term " glycan" to "refer to any kind of mono-, oligo-, polysaccharide," either free or covalently attached to another molecule. Glycans occur as (i) oligosaccharides (comprising 2–10 monosaccharides linked together either linearly or branched); (ii) polysaccharides (for glycan chains built up from more than 10 monosaccharides but the distinction with oligosaccharides is not strict); (iii) glycoconjugates (when there is a covalent linkage between the glycan chains and the proteins (glycoproteins), lipids (glycolipids) or naturally occurring aglycones (e.g. in antibiotics, saponins, alkaloids)).

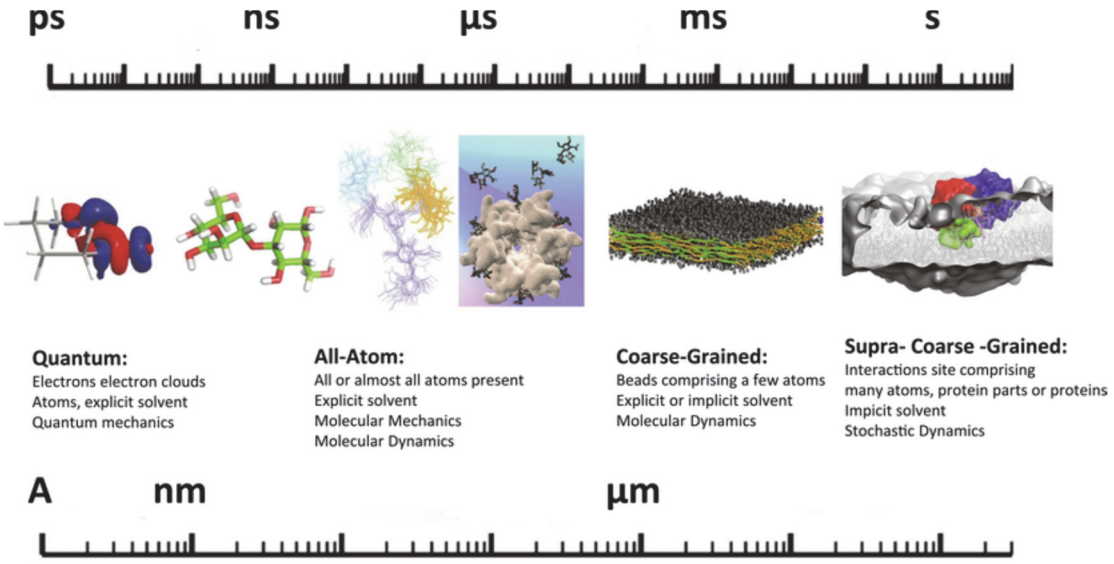

Fig. 1 Multiscale modeling in glycoscience. Applications range for molecular modeling at different resolutions: Quantum; All Atoms; Coarse-Grained and Supra-Coarse Grained. The axes display approximate ranges of times scales and system sizes.

Glycans have a potential of information content that is several orders of magnitude higher than any other biological molecules. The diversity of glycan structures results in part, from the broad range of monomers of which they are comprised (> 100) although some higher number has been reported.[3]

Monosaccharides with five or more carbon atoms generally can form both open and cyclic structures. Pentoses such as ribose form five-membered furanose rings, whereas hexoses such as glucose form mostly six-membered pyranose rings. Upon cyclization, the C1carbon atom becomes chiral, giving rise to either α- or β-anomer; an unbound monosaccharide can interconvert between the α- and β-forms, via the acyclic forms. As chemical compounds, monosaccharides and glycans different levels of descriptions are used for representations. Among the several descriptions, the use of a graphical representation called SNFG (Symbol Notation for Glycans) has been proposed and widely accepted as the result of a collective international agreement.[4] Such an extension and utilization of the graphical representation of glycans is a remarkable milestone that offers a unified way to communicate and describe pictorially, the diversity of glycan structures. The structural information encoded in the SNFG representation of glycans.[4] is not sufficient for characterizing, building and handling three-dimensional structures. For a given configuration (D) or (L) monosaccharide can occur as α-pyranose, β-

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

pyranose, α-furanose, β-furanose, i.e. under eight different 3-D structure. Such an extension initially presented in *Glycopedia*,[5] requires a limited set of rules (as illustrated in **Fig. 2**). While maintaining the spirit of using the symbolic representation for monosaccharides (and towards glycans) this set of rules provides the necessary extension to the construction of three-dimensional structures, allowing encoding for computational manipulation while maintaining IUPAC nomenclature. Because of differences in anomeric configurations (α or β), stereochemistry (D or L), and the variety of possible linkages (regiochemistry), the resulting number of different complex glycans and polysaccharides with the same sequence is staggering. Interestingly, not all possible sequences occur naturally, strongly indicating that specific glycoforms may result in precise biological functions.

For example, there is a limited number of monosaccharides occurring in mammalian glycans (see **Table 1**). Nevertheless, even a small number of monosaccharide units can provide a large number of different oligosaccharides, including branched structures, a unique feature among biomolecules. The reasons for the occurrence of so many three-dimensional structures made from a small number of monosaccharides come from the extraordinary conformational flexibility of glycans. The accumulations of experimental data, along with the detailed structural analysis of the conformational features, have provided some general rules that govern the structure and flexibility of glycans

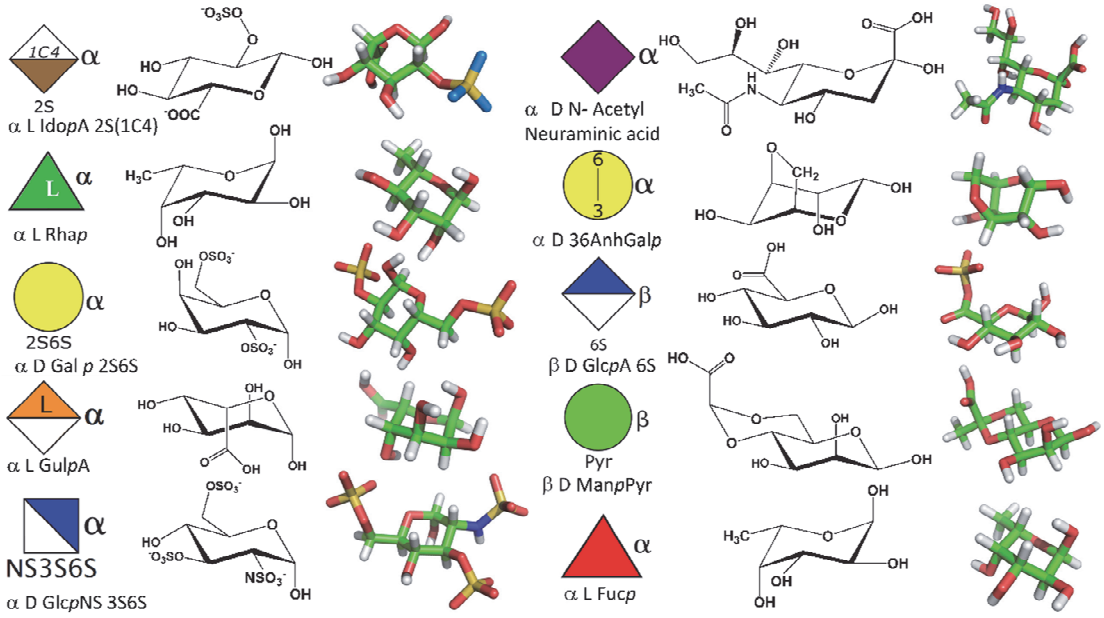

Fig. 2 From symbol representation to 3-D structures and its extension to 3-D structures Extension of the SNFG cartoons to include O-esters and ethers, which are attached to the symbol with a number (e.g., 3S for 3-O-sulfate groups, 2P for the 2-O-phosphate group), and the nature of the absolute (D or L) and anomeric configurations (α or β). All pyranoses in the D configuration were assumed to have the $^4C_1$ chair conformation, whereas those in the L configuration were assumed to have the $^1C_4$ chair conformation. The descriptors of the ring conformations adopted by idopyranoses ($^1C_4$, $^4C_1$ and $^2S_0$) were included within the monosaccharide symbol. The addition of the ring conformation in the monosaccharide symbol is the only discrepancy with the version of the SNFG nomenclature.

Table 1 The occurrence of the monosaccharides in mammalian glycans.[6]

Monosaccharide	Abundance	Monosaccharide	Abundance
D-Glc*p*NAc	31.8%	D-Gal*p*	24.8%
D-Man*p*	18.9%	Neu*p*5Ac	8.3%
L-Fuc*p*	7.2%	D-Gal*p*NAc	4.8%
D-Glc*p*	2.5%	D-Glc*p*A	0.3%
D-Xyl*p*	0.1%	L-Ido*p*A	0.1
Others	1.2%		

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

3. Computational concepts and tools

3.1 From quantum chemistry to coarse-grained calculations

Characterization of the structural and dynamic features of carbohydrates constitutes a challenge, both from the theoretical and experimental point of view. The conformations of complex carbohydrates depend on (1) the sequence and nature of the monosaccharides in the complex glycan (i.e. glucose vs. mannose), (2) the anomeric centers (i.e. α vs. β), (3) the linkage positions (i.e. 1–3 vs. 1–4), and (4) the chemical modification of the core structure (i.e. sulfation, phosphorylation, methylation, acetylation). There is a vast number of Quantum Mechanics (QM) methods available, and their detailed description is beyond the scope of this review. Many of them applied to carbohydrates of limited molecular size. In the last decade, density functional theory (DFT) methods replaced standard Hartree Fock methods. These methods were instrumental in dealing with the highly polar and the consequences of such stereoelectronic effects as the anomeric, *exo*-anomeric and *gauche* effects on the structure and reactivity of small carbohydrates.[7]

As carbohydrates and their derivatives possess many hydroxyl groups, their structures offer a large number of rotatable bonds, which, on top of the torsional movements occurring at the glycosidic linkages (Φ, Ψ, OME torsion angles), are sources of conformational flexibility. Besides, the orientation of such hydroxyl groups relative to the sugar ring is at the origin of the existence of hydrophilic patches (formed by polar hydrogens) and hydrophobic patches (formed by nonpolar aliphatic protons). Such features result in an anisotropic solvent density around carbohydrate molecules.

Computational methods are applied extensively for the exploration of the conformational space of glycans and polysaccharides. Appropriate energy functions and parameter sets are available in the literature. Some of them have the capability of treating carbohydrates in interactions with proteins considering solvation.[8] These methods encompass different approaches such as the Heuristic approach, Monte Carlo method and genetic algorithm. Over the last decade, molecular dynamics became the method of choice, in its All-Atoms representation and Coarse-Grained approximation. The theoretical and technological advances, often used in conjunction with diffraction methods, high-resolution spectroscopy, and other spectroscopic methods, provide a way to reconcile the experimentally available data and to predict structural and dynamical features that might not be accessible, yet. In such a context, molecular modeling should address three fundamental questions:

1. What are the most appropriate force fields and concomitant parameters to use?
2. What is the most satisfactory and efficient way to travel through the conformational hyperspace?
3. What is the appropriate way to calculate, from a modeling study, the spectroscopic observables for which experimental data are available?

3.1.1 The all atoms representation in molecular mechanics and molecular dynamics computations

Within the approximation that molecular structure and dynamics can be described sufficiently accurately by the laws of classical mechanics, atoms are represented as hard, impenetrable spheres, characterized by a mass, size (van der Waals radius) and electrostatic point charge, different for each specific atom type (**Fig. 3**).

The force field parameters values determine the molecule's potential energy corresponding to each specific configuration. The choice of a carefully developed and thoroughly tested force field is at the very core of the validity of any Molecular Mechanics or Molecular Dynamics simulation. In the case of complex carbohydrates, the choice of all-atom empirical force fields rests between two types of parameter sets, namely Amber/GLYCAM,[11] and CHARMM.[12] They are both reasonably complete, in terms of the ability to cover the vast majority of common monosaccharides and to treat different glycosidic linkages and branching complexity.[13] To also note the recent advancements in the development of the GROMOS 53A6 (GLYC) parameter set for the simulations of carbohydrates within a united-atoms CHARMMframework [14–16] and also the availability of Drude polarization models [17-19]

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

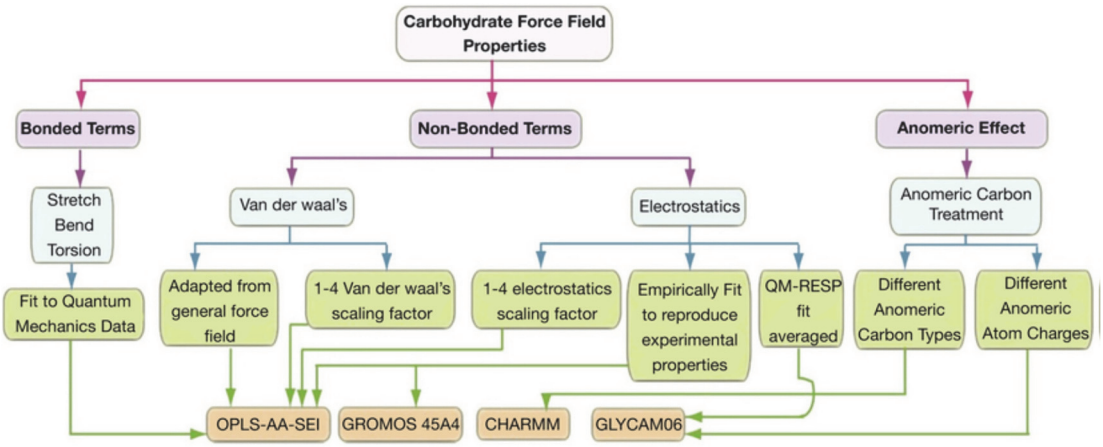

Fig. 3 Principle of calculation of the potential energy of a molecule for a Molecular Mechanics (MM) Molecular Dynamics (MD) investigation, along with some parametrization protocol comparison between the carbohydrate force-fields. Within the MM formalism, the molecule potential energy V(r) is a function of the positions of the N atoms that make up the system. It can be defined by an empirical force field, which general functional form is given by

$$V(r) = \sum_{bonds} k_b\,(r - r_0)^2 + \sum_{angles} k_a\,(\theta - \theta_0)^2 + \sum_{torsions} \sum_{n} \frac{1}{2} V_n [1 + cos(n\omega - \gamma)]$$
$$+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\}$$

The first three terms correspond to the potential energy contributions from bonded interactions, such as covalent bonds, bond angles and torsions. These are all represented by Hooke-type potentials, where bond vibrations, angle bending and dihedral angle potentials are regulated by classical spring constants ($k_{b/a}$) that modulate the stretch from an equilibrium distance (r_0) and an equilibrium angle (θ_0). Torsion potentials are generally represented through a Fourier series. The last term is the contribution to the total potential energy from nonbonded (or noncovalent) interactions, namely dispersion and electrostatic interactions, represented by a Lennard-Jones (L J) and by Coulomb potentials, respectively. In the L J term, ϵ is the energy well's depth at the equilibrium distance σ, while in the Coulomb term q values indicate the partial charges on the atoms i and j, and ε_0 is the permittivity of vacuum.The treatment of dispersion interactions with alternative formalisms to the LJ potential, such as the Buckingham potential,[9] is also possible in some software packages and can be used, provided that ad hoc parameters are available. Nevertheless, for most applications in biomolecular structure and dynamics, the L J potential is sufficiently accurate and also generally preferred to the Buckingham more rigorous treatment of dispersion for the sake of computational speed. Long-range electrostatic forces are treated within periodic boundary conditions[10] within the framework of the Particle-Mesh Ewald (PME) approach. PME electrostatics eliminates artefacts due to the abrupt truncation of the Coulomb potential through the introduction of cut-off values and/or effects due to the finite size of the simulation box.

3.1.2 Molecular mechanics

The applications of the principles described previously lead to the investigation and the description of the essential parameters of the shape and conformation of monosaccharides and disaccharides. Ring shapes can be defined in terms of reference conformations (chair, C, twist, T, boat, B, envelope, F, skew, S) or by the so-called puckering parameters. The glycosidic torsion angles Φ and Ψ describe the relative orientation of two consecutive monosaccharide units in a disaccharide moiety. For a (1-x) linkage Φ = O5-C1-O-C'x and Ψ = C1-O-Cx'-Cx'+ 1. The description of the conformation about α (1–6) linkage, requires a third descriptor, OME, referring to the O6-C6- C5-O5 torsion angle. The sign of all torsion angles lies within − 180 degrees to + 180 degrees. The energetically favorable conformations of a carbohydrate dimer may be shown on energy plots called (Φ, Ψ) maps which are somewhat similar to the Ramachandran plots used to visualize the backbone dihedral angles of the constituent amino acids in proteins. These plots feature multiple minima with the separating energy barriers being over 10–15 kcal/mol (**Fig. 4**).

Other critical difficulties, albeit less frequently addressed by an average practitioner, are the different ring shapes that might take either five or six-membered rings as a result of conformational flexibility. A set of two or three parameters (the so-called puckering parameters) describes five and six-membered ring, respectively. In the case of five-membered rings, the pseudo-rotational wheel represents the 20 twists and envelop shapes. As for six-membered rings, three parameters control the shape of the ring. Deviations occur away from the most commonly observed $^4C_1$ conformation, as in the case of Idose configurations, or the ring distortion of glycosides upon the mechanism of their enzymatic hydrolysis[21] (**Fig. 5**).

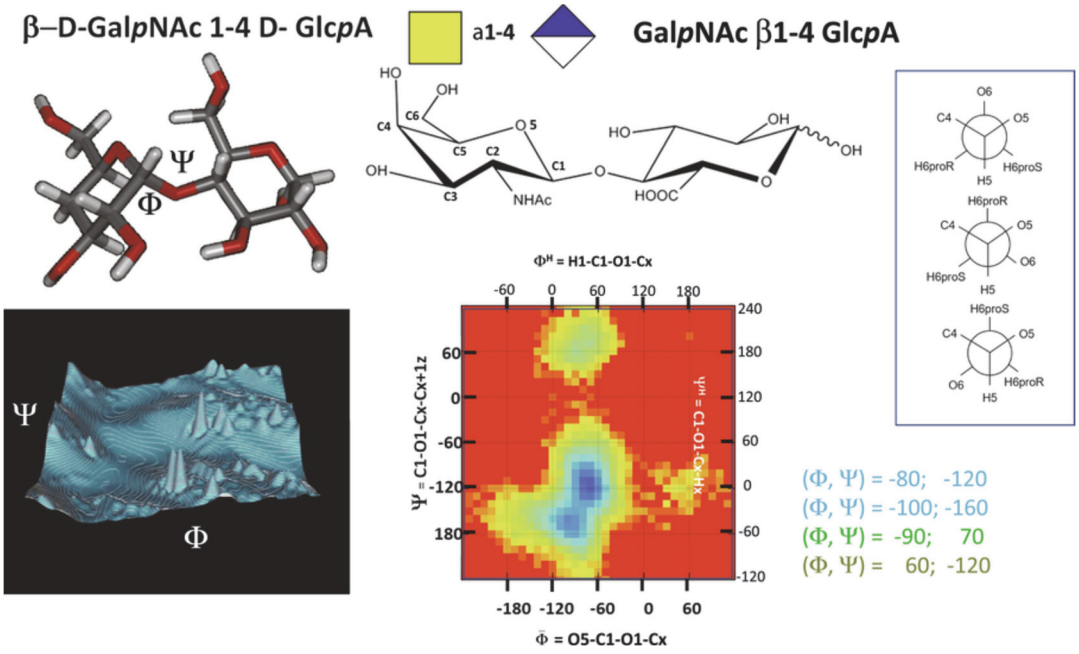

Fig. 4 Molecular representation of the structural descriptors of a disaccharide (conformation of the *exo*-cyclic and the glycosidic torsion angles. On the right-hand side are displayed the conformations of the exo-cyclic primary hydroxyl group of a pyranose is locked to the C6 carbon atom (whereas the secondary hydroxyl groups are linked to the ring carbon atoms). For such an exocyclic hydroxymethyl group, three staggered situations about the CH-CH_2OH bond (the C5-C6 bond in aldohexoses) can be considered. As the CH-CH_2OH bond is prochiral, the two hydrogen atoms need to be differentiated based on the R/S system. The description of the three rotamers orientations of O6 concerning O5 and C4 is GG (gauche-gauche), GT (gauche-trans) and TG (trans-gauche) depending upon the choice of an atom of reference.[20] The central part of the figure displays the potential energy surface showing conformational energy for the Φ and Ψ torsion angles. The favored low-energy Φ/Ψ combinations are shown in a light color, while the high-energy regions are in red.

3.1.3 The all atom molecular dynamics

Once the choice of force field sets the potential energy of the system, the conformational space accessible to the glycan is studied over a time coordinate through MD simulations. MD simulations generate an ensemble of configurations by applying the laws of motion to the atoms of the molecule. The concept behind MD simulation involves calculating the displacement coordinates in time (trajectory) of a molecular system at a given temperature. Finding positions and velocities of a set of particles as a function of time is done classically by integrating Newton's equation of motion in time. Within the context of MD simulations, the dynamic evolution of the system is obtained through a deterministic approach, namely by solving Newton's equations of motion, through numerical integration.

$$m\vec{a} = -\frac{dV(\vec{r})}{d\vec{r}}$$

The integration time step is an infinitesimal δt to guarantee the conservation of energy. In practical terms, the δt value has to be as one order of magnitude lower than the fastest bond vibration, which is approximate of 10 fs for a C- H bond stretching mode.[23]

Constraining the distance of all bonds to hydrogen atoms, including water molecules, with schemes such as SHAKE[24] or LINCS[25] allows to increase the time step to 2 fs, thus to increase the sampling capability. Several different numerical integrators are available for all-atom deterministic simulations. Such choice determines the reproducibility of the data collected, more specifically the reversibility of the integration discrete steps, and ultimately the conservation of the total energy.

Molecular simulations are usually carried out as a micro-canonical (constant-NVE) or canonical (constant-NVT) ensemble. As a consequence, all other thermodynamic quantities must be determined by ensemble averaging. In a classical system, Newton's equations of motion conserve energy and thus provide a suitable scheme for calculating a micro-canonical ensemble. However, the canonical ensemble is performed by coupling the molecular system to a constant-temperature bath, which rescales the atomic velocities according to the desired temperature. In Constant-pressure simulations can be performed by scaling through coupling to a constant-temperature position, as the pressure can be calculated from the virial theorem. For an overview of the different types of

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

integrators and temperature coupling schemes available, we refer the reader to a recent comprehensive and very insightful review of molecular simulation methods[11] and references therein.

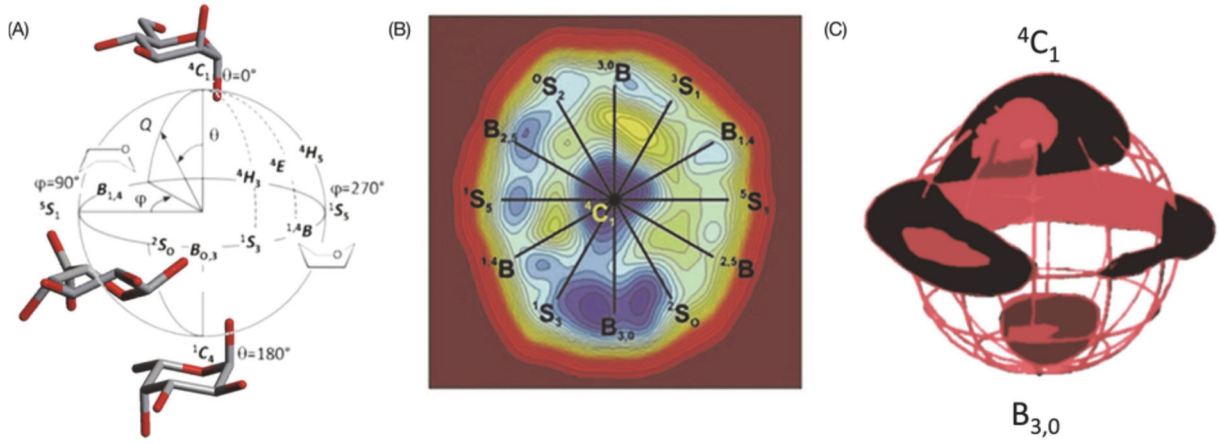

Fig. 5 Puckering Parameters. (A) Stoddard representation of the puckering sphere for six-membered rings, with the depicition of three low energy conformations: $^4C_1$, $^1C_4$ and $^2S_0$. (B) Conformational free energy landscape of an isolated β-D-glucose monosaccharide in Stoddard representations.Three carbohydrate force fields, namely GLYCAM06, GROMOS 45a4 and OPLS were evaluated. Free energies of different ring conformations of β-D-glucopyranose were calculated using metadynamics in vacuum as well as in explicitly modeled water. All three force fields model the $^4C_1$ conformation as the most stable by at least 6 kJ/mol, as compared to other conformations. Interconversion from the $^4C_1$ to any other conformation is associated with a barrier of no lower than 26 kJ/mol. The free energy surface calculated in the GLYCAM06 force field is in remarkably good agreement with the recent Car-Parrinello metadynamics study. The effect of a water environment is relatively low and analogous in all tested force fields. Namely, the presence of water stabilizes the upper-left (^{3,O}B) versus bottom-right ($B_{0,3,O}$) area of Stoddard's representation, relative to the situation in a vacuum. Comparison of free and potential surfaces is also provided for vacuum calculations[22] (C) same as in (B) depicted on the puckering sphere.

Once the mechanistic framework for the propagation of the system's dynamics is set, the question we often ask, or should ask, ourselves about our simulation is "*how long is long enough*?." Deciding on the optimal length of MD sampling is a complex, highly system size-dependent problem that is generally limited by the computational resources available to the researcher. Because most glycans are very flexible and dynamic, when unbound, they can potentially occupy a very high number of conformational states with nonnegligible populations. Therefore, to understand their molecular recognition by different receptors, it is essential to determine the correct statistical distribution of the different 3-D structures they can adopt, which requires exhaustive sampling of the conformational space. Because the transition between different free (unbound/unlinked) glycan conformers is generally not energetically hindered at room temperature (300 K), conventional MD schemes are quite capable of delivering the required sampling.

Nevertheless, it is highly advisable to run different MD simulations of the same glycan started from all potentially relevant conformations, especially if highly flexible linkages are present. Meanwhile, when studying the dynamics of glycans linked to glycoproteins or bound to lectins/enzymes, it is of paramount importance to realize that the starting conformation may be biasing sampling due to direct protein–carbohydrate interactions that may prevent transition to other relevant conformations. This should be especially concerning in (widespread) cases when the starting conformation of the bound/linked glycan is built from scratch, due to lack of experimental information. In these cases, the use of accelerated MD schemes, such as temperature replica exchange MD[26] (REMD), Hamiltonian REMD,[27] swarm MD,[28] or metadynamics[29] is highly advisable as it ensures a more exhaustive coverage of the conformational energy hypersurface.

The available computer power severely limits MD. It became feasible to perform a simulation with several thousand explicit atoms for a total time of up to the microsecond scale. Most of the published simulations have a duration of less than a microsecond. It is necessary to perform many such simulations to explore the conformational space adequately. Carbohydrate molecules may undergo dynamics event on longer time

scales. Standard molecular dynamics techniques cannot investigate such motions, and this limitation makes it difficult to compare situations that occur on a much larger time scale that generally occur throughout NMR experiments. Nevertheless, Molecular dynamics simulation is often used in conjunction with NMR experiments (primarily NOE, and residual dipolar coupling-based experiments) to generate likely carbohydrate structures informed by the experimentally derived constraints.

At present, the best approach is the inclusion of the environment in the simulation; that is, a molecular dynamics simulation with explicit water molecules or other surrounding molecules. Carbohydrates have a very high affinity towards water, with the majority of hydrogen bonding between water and carbohydrates occurring throughout their hydroxyl groups. The carbohydrates affect the surrounding water structure, and, in return, the water affects the structure of the dissolved carbohydrate molecules. Molecular dynamics provides the most promising way to investigate the hydration features of carbohydrates and set up a firm basis for docking simulations.

3.1.4 Molecular dynamics: Coarse-grained simulations

MD simulations scale linearly with the number of atoms in the system. Despite the enormous advancements in high-performance computing (HPC) technologies[30–33] the study of extensive molecular systems with atoms exceeding the million, is still prohibitive for routine simulations. The substantial computational costs result from a large number of degrees of freedom that require complex sampling schemes. A coarse-grained (CG) description of molecules which offers a progressive way of augmenting the spatial and temporal scale of simulations may be suitable and informative. In a CG representation, there is a reduction of the degrees of freedom. This reduction results from the merging (or graining) of multiple atoms together into “pseudo-atoms” or “beads” (**Fig. 6**).

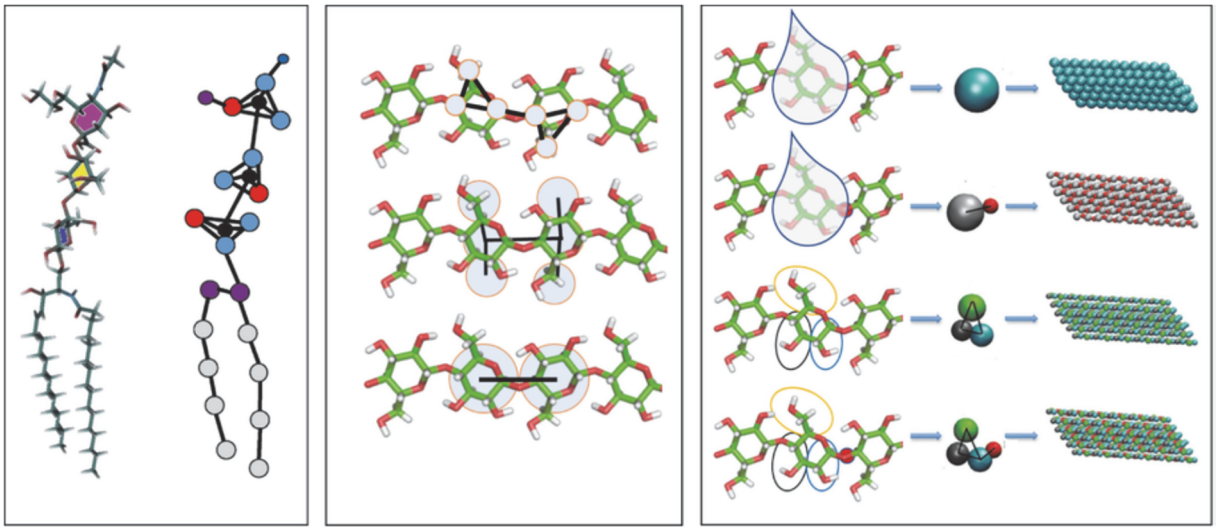

Fig. 6. Graining multiple atoms together into ≪pseudo-atoms≫ or ≪beads≫. The merging schemes vary depending on the specific approach used and on the level of precision required.34–37 The least invasive approach is to group the nonpolar hydrogen atoms with the heavy atoms they are bound to. A framework, called “united atoms,” yields up to a 10-fold reduction of the computational effort. Merging schemes can accelerate simulations by three to four orders of magnitude in comparison with classical all-atom MD simulations.[34,38,39]

In CG simulations the potential energy is determined by a force field defined through a set of equations analogous to those used in all-atom force fields. A CG force field contains terms for covalent interactions, namely bonds, angles and torsions between pseudo-atoms. Noncovalent terms account for dispersion interactions through an L J term, and electrostatic interactions through a Coulomb term. To represent covalent interactions between beads, the potential energy in the function of the bond length, angle and torsion, is fitted by a harmonic function.[40]

Owing to the merging of atoms into beads, the description of the energy of CG models requires additional terms beyond the classic empirical force field representation. Internal correlations between groups of atoms are introduced explicitly in the form of multi-body terms, which are computationally expensive.[41] Such depictions contribute to smooth the potential

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

energy landscape and thus accelerates sampling (Fig. 7).[38,42,43]

Most approaches keep the distinction between local energy terms and so-called contact potentials. The local energy terms describe spatial correlations between pseudo-bond vectors, which no longer follow classical harmonic behavior, and are often expressed by arbitrarily chosen functions. A single formula describes the nonbonded terms with proper formulation depending on the type of interacting atoms, their distance, and sometimes their mutual orientation and local neighborhood. Due to the diffused nature of a spherical cloud representing a group of physical atoms, these interactions tend to be softer, and the 12–6 exponents in the L J potential are parametrized accordingly. Finally, in CG simulations the solvent, most commonly water, can be represented implicitly,[44] through continuum electrostatics, further reducing the complexity of the system or explicitly.[45]

The MARTINI Force Field[36,46] is widely used for simulation of glycolipid membranes in conjunction with both monotopic and transmembrane proteins and also an extension for carbohydrates.[47] The philosophy behind MARTINI is to use universal modular building blocks, which makes relatively straightforward and effective to change the representation from all-atom to coarse-grain (CG) in a wide range of biological systems. The MARTINI force field uses a one-to-four mapping approach, where a single bead represents a group of four heavy atoms. Within this framework, one bead represents four water molecules. Small ring-like fragments, e.g. aromatic amino acid side chains, sugar residues, cholesterol, …, are mapped with a slightly higher resolution of up to two heavy atoms per bead. Overall, four main types of coarse-grained particles are defined according to their chemical properties, namely, polar (P), nonpolar (N), apolar (C), and charged (Q). They are divided into subtypes based on hydrogen-bonding capabilities, i.e. donor, acceptor, both or none, and polarity, ranging from 1 to 5. Their combination gives a total of 18 unique “building blocks.” Interactions between CG particles are modeled according to a classical mechanic’s approach with an interaction potential equivalent to the one used in atomistic simulations. Bonded interaction parameters are derived based on atomistic structural data or from iterative fitting to atomistic MD simulations of corresponding atom groups.

Nonbonded interactions are described by a Lennard-Jones (L J) 12-6 potential for dispersion and by a Coulomb potential for electrostatics, with parameters adjusted to reproduce experimental thermodynamics data of the free energy of hydration, free energy of vaporization and partition free energies between water and several organic phases for each of the 18 types of coarse-grained particles.[36,46] The development of a MARTINI force field follows the iterative fitting of the CG parameters that define bonded interaction potentials to atomistic trajectory data. At the same time, nonbonded terms are derived to reproduce experimental data. MARTINI combines bothstructure-based (or top-down) and thermodynamic-based (or bottom-up) parameterization strategies.

Within the MARTINI representation, water can be treated explicitly at the same level of coarse-graining as all other molecules, namely with four water molecules per bead. These water beads, just as many other CG water models, do not bear charges and consequently are blind to electrostatic fields and polarization effects. To compensate for the neglect of explicit polarization, screening of electrostatic interactions is done implicitly, assuming a uniform relative dielectric constant. While this is a reasonable approximation for bulk water, problems arise at the interfaces between water and other phases and in the vicinity of charged particles. Because of the implicit screening, the interaction strength of polar substances is underestimated in nonpolarizable solvents. A MARTINI-type polarizable water model was introduced by Yesilevskyy *et al.*,[48] where a three-bead model representing four water molecules, accounts for a reorientation of water molecules due to the polarization effects. Charged particles, referred to as Q type, represent ions. In the case of single atom ions, such as sodium and chloride ions, the first hydration shell is included in the CG representation, to which the full charge is assigned. The coordination numbers for ion pairs and ion solvent remain in reasonable agreement with atomistic data. Keeping in mind the difficulty of modeling of ions already with all-atoms force fields, the CG ion force field is only qualitatively accurate.

The loss of directionality of hydrogen bonding interactions prevents from using CG and MARTINI in protein folding studies. Depending on the level of ≪ graining ≫, the loss of atomistic resolution and dihedral structure within the CG bead prevents obtaining

information on stereoisomers and of puckering of the carbohydrate ring. However, providing a cautious CG parametrization, CG models can capture time-averaged properties and conformational space of atomistic systems of interest.[49] Nevertheless, careful parameterization protocols are of the essence as, for example, nonphysical aggregation of glycans represented by MARTINI can be caused by an imbalance of the nonbonded solute-solute, solute-water, and water-water interactions.[50] The latter can be fixed by reducing the depth of the potential well of solute-solute L J interactions. The efficient scaling relays on the experimental values of the second virial coefficient of osmotic pressure, which describes the deviation from the ideal behavior of a solution with solute molar concentration.[50]

The coarse-graining affects the calculation of the thermodynamic properties of a modeled system and in particular, the balance between enthalpic and entropic contributions. The reduction of the degrees of freedom, inherent in a CG model, affects the entropy of the simulated system. A reduction of the enthalpic terms compensates it. In turn, a CG model may accurately reproduce free energy differences, but contributing enthalpy and entropy values may be inaccurate. The main advantage of a CG model remains in the flexibility of the representation, in other words, in the number of ≪beads≫ chosen per carbohydrate unit. Going to a coarser representation, one sacrifices the more detailed features of the modeled process and regulate the thermodynamic contributions by appropriate scaling of the nonbonded interaction term. However, the transferability of the CG model to a wide range of concentrations and of molecule types, i.e. glycolipids and glycoproteins, to describe diverse and biologically relevant systems is essential. Up till now, except the generic MARTINI model, none of the CG force fields schemes can describe the precise arrangements of real biosystems.[51,52] MARTINI provides an intrinsically consistent CG approach, which and an extension to including different lipid types, sterols, sugars, peptides and polymers. This flexibility is especially important when dealing with complex sugar-based systems as functionalized glycomaterials and glycoconjugates or to explore the protein-carbohydrate interactions.

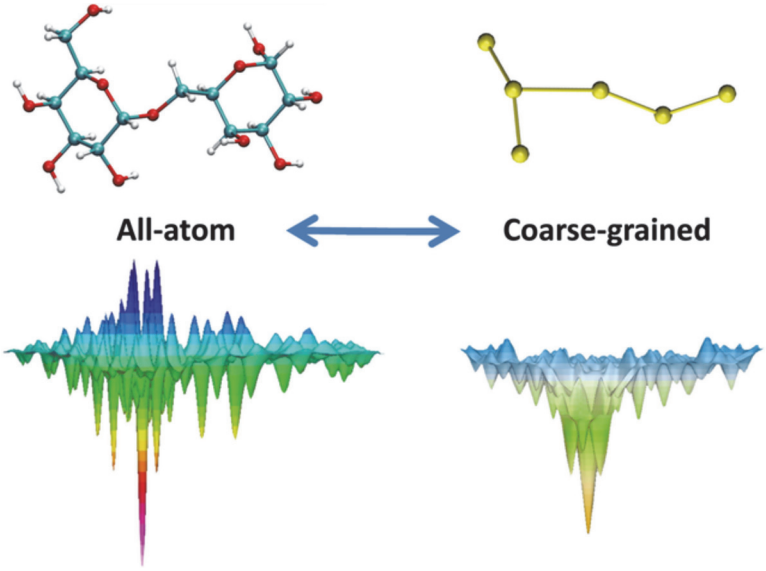

Fig. 7. Schematic view of the different all-atom and coarse-grained (CG) representations of a carbohydrate fragment and the corresponding potential energy surface. Note: lowest energies are shown in red and highest energies in blue

.

3.2 Heuristic approach

The potential energy surface is explored using single-coordinate driving approach with the CICADA[53] (Channels in Conformational Space Analyzed by Driver Approach) method. The potential energy surface is explored using single-coordinate driving approach [54]; each selected torsion angle is driven with a concomitant full-geometry optimization at each increment (except for the driven angle). It displays several advantages over other conformational searches. It has polynomial dependence of dimensions on computer time, in contrast to the systematic grid searches which have exponential dependence. The conformations are free of artificial harmonic constraint potentials. It overcomes all barriers among families of conformations on the conformational hypersurface but spends almost all of its time in the essential highly

populated areas. The inherent properties of the algorithm make rigorous optimization superfluous and provide good convergence behavior. It provides low-energy conversion pathways for estimating adiabatic rotational barriers.[55,56]

3.3 Monte Carlo method

The Monte Carlo method is essentially a random search method. From a starting configuration (A), a random displacement of one or more atom(s) generates a new configuration (B). The new configuration is either accepted or rejected, based on an energy criterion. When the energy of B is lower than or equal to that of A, B will be accepted. When it is higher, it will be accepted only if the Boltzmann factor (for the desired temperature) is higher than a random number taken from a uniform distribution between 0 and 1. When B is rejected, A is counted again before a new configuration will be generated; when B is accepted, it will serve as a new starting configuration. The process is repeated many times and results in a large number of configurations, which should be representative of the system. The method is more efficient for atomic or simple molecular systems than for complex (macro)molecular systems since a random displacement in the latter case will generally lead to such distortions of a molecule that the energy of a new configuration will usually be very high. Metropolis Monte Carlo methods have been applied to the conformational analysis of oligosaccharides with the aim of deriving ensemble average parameters.[57,58]

3.4 The genetic algorithm search

The field of genetic algorithms (GAs), evolutionary programming, and similar areas of computer science take their inspiration from biological evolution: survival of the fittest, the inheritance of traits from parents to offspring, mutation, crossover, and population genetics. Lamarckian evolution is based on the idea that the parent can pass on acquired traits to the offsprings. When applied to the 3-D structure prediction of complex carbohydrates, some of the terms translate to more common descriptions used throughout the modeling literature. A “population” is a set of “ individuals.” Each individual represents one specific 3-D conformation of an oligosaccharide. An individual can be said to carry several “genes.” Each gene represents one or several properties of that individual conformation. The genes translate to specific conformational variables, such as glycosidic linkages, torsion angles, primary hydroxyl conformations. The passing of genes between generations is done with a degree of “mutation,” that is, random changes to some conformation values. It can also be done with crossover, where the offspring receives its gene set from more than one parent. This crossover function, which combines whole segments of the parents’ conformational traits when creating the offspring, is one of the most powerful features of GA in comparison with other stochastic search methods.

The GA has been implemented in the GLYCAL software application[59] and for the automatic conformation prediction of carbohydrates.[60] They implement several different types of GA searches: standard GA and parallel GA with Lamarkian and natural evolution.[61,62] It has been applied to the 3-D characterization of complex oligosaccharides: some of the *Shigella dysenteriae* and *Escherichia coli* O-antigens,[63,64] one exopolysaccharide from *Burkholderia cepacia*,[64] and the structure of a *Schistosoma mansoni cercarial* surface polysaccharide.[65]

3.5 Carbohydrate–water interactions

Molecular Dynamics simulations of solute in water provide a unique possibility for a detailed examination of water–solute interactions. A large number of water molecules, the complexity of the solute, and the high degree of mobility in such simulations require a statistical approach to describe and analyze the hydration. The statistical approach commonly employed to examine water–water and water–solute interactions is a radial pair distribution function that calculates the probability of finding a pair of atoms at a distance r apart, relative to the probability expected for a random distribution at the same density. In complex aqueous simulations, the radial pair distribution function of specific solute nuclei and the oxygen nuclei in water is used to describe the significant differences in the first and second hydration shells of different types of solute atoms. The concept of the first hydration shell is very informative as it provides answers to two questions: (1) What is the total number of water molecules inside the first hydration shell of all solute oxygens? (2) Are shared water molecules a dominant feature of the solute structure?[66]

MD simulations provide a simple definition of the hydration number. From the trajectory, the number of water molecules on the first hydration shell (less than

3.5 Ang.) associated with the solute oxygen atom is counted in each phase point, yielding to an average value. Typically, for a disaccharide, the first hydration shell water molecules per solute molecule is about 27. To quantify and visualize shared water density among solute oxygen, it is possible for rigid molecules to calculate and contour 3-D water densities in a fixed frame defined by the solute. As for flexible oligosaccharides, the normalized 2-D radial pair distribution function is one such tool for analyzing the solute surrounding for localized water densities; this is particularly adequate for investigating bridging water molecules between two monosaccharide rings in carbohydrates.[67]

One can also assess the dynamic aspect of the first hydration shell as maximum, and calculate an average residence time for all water molecules around specified atoms. Seemingly residence times characteristic of water–water interactions can be assessed; they are typically about 0.5(2)ps. The average residence time of water molecules in their interactions with hydroxyl oxygen of carbohydrate is significantly higher (0.6 (2)–0.7 (2) ps), with some notorious exceptions, as in sucrose and trehalose, where residence times of some hydroxyl groups are 2.7–3.0 ps.[66] Adding water to an MD trajectory gives new properties to the solute, such as rotational and translational diffusion. By calculating the center-of-mass square displacement auto-correlation, the Einstein relation yields the self-diffusion of the solute. Typical values for disaccharides are (5.0–5.2) $10^{-6}cm^2s^{-1}$. The rotational diffusion and overall tumbling can be assessed in an analogous manner to the translational diffusion.

3.6 Description of the available software and applications

The complexity of the topology of complex glycans (and polysaccharides) requires the design of dedicated molecular building procedures that can rapidly convert the commonly used sequence information into preliminary but reliable 3-D Models. Particular procedures are available for such a task using libraries of constituent monosaccharides.

Besides several molecular builders implemented in stationary software as, for example, HyperChem and Maestro (Schrodinger software), some (free of charge) tools to build-up preliminary carbohydrate structure in All-Atom representation are currently available online.

Sweet (part of the Glycosciences.de portal) is a software for constructing 3-D models of saccharides from their sequences using standard nomenclature.[3]

Glycam-Web monosaccharides builder allows to constructing oligosaccharides from a library of monosaccharides, to introduce branches and derivatives. The structures are constructed using the parameters of GLYCAM06 force field.[11] To note a somewhat limited range of monosaccharides (almost the lack of bacterial-specific units). A new library for glycosaminoglycans now is available.[68] There is also a facility to attach a pre-built oligosaccharide to a protein. For all the structures, the server provides the file with GLYCAM06 topology in Amber format.

Polys2.0 contains a vast series of monosaccharide units in its database. It allows to build up complex carbohydrates and polysaccharides based on a set of allowed torsion angles for each pair of constructing, which are available from the illustrative phi-psi maps of disaccharides, i.e. in adiabatic approximation.[69]

CarbBuilder supports the building of polysaccharides with a constantly enriched list of monosaccharides[70]

CHARMM-GUI[71] is a graphical user interface that consists of several modules for the modeling of biomolecules in the CHARMM force field. The modules allow both read and build up structure from/to PDB format and generate inputs for popular MD simulations programs.

Besides, glycans can be assembled with a lipid core to produce both preliminary and lipopolysaccharides.[72] *do-glycan* is a desktop application that allows users to prepare carbohydrate structures. doGlycans toolset is composed of two features: (i) prepreader.py, for preparing the carbohydrate chains for polymer simulations and; (ii) doglycans.py that prepares models for glycoproteins and glycolipids. The toolkit creates topologies of carbohydrates in the GROMACS format and then provides 3-D structures of carbohydrates covalently linked to the given lipids and proteins.[73]

To construct the initial geometry of carbohydrates and glycoconjugates in Martini representation, one may use Martini Maker module CHARMM-GUI. This toolkit assists to convert All-Atom to Coarse-Grained representation or to construct monosaccharides structures from building “beads”. It also helps to pre-build monosaccharides lipid systems and pre-assemble glycoconjugates in the form of symmetrical or asymmetrical bilayers, vesicles, micelles, monolayers and nanodiscs. Another possibility to construct a

Coarse-Grained glycolipid bilayer with a possibility to include proteins in it is to use the INSANE (INSert membrANE) python script. [74]

4 Glycans and glycoconjugates

4.1 N- and O-linked glycans

Glycosylation is the most common and complex post-translational modification of biomolecules. In the specific case of protein glycosylation, the glycans that decorate proteins surfaces can be extremely different, in terms of size, branching complexity and monosaccharides sequence. Overwhelming evidence gathered through glycobiology research over the past 30 years shows that specific types of glycosylation are essential for correct and efficient protein function and the alteration of such motifs modulates protein structure and activity.[75–77] Unfortunately, the molecular basis underlying the different roles that specific glycoforms play is often unknown, while having such information would be a remarkable step forward in understanding the glycobiology of health and disease states. To this end, atomistic MD simulations provide unique insight, giving an accurate description of the 3-D structure and real dynamics of glycan motifs at the actual timescale when molecular events take place. Below we will briefly discuss a few examples of selected All-Atom MD simulation studies that contributed to our understanding of severe problems in glycobiology, such as carbohydrate recognition, of how specific N-glycosylation patterns modulate protein function, structure and dynamics and the structure and function of the human immunodeficiency virus 1 (HIV-1) glycan shield.

A glycobiology research area where atomistic simulations have great potential for discovery is in understanding how glycans' sequence affects the glycoprotein function. It is a particularly thorny problem for classic experimental structural biology approaches as glycans are highly dynamic when unbound. This conformational disorder is retained by protein-linked glycans that do not form extensive interactions with the protein surface. As a result, experimental data obtained from highly dynamic glycans are often sparse and underdetermined, and as a consequence, the corresponding glycan 3-D structures can be somewhat speculative, with highly questionable ring and glycosidic linkages conformations. Additionally, the very nature and degree of the glycans conformational dynamics depends on their sequence,[78] and that is an aspect that atomistic MD simulations can address quite exhaustively.

4.1.1 How core fucosylation affects the structure and ADCC function of IgG1s

One emblematic case where glycan sequence affects protein function is in immunoglobulins G1 (IgG1). The core fucosylation of the fragment crystallizable (Fc) N-glycans results in a dramatic quenching of the antibody-dependent cellular cytotoxicity (ADCC) with a consequent reduction of the antibody therapeutic efficacy.[79–82] A recent work assessed the role of sequence in unbound complex N-glycans 3-D structure through extensive sampling with a cumulative simulation time exceeding 64 μs.[83] The results of the work show that while core fucose does not affect the glycans' intrinsic dynamics, galactosylation of the α (1–6) arm does promote folding of the arm over the chitobiose core. This explains the known greater difficulty of sialylating the α (1–6) relative to the α (1–3) arm in free glycans and the differential recognition of positional isomers in glycan arrays.[84] Based on these results, it was clear that the effect of core fucose on ADCC quenching was dependent on how the architecture of the Fc glycosylated region affects the glycans dynamics and their availability for interactions with the FcγRIII receptors. Earlier crystallographic work[85] shows that core fucose weakens glycan–glycan interactions in the complex between the IgG Fc and FcγRIII. The work based on accelerated sampling through temperature replica exchange MD (REMD) simulations of Fc-linked complex N-glycans shows that, not only core fucosylation alters the dynamics and exposure to the solvent of the sialylated α (1–3) arms, but also that in doing so it enhances the overall Fc region dynamics.[78] Furthermore, the simulations show that core fucose residues in position within the Fc that in the complex with the FcγRIII is occupied by the FcγRIII N162 glycan (**Fig. 8**). It suggests an increased binding free energy required to clear the steric hindrance and displace the fucose upon FcγRIII binding[83] in agreement with earlier crystallographic work showing a clash-free IgG1 in complex with a nonglycosylated FcγRIII70.

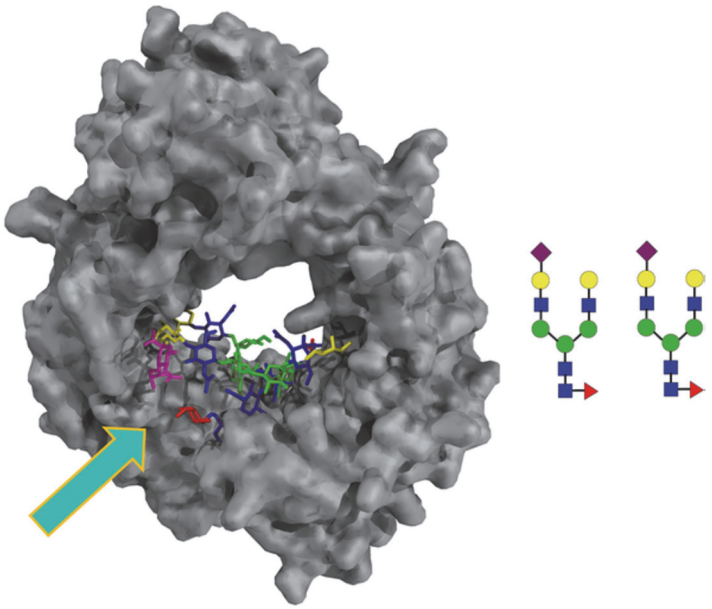

Fig. 8 Structure of the IgG1 Fc region (grey) from REMD sampling showing the conformation of the two N-glycans, sequence shown on the right-hand side, colored according to the SNFG nomenclature with the fucose obstructing the extension of the symmetric α (1–3) N-glycan's arm, in the Fcγ receptor's binding region, indicated with an arrow.

4.1.2 Structure and dynamics of the HIV-1-Env glycan shield

The ability of HIV to escape the humoral immune response is likely due to the high-density glycan cover of the trimeric envelope (Env) proteins complex that masks the virus from recognition.[86] Over 90 high-mannose N-glycans decorate the complex surface, an amount that corresponds to about half of the molecular weight of the entire glycoprotein. The majority of these N-glycans are Man-5 and Man-8/9, with a small population of Man-6/7. The crystal structure of the trimeric HIV-1 Env shows that these N-glycans outstretch from the gp120 surface through their chitobiose core, while the α (1–3) and α (1–6) arms extend perpendicularly interacting with neighboring N-glycans and forming a dense glycan–glycan interaction network that effectively shields the protein surface[87] (**Fig. 9**). This seminal crystallographic work was complemented by canonical MD simulations, more specifically by three 500 ns trajectories. The complexity of the system, in terms of its sheer size and of the intricacy of the N-glycans interaction network, makes the discovery potential of these simulations rather limited, mostly because of lack of sampling. For this reason, the structural and dynamic insight provided is not substantially incremental to the analysis of the crystallographic data. These canonical MD simulations were extended to 2 μs in later work,[89] which delivered complete information on the collective dynamics of the shield and protein surface accessibility. Notable in this particular context is the work by Yang *et al.*[90] that shows how enhanced sampling MD achieved through Hamiltonian Replica Exchange (HREX) is a much more informative approach. Indeed, the HREX simulations reveal a more complete image of the N-glycans conformational space and its heterogeneity, providing valuable insight on the pre-structuring of the N-glycans for recognition and binding by broadly neutralizing antibodies (bNAbs) and of their accessibility, as well as the accessibility of the protein surface to CD4 receptors and CCR5 and CXCR4 co-receptors.

4.2 Structure, function and dynamics of glycolipids in the plasma membrane

Glycolipids are vital components of cellular membranes. A hydrophobic lipid tail and one or more hydrophilic carbohydrate groups linked by a glycosidic bond make glycolipids. Their structures consist of a mono- or oligosaccharide moiety attached to a sphingolipid or a glycerol group with one or two fatty acids. These are glycosphingolipids and glycoglycerolipids, respectively. Glyceroglycolipids are characterized by acetylated or nonacetylated glycerol groups with at least one fatty acid as ceramide as the lipid complex. The acyl group of ceramides is a long chain saturated or monounsaturated fatty acids. The specific nature of the carbohydrate moiety defines further classes and subclasses, namely cerebrosides

(galactocerebroside, glucocerebrosides, sulfatides), gangliosides, globosides, glycophosphosphingolipids and glycophosphatidylinositols.

The carbohydrate moiety is the most exposed structures on glycolipids, sitting on the extracellular surface of cells. Their flexible structure with numerous binding positions makes them highly suitable for recognition and cell signaling

Since the lipid moiety is buried within the membrane, carbohydrate–carbohydrate interactions are the predominant interactions that occur between glycolipids. Because of their biophysical properties, glycolipids are also essential for stabilizing the membrane bilayer.

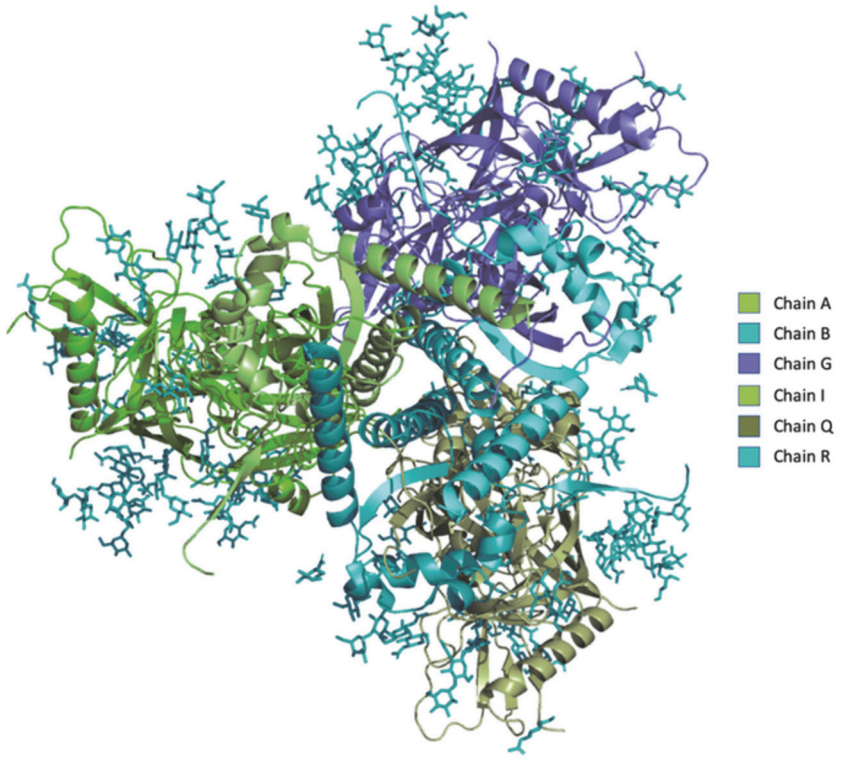

Fig. 9 Structure of the HIV gp120 (chains A, G, Q) and gp41 (chains B, I, R) from the cryoEM PDB 6NF2.[88] Glycans are shown in "sticks "mode and highlighted in cyan. Graphics through the PyMOL Molecular Graphics System, Version 1.2r3pre, Schrodinger.

.

Because of the reduced complexity inherent to the model, Coarse-Grained computer simulations have the potential to bridge the gap between experiment and All-Atom calculations for systems of high complexity, more specifically they are well-suited to study the dynamic nano-clustering of glycolipids in which the carbohydrate–carbohydrate interactions play a crucial role. A seminal work,[42] studied an idealized mammalian plasma membrane comprising 63 different lipid species, combining 14 types of headgroups and 11 types of tails asymmetrically distributed across the two leaflets. The 40 μs CG-MD trajectory, obtained based on the MARTINI representation, showed a general nonideal lateral mixing of the different species. Glycolipids, monosialotetrahexosylganglioside (GM1) and monosialodihexosylganglioside (GM3), also referred to as gangliosides, showed the highest non ideal mixing behavior, consistent with both in vitro and in vivo experimental data. The size of nano-domains in the CG model had a broad distribution of up to approximately 50 glycolipids, which were highly dynamic with individual clusters breaking apart and reforming at the microsecond time scale. Another group[91] reported that the formation of such GM3 nanoscale clusters correlated with spontaneous curvature of the membrane. The interactions between the gangliosides GM1 and GM3 were further studied using equilibrium MD simulations at both CG and atomistic levels. In this work, the MARTINI ganglioside force field was re-parameterized to reduce self-interactions of GMs and to bring the radial distribution function (RDF) closer to the atomistic results. For more information on mammalian cell membrane CG models, we refer the reader to a mini-review.[92]

5. Polysaccharides

Polysaccharides are the most abundant and diverse family of biopolymers. With several hundreds of known examples, they offer a great diversity of chemical structures. They range from simple linear homopolymers to branched heteropolymers, having repeating units that consist of up to octa-saccharides. Polysaccharides may also be branched, which is a unique feature among naturally occurring macromolecules. Depending on their primary structures, polysaccharide chains adopt characteristic shapes, such as ribbons, extended helices, hollow helices, ⋯, which characterize their secondary structures. Biosynthesis may also result in the formation of multiple helices. Also, a metastable structure occurs whenever polymerization and crystallization are concomitant. Some of these features may persist locally in the diluted state and may be directly responsible for the solution properties of some polysaccharides. Depending upon their origin, polysaccharides display a wide range of structures and architectural organizations that may develop over several orders of magnitude. The computational methods that have evolved to study small to medium size glycans have to extend to cope with such a diversity of situations. Simulations based on the principle of all-atoms representation are still used.

Nevertheless, the characterization of the structural features of polysaccharides has benefited from the development of Coarse-Grained in its capacity of simulating large-scale biomolecular system and eventually time scale, which are inaccessible to the All-Atom models. These applications concern mainly polysaccharides (i) from biomass origin, i.e. cellulose,[93–102]nano-cellulose,[103–105] interactions cellulose-polysaccharides,[106,107] chitin [40,108,109] (ii) from the extra-cellular-matrix; Glycosaminoglycans: GAGs. [110–114] and other polysaccharides such as α (1–3) glucan[115,116] In most cases, these applications investigate the interactions of these polysaccharides with proteins.

5.1 Polysaccharides in solution

The polysaccharide chains in solution tend to adopt a more or less coiled structure. Such a dissolved random coil would fluctuate between local and overall conformations. Polysaccharides assume an enormous variety of spatial arrangements around the glycosidic linkages because these molecules have extensive conformational freedom. Theoretical polysaccharides models are based on studies of the relative abundance of the various conformations, in conjunction with the statistical theory of polymer chain configuration.[117] Possible interactions between residues of the polysaccharide chain that are not nearest neighbors in the primary sequence of the polymer are ignored. A Monte Carlo sample reflects the range of conformations of polymer molecules. The observable properties of dissolved polysaccharides are averaged over the entire range of conformations accessible to the chain, and they may be determined from conformational states derived from the potential energy surfaces of the consecutive disaccharide fragments. This approach yields properties corresponding to the equilibrium state of the chain. Results refer to a model for an unperturbed chain that ignores the consequences of the long-range-excluded volume effect because only nearest-neighbor interactions are accounted for in the computation of the Φ and Ψ potential energy surfaces.

The following example illustrates the application of All-Atoms modeling to the characterization of one component of pectic substances (**Fig. 10**). Pectins are a family of polysaccharides that constitute a large portion of the cell wall of many higher plants where they influence growth, development, and senescence. They are extensively used as gel formers and thickening agents in the food industry. The backbone of pectin polysaccharide is formed by (1–4)-linked α-D-galacto-pyranosyluronate residues, either free or in ester form. These homogalacturonan sequences may be interspersed at intervals with rhamnopyranosyl residues carrying the major part of neutral sugar side chains, mainly arabinans, galactans, or arabinogalactans. These so-called rhamnogalacturonans-I moieties are the most diverse and structurally complex members. The characterization of their 3-dimensional structures and dynamic features of the constituents of RG-I highlights the occurrence of an extended threefold helical structure of the rhamnogalacturonan linear backbone. Branching helps to stabilize a conformer of the backbone twisted along 1 → 2 glycosidic linkages triggering the orientation of long side chains without altering the extended overall backbone chain conformation. Formation of an antiparallel pairing of

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

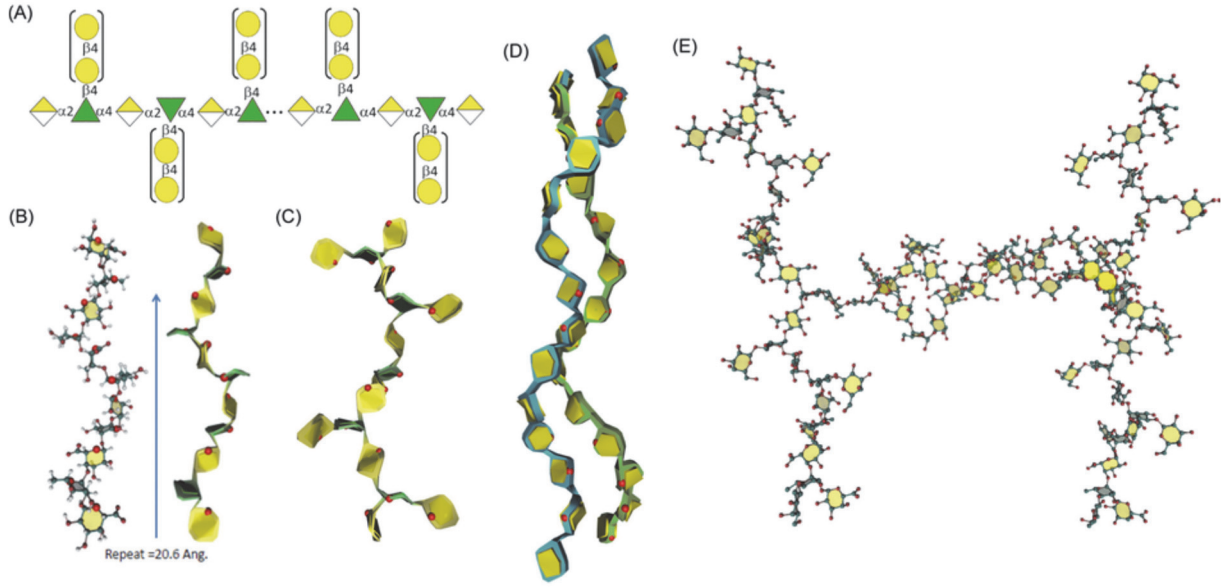

Fig. 10. Rhamnogalacturonan 1. From sequence to three-dimensional arrangement. (A) The SNFG representation of the primary structure of RG-1 showing the repeating units [→ 2)-α-L-Rha*p*-(1 → 4)-α-D-Gal*p*A-(1 →] of the backbone and the possible sites of branching, throughout the [β-D-Gal*p*-(1 → 4)-α-L-Rha*p*]-linkage of long galactan chains of different lengths. The lowest energy minima of the constituting elements along the backbone generate a threefold helical structure (B) which is not altered by the α-D-Gal*p* residue linked to O-4 of α-L-Rha*p* (C). (D) The β-D-Galactan chain form a sixfold helical structure which interact with another chain to form an association of anti-parallel galactan chains. (E) Summarizes the structural elements to suggest one possible mode of interaction of RG-1 in planta.[118]

the β-galactan side chains allows us to suggest a novel mode of noncovalent cross-linking in pectins.[118]

5.2 Lipopolysaccharides in membranes

The lipopolysaccharides (LPS) in the bacterial outer membrane present the most extensive and complex chemical diversity. Ma et al.[49] developed an LPS parameter set for a MARTINI representation that was benchmarked against available experimental data and atomistic simulations. This model was based on several approximations concerning primarily the oligosaccharide domain of the LPS molecule, which contains anomeric centers. Due to the loss of atomistic resolution within the CG bead, this model does not provide information on stereoisomers. The model was applied to more than 27 different membrane compositions for a cumulative simulation time of 100 μs. Increasing the LPS to 1,2-dipalmitoyl-3-phosphatidylethanolamine (DPPE) composition in the outer leaflet contributes to decreasing the phase transition values from 346 to 290 K, indicating a less ordered packing of the LPS molecules and increased area per lipid for the membrane. Such simulation outlines how chemical heterogeneity in membrane composition may result in significant variation in membrane properties such as fluidity and phase transition temperatures, which in experimental systems may vary by +/- 15 degrees. Also, to note, the response to the mechanical stress of the *E. coli* outer membrane was investigated by Jefferies et al.[119] The authors found that the different packing of terminal O-antigen chains affected lipid mobility and the mechanical strength of the Gram-negative membrane models. For a more detailed overview of recent progress in the understanding of bacterial cell envelope from CG modeling the reader is referred to a comprehensive perspective by Khalid *et al.*[120] Notably, the authors underline that the motions of LPS molecules are highly correlated witheach other and also with the outer proteins embedded within the membrane (**Fig. 11**).

5.3 Cellulose, nano celluloses and hemicelluloses

The degradation of cellulose and its applications to biofuels of the second generation are attracting considerable attention in developing computational methods able to construct very complex architecture and at the same time understand and predict mechanical properties. In this endeavor, several Coarse-Grained models using distinct ≪ beads ≫ representations provided important insights. For example, a simplified representation using ≪ bead ≫ for every monomeric glucose subunits, described the intrinsic conformational transition of long cellulose macro fibrils between crystalline and amorphous phases at long time scales.[94] Extension of the investigation examined the significance of the presence of an explicit solvent that showed the persistence

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

length of cellulose fibril in the transition region between fully crystalline and fully amorphous that corresponds to that of native cellulose fibrils.

Detailed analysis of the individual energetic contribution to the transition revealed that the nonbonded interactions, in particular, that of cellulose – water interaction, played a significant role in the observed crystalline to the amorphous transition of cellulose fibril.[95] A group of authors[99] derived a set of MARTINI coarse-grained force field parameters for the simulation of crystalline cellulose fibers. The model is adapted to reproduce different physicochemical and mechanical properties of native cellulose. The model is able not only to handle a transition between cellulose allomorph but also to capture the physical response to temperature and mechanical bending of longer cellulose nanofibers.[99]

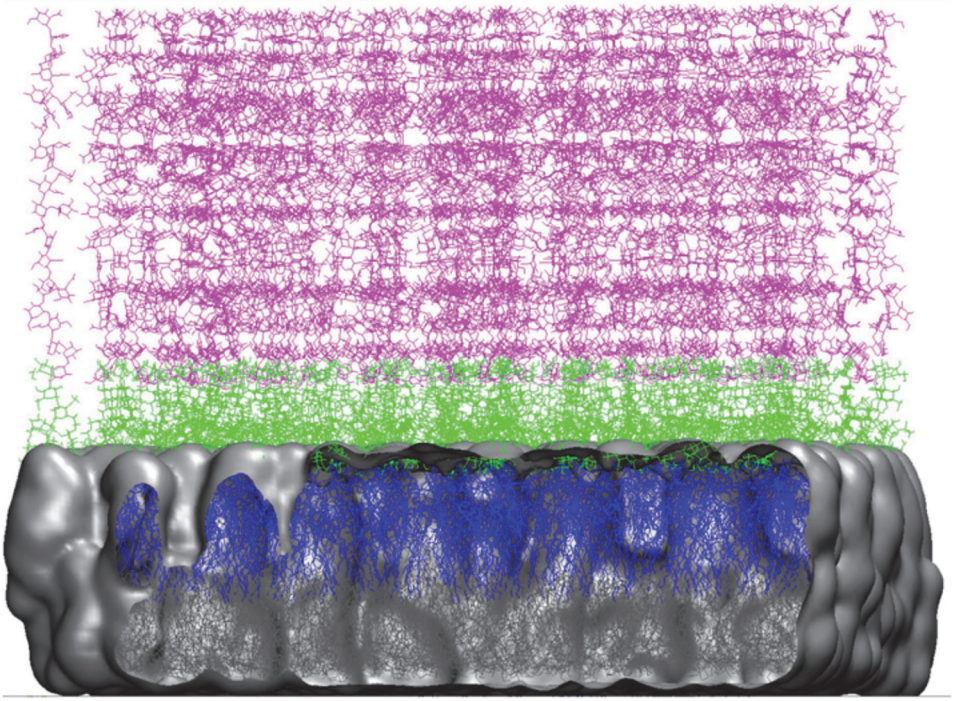

Fig. 11. All-Atom representation of model membrane of *Escherichia Coli*, DPPE lipids are shown as grey, for LPS moieties the following color coding was used: Lipid A (Type 1)—blue, Core oligosaccharide (R1,→ 3)βDGlc(1 → 3)[αDGal(1 → 2)αDGal(1 → 2)]αDGlc(1 → 3)αDGlc(1 → 3)[αDLHep(1 → 7)]αLDHep(1 → 3) αLDHep(1 → 5)[αDKdo(2 → 4)]αDKdo(2 →)—green, O-antigen (O1, → 3) [βDManNAc(1 →2)]αLRha(1 → 2)αDGal(1 → 3)βDGlcNAc(1 →)—magenta.

These few examples are among the numerous investigations where the developments of computational methods decipher complex macromolecular architectures which are related to physical and mechanical properties. At the same time, they lay the foundations to explore further the relationship between the crystalline morphology on topo-chemistry and topo-enzymology.

As biodegradable nanomaterials, Cellulose Nano Crystals (CNC) offer outstanding physical, chemical and mechanical properties for many applications. Coarse-Grained simulations provide a way to construct reliable structural models that can extend to the prediction of mechanical properties. For example, a Coarse-Grained model based on a representation of cellobioses units as one ≪ bead ≫ indicated what well-aligned CNCs lead to a more brittle and catastrophic failure mechanism, whereas naturally twisted interfaces promote toughening mechanisms that help attained optimal mechanical performance.[104] The same model was used to realize the effect of interfacial energy and twist on the mechanical performance shows that elastic modulus, strength, and toughness are more sensitive to twisted angle than interfacial energy.[105]

The modeling of the interactions occurring in the plant primary cell-walls starts with an elucidation of the interactions occurring between cellulose and hemicelluloses such as xylan and xyloglucan. The presentation of a Coarse-Grain model describing xylan and its interactions with crystalline cellulose highlighted how the complementarity of the chains

drove the interaction. The extended modeling revealed that the adsorbed xylan could also adopt coiled structures, especially when laying on the hydrophobic cellulose surfaces.[107]

5.4 Polysaccharide–protein interactions

Many polysaccharides occur in the form of highly packed 3-D arrangements as a result of extensive inter- and intra-molecular hydrogen bonding networks and van der Waals interactions. These features render the structures utterly insoluble in water (e.g. cellulose, chitin) and provide them with considerable recalcitrance from attack by most enzymes. Degradation of cellulose to glucose requires the cooperative action of three classes of enzymes, collectively known as cellulases. Spatial models of cellulose degradation must capture effects such as enzyme crowding and surface heterogeneity, which lead to a reduction in hydrolysis rates. (**Fig. 12**).

The first QM/MM study of the mechanism of cellulose hydrolysis dealt with the mechanism of endoglucanase (family GH8). DFT was used to describe the QM atoms and metadynamics to drive the chemical reaction and obtain a free energy landscape. A reduced number of collective variables was used. As a result, quantification of the reaction free energy barrier could not be assessed.[126] The simulations reproduced the concerted one-step general inversion mechanism. It confirmed the identity of the general base residue and the boat-type conformation of the transition state.

The complete reaction pathway for cellulose hydrolysis was investigated using transition path sampling. The outcomes of the investigation indicate that the deglycosylation proceeds via a product-assisted mechanism in which cellobiose interacts with a water molecule for nucleophilic attack on the anomeric carbon atom of the glycosyl-enzyme intermediate.[127]

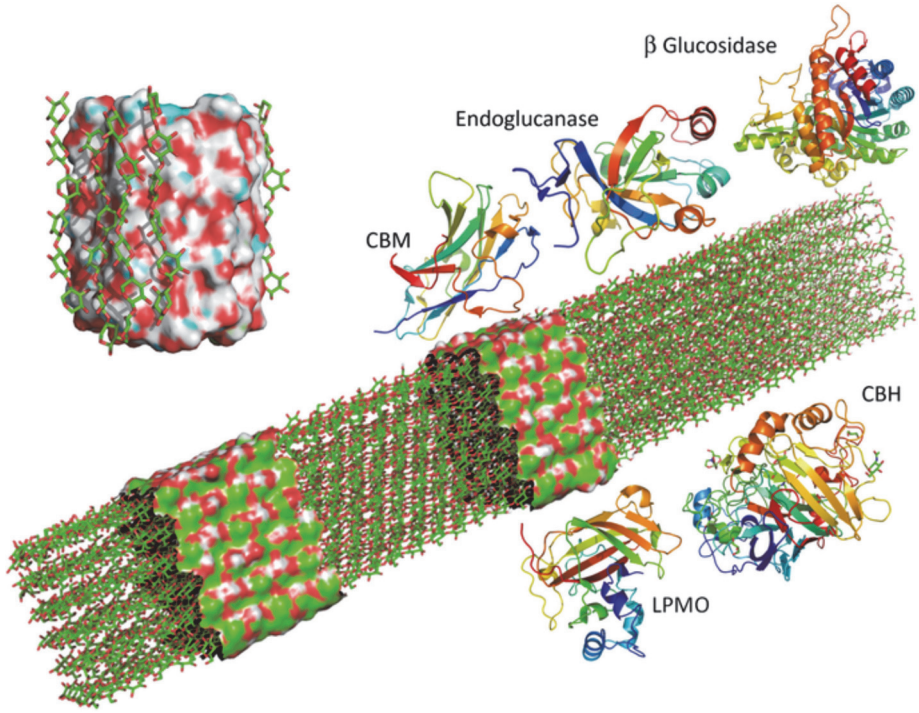

Fig. 12 Interaction and deconstruction of cellulose. The central part of the figure displays the three-dimensional structure and arrangement of 36 cellulose chains, constructed for an MD simulation surrounded by water molecules (not shown). Two sections of the microfibrils display the vander Waals surface to enlight the potential of interactions. The panel in the left-hand side part of the figure is a representation of the interactions occurring between the microfibril and xylan (one of the plant cell constituent) resulting from a molecular modeling investigation.[107] The rest of the figure displays the three-dimensional structures of the main categories of enzymes that digest crystalline cellulose to break it down into glucose: these are endoglucanase, cellobiohydrolase, β-glucosidase following the action of carbohydrate-binding module and lytic polysaccharidemonooxygenases (LPMO) that dramatically enhance the breakdown of cellulose. CBM PDB 4B96[121]; Endoglucanase PDB 5XBU[122] β-glucosidase PDB[123]; CBH PDB 505D[124]; LPMO PDB 6RW7.[125]

The recent years have witnessed the discovery and characterization of lytic polysaccharide monooxygenases (LPMO) that dramatically enhance the breakdown of cellulose.[128] The first QM/MM investigation of H_2O_2-dependent activity in family AA9 LPMO, showing that catalysis involves the formation of a caged hydroxyl radical and a Cu(II)-oxyl intermediate that oxidizes the C4 H bond of the polysaccharide substrate.[129–131]

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

In parallel of these investigations aiming to understand GH mechanisms and capturing complex reaction coordinates and conformational itinerary that substrates follow during the whole catalytic pathway, by Quantum chemical method, another direction of research explores the potential of Coarse-Grained method. Indeed the size of the full system to investigate requires the development and applications of such low resolution, coarse-grained representations.[132]

They are using MARTINI representation,[133] re-parametrized the cellobiose to make nonbonded interactions suitable to reproduce the partitioning free energies between water and cyclohexane for a series of cello-oligomers. By extrapolating the model to longer cello-oligomers, and by assigning particular cellulose-cellulose nonbonded interactions, a model was obtained which gives a stable, ordered structure in water that closely resembles the crystal structure of cellulose Iβ. The simulation of the motion of the carbohydrate-binding domain of the fungal cellulase Cel7A from *Trichoderma reesei* on a crystalline cellulose surface illustrates the potential of the method.

The representation of the cellulose surface layer as a two-dimensional grid is another way to capture the key events associated with the enzymatic degradation of cellulose at the mesoscopic level. The calculation includes free and bound states of both *endo*- and *exo*-cellulases with explicit reactive surface terms (e.g. hydrogen bond breaking, covalent bond cleavages) and corresponding reactions rates.[134]

6 Carbohydrate–protein interactions

As with other types of macromolecular interactions, favorable changes in enthalpy (ΔH) and entropy (ΔS) drive the formation of carbohydrate-protein complexes. Thermodynamic measurements have indicated that the binding free energy of monosaccharide to proteins is relatively small. ΔG increases in a significant manner whenever disaccharides or higher oligosaccharides are interacting with proteins. Whenever such proteins are interacting with carbohydrates, the multivalent effect generates high “avidity.”[135] The binding free energy between a carbohydrate molecule and a protein partner (ΔG) is indeed the variable of interest to be assessed. It is assumed to be composed of independent contributions in terms of van der Waals forces, electrostatic interactions with or without encompassing hydrogen bonding, the hydrophobic effect, etc.

Along with van der Waals and electronic interactions, a large number of hydrogen bond donors and acceptors present in carbohydrates, play a significant role in the formation of the complex with proteins. The complexity of such networks is enhanced by the competition occurring between the water molecules to form hydrogen bonds. Some entropic cost may counterbalance the overall enthalpic gain from hydrogen bonding. The occurrence of CH/π interactions characterize the enthalpy of binding of carbohydrates to protein. It is defined as a type of hydrogen bond occurring between a hydrogen atom attached to a carbon and the π systems of arenes. Typically, this effect is weaker. Despite the full recognition of this effect, its computation requires a high level of theory and most of the computational procedure do not take them into account.

The interactions between proteins and carbohydrates play a role in numerous biological processes such as protein specificity in antibody-antigen recognition, cell-cell adhesion, enzyme-substrate specificity, molecular transport, etc. (**Fig. 13**). They are essential to the onset, detection, and, potentially, also the prevention of human diseases such as cancer, inflammation, diabetes, neurodegenerative diseases, and bacterial and viral infections. The interactions between proteins and complex carbohydrates such as polysaccharides are also involved in the biosynthesis and biodegradation of the primary raw materials on Earth. Determination of the three-dimensional (3-D) structural and dynamical features of complex carbohydrates, carbohydrate polymers, and glycoconjugates, along with the understanding ofthe molecular basis of their associations and interactions represent a crucial challenge in structural glycoscience.[7,136]

Difficulties of co-crystallizing proteins and carbohydrates impede experimental assessment of the carbohydrate–protein interactions by X-ray crystallography. Nevertheless, highly resolved protein-carbohydrate complexes gathered from X-ray synchrotron investigations have accumulated to the point where it has been possible to compare the experimentally derived structures with those predicted from computational methods.[137] Some general features governing the protein–carbohydrate interactions are known, and computational tools have evolved and improved accordingly. These tools provide efficient

ways to increase our understanding of the different contributions to the binding energy. These developments allow searching the conformational space efficiently and yield reliable estimates of the binding free energy. They allow exploring in silico cases where the experimental data are lacking and provide sound structural information for a rational design of bioactive carbohydrates or carbohydrate mimetics.

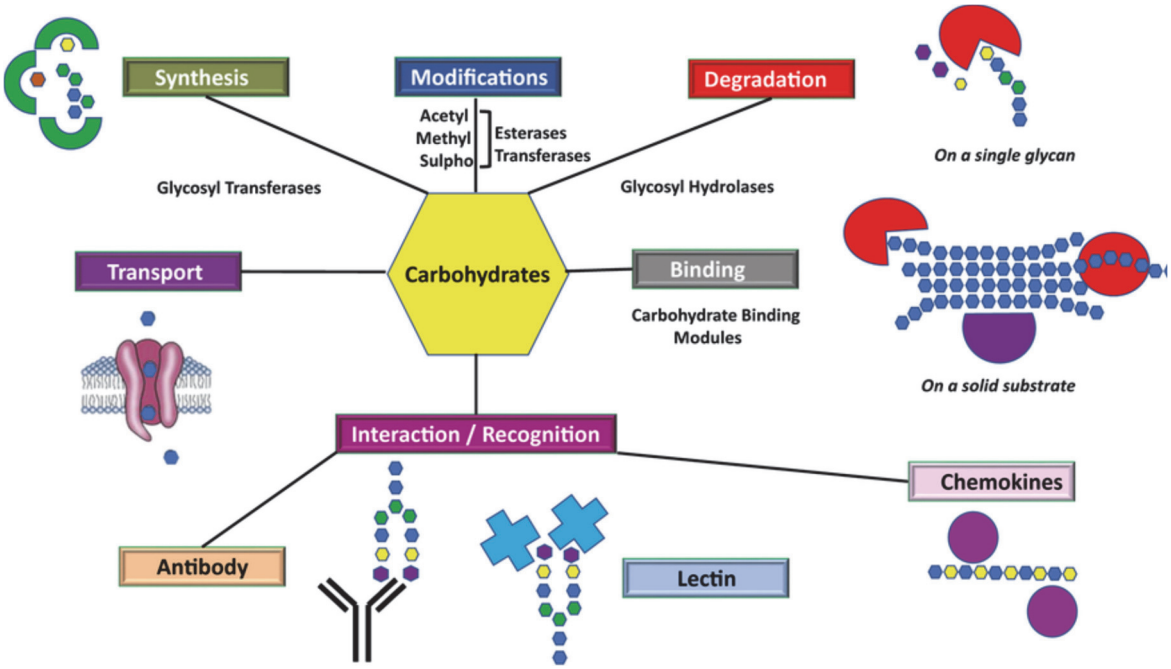

Fig. 13 Synopsis of the families of proteins interacting with carbohydrates along with their functions: transport; synthesis (glycosyl transferase) modifications (auxilliaries enzymes); degradation (glycosyl hydrolase, on single glycan and on semi-crystalline and crystalline glycan); carbohydrate binding modules; antibodies, lectins, and chemokines.

A myriad of glycan structures found in nature is derived from the enzymatic formation and the breakdown of glycosidic linkages achieved by carbohydrate processing enzymes, such as glycoside hydrolases and glycosyltransferases. Glycosylation proceeds in a stepwise manner, and therefore, the expression and specificity of the enzyme represent fundamental regulatory factors in defining the repertoire of biosynthesized glycans. The covalent addition of glycan to proteins and lipids represent not only the most abundant posttranslational modification but also by far the most structurally diverse. Structural changes in cell surface glycans accompany many physiological and pathological cellular processes. The functional significance of these changes is still not fully understood. It is, therefore, of high interest to elucidate the mechanisms utilized by these carbohydrate-acting enzymes.

6.1 Insights into enzymatic catalysis

Hybrid quantum mechanics and molecular mechanics (QM/MM) methods have become a powerful tool to give an accurate description of the catalytic events (hydrolysis or transfer) that occur either in Glycosidic Hydrolases (GH) or Glycosyltransferases (GT) (**Fig. 14**). The QM treatment of the electronic structure of an active site region and the rest of the enzyme by molecular mechanics, allow the modeling of enzymatic reactions, taking into consideration the impact of the surrounding.

Different reaction pathways of the enzymatic reaction mechanism can be captured along with transition-states. There are several programs available for calculations using QM/MM methods, as described in a review chapter that illustrates their applications to the area of GT and GH.[7] Another article describes how the modeling of the transition states and the subtle structural and electronic states follows the reaction coordinates in the case of GH.[131]

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

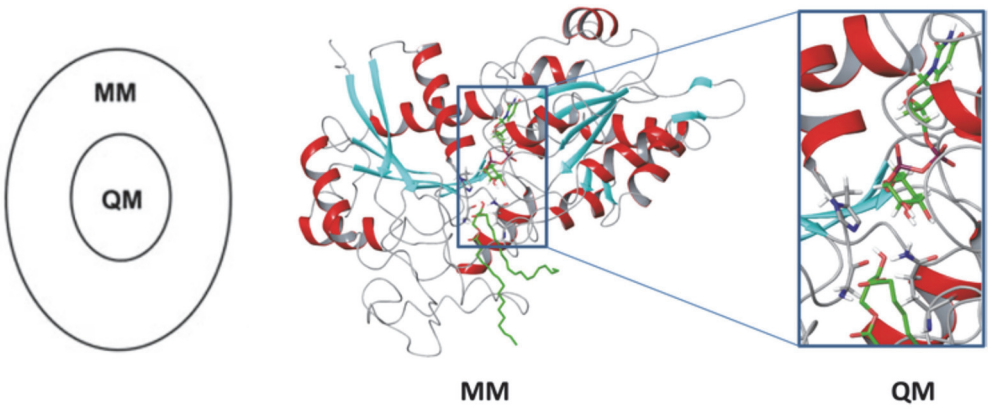

Fig. 14 Principles for the QM/MM method; (A) system is divided into QM and MM region; (B) the MM region of the overall structure of human OGT–UDP-GlcNAc–CKII-peptide complex is shown in ribbon representation and residues included in the QM region are shown in stick representation. Water molecules are not shown for clarity.

An investigation of the catalytic mechanism of carbohydrate acting enzymes using a QM/MMmethod is not a trivial task. The procedure requires three steps (1) preparation of a structural model (2) building of a QM/MM model and (3) mapping the enzymatic reaction. The last step uses a so-called stationary search in which a scan of one or two-dimensional search is performed along with selected reaction coordinates. Accordingly, hundred of energy minimizations are run, that eventually lead to transition states and other stationary points on the Potential Energy Surface.[138] Within the family of glycosyltransferase, the inverting O-GlcNAc glycosyltransferase (OGT) is a critical post-translation enzyme. It catalyzes the transfer of *N*-acetylglucosamine from UDP-*N*-acetylglucosamine (UDP-GlcNAc) to the hydroxyl group of the Ser/Thr of cytoplasmic, nuclear, and mitochondrial proteins. The authors of this study used the so-called string method for free energy estimation. The string method was able to optimize the reaction path efficiently for all three putative mechanisms studied in this work, and it was able to find the most probable mechanisms.[139]

6.2 Glycosyltransferase at work

Glycosyltransferases (GTs) are carbohydrate processing enzymes that transfer glycosyl residues from a donor to other molecules. Glycosyl donor substrates are mostly sugar nucleotides, such as UDP-GlcNAc, UDP-Gal, GDP-Man. However, lipid-linked sugars, e.g. dolichol phosphate saccharides and unsubstituted phosphates. Acceptor substrates are carbohydrates, proteins, lipids, DNA, antibiotics, or other small molecules. The chemistry of the catalytic reaction can be regarded as a nucleophilic displacement of the substituted phosphate leaving group e.g. UDP functional group at the anomeric carbon C1 of the transferred saccharide residue of a donor by a hydroxyl group of a specific acceptor. The formation of a new glycoside linkage during this reaction can mechanistically proceed with either inversion or retention of stereochemistry at the anomeric carbon C1 of the donor sugar. Thus, glycosyltransferases belong to either retaining or inverting enzymes, depending on the stereochemical outcome. Several detailed reviews of mechanistic and structural studies of glycosyltransferases were published, some summarize the outcome of computational investigations which usually ignore the physiological in which the reaction occurs.[140,141]

Interestingly, chloroplast offers a challenging example to investigate the enzymatic catalytic event in a complex organisation.[142] As a molecular machine, chloroplast converts the harvest photons into chemical energy. There is a unique spatial architecture of chloroplast that results from the presence and organization of two galactoglycerolipids whose content reaches 80% of overall lipid amount. Monogalactosyldiacylglycerol synthase (MGD1) performs the addition of one galactose from water-soluble donor substrate, UDP-α-D-galactose to hydrophobic acceptor substrate, diacylglycerol (DAG). The transfer proceeds with inversion of the anomeric configuration to the donor substrate and product of the reaction (**Fig. 15**).

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

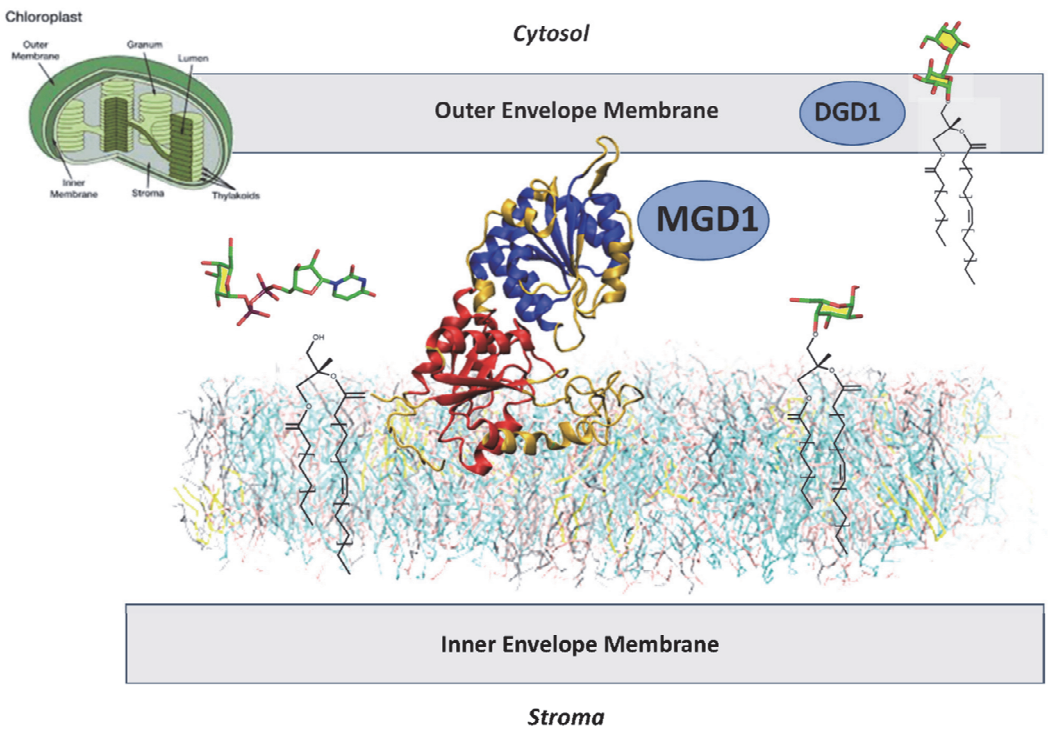

Fig. 15 Synopsis of the leading molecular players involved in the biosynthesis of galactoglycerolipids. Two monogalactosylglycerols (MGDG) and digalactosyldiacylglycerol (DGDG) galactoglycerolipids are the main lipids. The bulk of MGDG is synthesized in the inner Envelope Membrane (iEM) of chloroplast by a glycosyltransferase, monogalactosyldiacylglycerol synthase, MGD1. The conundrum of the MGDG synthesis arises from the complexity of the active machinery formation on the surface of iEM. For the catalysis, two substrates, hydrophilic sugar-bearing UDP-Galactose, and hydrophobic fatty acid tails bearing diacylglycerol, DAG, and at least one activator, anionic lipid molecule, phosphatidylglycerol, PG, are necessary to be bound together by MGD1. How MGD1 fishes the activator molecule and the DAG, which amounts to less than 1% of overall lipid content in the iEM, is still an open question.

Molecular dynamics simulations were performed at different levels of molecular representation, namely Coarse-Grained and All-Atom, to cover large spatial and temporal scales of the process of active complex assembly on the membrane surface. It revealed the compelling features of the interactions between MGD1 and the lipid bilayers and the lipid capture. The analysis of MD trajectories revealed the mutual influence of the membrane and the protein. The protein induces the lipid re-organization in the membrane. Such a reorganization facilitates the capture of both the substrate and the activator. At the same time, the membrane modulates the intrinsic dynamics of the protein, which is essential for the enzyme activity. The following self-assembly process can be drawn: (i) in a membrane without protein there occurs a spontaneous formation of rafts made of PG and DAG molecules. (ii) When MGD1 is embedded in the membrane, the protein interacts with the PG/DAG rafts with further accumulation of DAG. Such an accumulation of DAG results from lateral and transversal diffusion across the membrane. Further, All-Atom simulations of MGD1/lipid bilayer system were applied to elucidate the influence of membrane surface on protein intramolecular dynamics taking the hydrogen bonds explicitly. The results of the simulation shed light on the possible ways of the allosteric regulation of protein activity. Finally, from previously reported experimental data on MGD1 binding to iEM mimicking membranes,[143] the mutagenesis experiments suggested the crucial residues that might be responsible for the substrates and activators binding. The data derived from the accurate computer modeling provided the atomistic model of interactions in the active complex MGD1/UDP-Gl/DAG/PG.

6.3 Glycosyl hydrolases

Glycosidases or glycoside hydrolases (GH) hydrolyse glycosidic bonds in carbohydrates, polysaccharides, glycoproteins, glycolipids, etc. These enzymes are classified into ***endo***- and ***exo***-types. Exo-type glycosidases attack and hydrolyse mono-glycosides into free sugar and aglycon. When acting on oligo- or polysaccharides, they liberate a monosaccharide unit from the nonreducing end. Endo-type glycosidases act on oligo- and polysaccharides and catalyse the hydrolysis of an internal glycosidic

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

linkage thereby liberating two carbohydrate moieties, or in releasing an oligosaccharide (or polysaccharide) and mono-glycoside of the reducing end. Some glycosidases are capable of acting as both exo- and endo-types. The reactions resulting from the catalytic action of glycosidases is characterized by the anomeric configuration of the glycosidic bond of the substrate that the enzyme attacks, i.e. with retention or inversion of the anomeric configuration. Continuously published articles review the research on glycosidases and their mechanism.[144–147] One of the distinctive features in the catalytic mechanism of glycosidases is that the pyranose ring at subsite-1 occurs in a distorted conformation instead of the most stable $^4C_1$ conformation.

In the case of glycosyl hydrolases, of significant importance has been the discovery of the conformational changes that follow the sugar at a particular position in the enzymic site during the catalytic reaction; this is the so-called conformational catalytic itinerary. There does exist an interplay between protein and substrate conformational change, which may explain how an active enzymic site adapts dynamically to substrates of multiple sizes and multiple compositions.[148] A significant finding was the discovery of a processive catalysis by an *exo*-hydrolase with a pocket-shape active site.[149]

6.4. Insight into protein-carbohydrate recognition

When used in conformational studies of carbohydrates, computational molecular modeling methods offer alternatives for the study of protein–carbohydrate interactions. These simulations are probably the most powerful method we have nowadays to obtain atomistic-level information on the molecular recognition of highly dynamic systems, such as complex carbohydrates. The recognition pathway can follow two opposed processes described as “conformational selection” and “induced fit.”[150] Following the conformational selection theory, the receptor will bind selectively only the conformers that correspond to the final, bound conformation, which in practice defines an adequate concentration of the substrate. The induced-fit theory explains that recognition occurs regardless of the substrate’s specific 3-D conformation and that the substrate will fold in-place to match the spatial constraints of the receptor-binding site. Borrowing from the « intrinsically disordered proteins » field, an intermediate case scenario describes that the substrate can form local 3-D motifs, known as molecular recognition features (MoRFs),[151] that are recognized by the receptor. MoRFs act as nucleation sites initiating a folding in-place process Because of their highly dynamic architecture, carbohydrates are intrinsically disordered biomolecules.

Consequently, it is rather difficult if not impossible, to experimentally determine how they are recognized by lectins or by glycan-processing enzymes. Significant steps have been made, among which are the developments and implementations of force fields capable of accounting for the specificity of carbohydrates and their compatibility with those developed for proteins. The conformational flexibility of carbohydrates needs to be characterized and taken into account at each step of the investigation.

6.4.1 Protein-Induced conformational distortion of glycan

The resolution of the crystal structure of a β-propeller lectin from *Ralstonia solanacearum* (RSL) opened a very challenging case with the occurrence of a rare “open” conformation of the Lewis X (LeX) trisaccharide[152] (**Fig. 16**). It was the first observation of the occurrence of such conformation for the unbound LeX.[153,154] The crystallographic analysis of the open LeX complex[152] was accompanied by conventional (nonaccelerated) MD simulations of two LeX-RSL complexes, of 1 and 0.85 μs production, and by the umbrella sampling analysis[155] of the binding and unbinding pathways.

Additionally, the conformational equilibrium of the LeX unbound (free) in solution was studied through a set of 30, 1 μs trajectories run in parallel, started from uncorrelated conformations.[152] The MD results confirmed the stability of the LeX open conformation in the complex and highlighted the specific interactions with the binding site residues that stabilize it. Additionally, the MD simulations of the free LeX were extensive enough to capture opening events, thus to define a potential opening pathway that involves a concerted conformational change of the two glycosidic linkages and the Glc*p*NAc ring pucker, and also confirmed that the LeX structure opening does indeed happen in solution, but is too much of a rare occurrence to be captured by NMR. In summary, the extensive MD work[152] suggested that RSL binds LeX through an induced-fit mechanism, where the nature of the interactions and architecture of the binding site

compensate for the enthalpic cost of opening, namely 10.6 kJ/mol from sampling. Within this context, a notable contribution[156] showshow their enhanced MD scheme, namely a multidimensional variant of the swarm-enhanced sampling MD (msesMD) method, compares to conventional (unbiased) MD sampling, i.e. 3, 10 μs trajectories, to umbrella sampling and to accelerated MD[157] for characterizing the LeX/A and sialyl-LeX/A (sLeX/A) conformational spaces exhaustively.

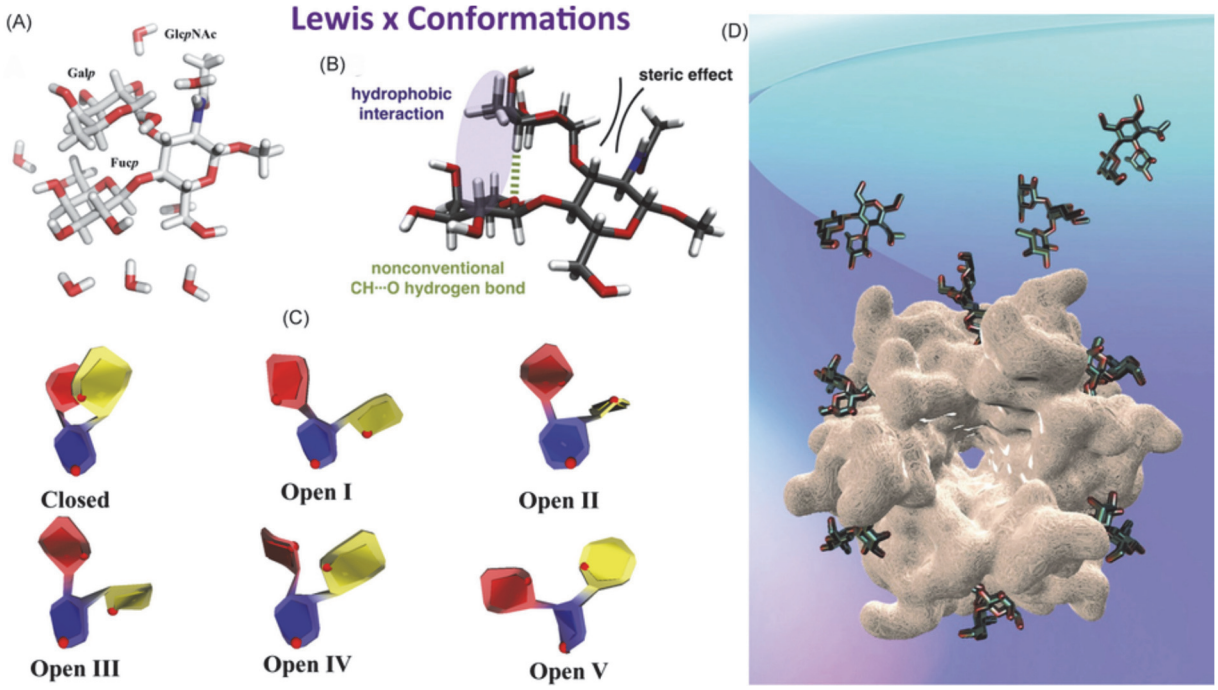

Fig. 16. The hidden conformations of LewisX: β-D Gal*p* (1-4) (α-L Fuc*p* (1-3) β-D Glc*p*NAc). The conformations of fucosylated Lewis oligosaccharides are considered to be rigid, adopting a single shape referred to as the "closed" conformation (A). This rigid shape is due to stacking between fucose (Fuc*p*) and galactose (Gal*p*) rings, by a nonconventional CH・・・O hydrogen bond and by steric hindrance of the *N*-acetyl group of Glc*p*NAc (B). The crystal structure of Lewis x (LeX) trisaccharide, together with NMR and modeling data, confirmed that the trisaccharide presents only limited conformational fluctuations around the closed shape. When bound to the *Ralstonia solanacearum* lectin LeX core adopts several very different conformations (referred to as Open) shown in (C). Extensive molecular dynamics simulations confirm rare transient LeX openings in solution, frequently assisted by distortion of the central *N*-acetyl-glucosamine ring. Additional directed moleculardynamic trajectories (D) revealed the role of a conserved tryptophan residue in guiding the fucose into the binding site[152] drawing Sweet Unity Mol.[177]

This study shows that enhanced sampling schemes provide the same structural, dynamics and energetic insight obtained from the microsecond-long multiple trajectories, at a fraction of the time while exhaustive sampling can be reached in the case of tri- and tetrasaccharides, even if it requires high-energy transitions, it becomes quickly unachievable as the number of atoms and the complexity of the system increases.

6.4.2 Docking carbohydrates on protein

Protein–carbohydrate docking has come of age; producing reliable and insightful results.[158] The question of choosing the appropriate software concerning the problem to be investigated still stands, and remains critical for the proposed solution. This is particularly true for cases of small ligands in large and poorly defined binding sites. Docking is a computational method that places a small molecule(ligand) in the combining site of its macromolecular target (receptor) and provides an estimate of the binding affinity. Molecular docking requires (at least some) 3 dimensional knowledge of the ligand and the receptor of interest. The carbohydrate ligands are typically built by using molecular mechanics methods or directly sourced from structural databases. Energy parameters suitable or energy minimization and/or molecular dynamics of protein-carbohydrate complexes are available or different force fields.[159] Receptors structures are currently obtained from X-ray crystallography and NMR spectroscopy; those that are unavailable can be generated by homology modeling, threading, and de novo methods. Even though several docking programs that operate in slightly different ways are available,

they all involve two main features, that is sampling and scoring. Sampling entails the conformational and orientational location of the ligand in the receptor-binding site. To predict the carbohydrate orientation in binding sites, flexible docking methods are used to account for possible orientations of pendent groups (i.e. hydrogen bond network directed by the orientation of hydroxyl and hydroxymethyl groups) and the conformational flexibility occurring at each glycosidic linkage. In most docking programs, the ligand is treated as flexible, whereas the protein conformation is often kept rigid. Programs exist that can carry such "soft docking." Proper accounting for receptor flexibility is computationally much more expensive, and it is not yet common practice. The docking algorithms can be grouped into deterministic approaches that provide reproducibility and stochastic approaches in which the algorithm includes random factors that do not allow for full reproducibility. Incremental construction algorithms consist of the division of a ligand in rigid fragments, as implemented in program DOCK.[159,160] One of the fragments is selected and placed in the protein binding site. The reconstruction of the ligand is performed in situ, adding the remaining fragments. Among the stochastic searching approaches, the genetic algorithm (inspired by evolutionary biology) is implemented in AutoDock.[160] A variety of other sampling methods are used in docking programs. Some of them include simulated annealing protocols and Monte Carlo simulations. The algorithm used in Glide[161] can be defined as a hierarchical algorithm. Scoring functions are used to evaluate the best conformation, orientation, and translation (referred to as poses), which classify the ligands in rank order. Energy scoring functions evaluate the free energy of binding between proteins and ligands, using the Gibbs–Helmholtz equation that describes ligand-receptor affinity. Empirical scoring functions use a set of parameterized terms describing properties known to be decisive in protein-ligand binding to formulate an equation for predicting affinities. These terms generally describe polar–apolar interactions, loss of ligand flexibility, and desolvation effects.

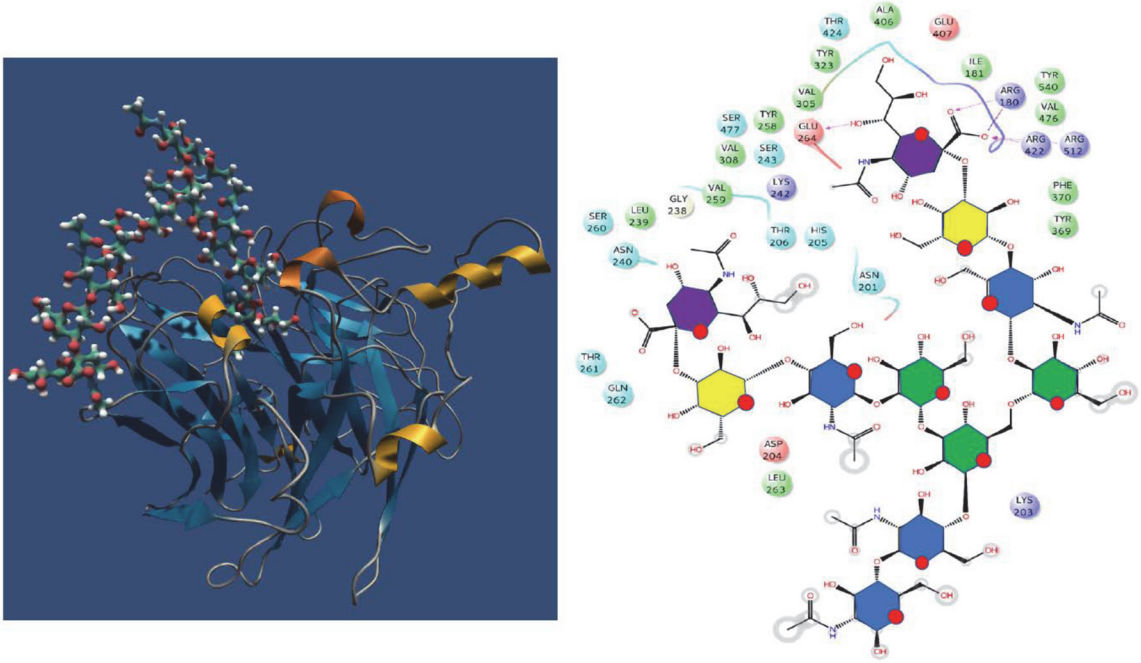

Fig. 17. Structural features of the glycan receptor binding by mumps virus hemagglutinin-neuraminodase. 3-D representation of the undecasaccharide in interactions with the receptor; and a two-dimensional plot illustrates the interaction of the sialylated undecasaccharide with the residues in the binding pocket. Drawn from structural data kindly provided by the authors.[164]

A distinct feature of protein-carbohydrate recognition is the interaction between aromatic side chains of the proteins and C H bonds of the carbohydrate's hydrophobic faces, which results in the formation of crucial CH–π contacts.[162] Widely used docking programs, which account differently for these types of interactions, may not perform as well for the protein-carbohydrate complex. Various docking programs and scoring functions perform differently for different targets, and that varying performance might occur for different ligand types. An accurate determination of carbohydrate–protein complexes

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

remains a non trivial matter. In the case of protein–lectin complexes, the difficulties arise from the shallow and multichambered binding sites of many lectins. Comprehensive studies on the validation of carbohydrate–lectin docking have been performed and compared to experimentally crystallographically determined complexes. In comparison with a large number of docking studies performed on carbohydrate–lectin complexes, there are relatively few published docking studies on carbohydrate–antibody recognition, reflecting the limited number of suitable validation tests (i.e. high-resolution carbohydrate–antibody crystal structure complexes) and the inherent difficulty in modeling such systems. Despite such difficulties arising from the challenges of protein-carbohydrate complexes, molecular docking has started producing reliable and insightful results. However, many challenges remain, and it is still a nontrivial exercise to perform and far from being a turn-key tool. In particular, the ability of docking programs to correctly score docking poses (especially in the cases of small ligands in large and poorly defined binding sites) calls for critical inspection of the results.

The full characterization of glycan recognition by proteins requires the quantitative measurements of the strength and specificity of the interactions, which can be assessed through the combined use of biophysical methods and computer simulation. NMR methods, in conjunction with docking simulation, have dealt with several classes of interactions at the recognition sites (see review[163]). The elucidation of the structural basis for glycan receptor bindings by mumps virus hemagglutinin-neuraminidase illustrates.[164] By combined use of NMR, docking, molecular modeling and CORCEMA-ST, the structural features of sialoglycans/MuV-HN complexes were revealed. Evidence for a different enzyme activity towards longer and complex substrates compared to unbranched ligands was also examined by an accurate NMR kinetic analysis (**Fig. 17**).

6.5. Insights into carbohydrate transport

Carbohydrates such as glucose, lactose, sucrose, raffinose, malto-, fructooligosaccharides, l-fucose, trehalose, oligo-alginate, oligo galacturonate and others, constitute a source of carbon for many organisms. These molecules need to travel across the channel and pores, and their motion is critically important for the proper functioning of many cellular processes. At the protein level, a family of proteins, collectively referred to as membrane transporters, belong to the Major Facilitator Superfamily (MSF), achieve such a role. These trans-membrane proteins allow permeation of carbohydrates. Their structures, along with the elucidation of the mechanistic transport model, are the subject of intense research. Some high-resolution structural elucidation of transporters has enabled investigation into the MD of fundamental transport processes. Glucose transporters belong to one of the largest family of the membrane transporters, which are present in all the kingdom of life. They provide the pathway to transport glucose (and other mono and disaccharides) across the membrane.[165,166]

Long MD simulations provide a way to investigate the mechanism of the conformational transition of the human glucose transporter in the absence and the presence of glucose. Many features result from these endeavors. Some characterize the behavior of the glucose in the central cavity of the protein. Once entering the protein cavity, the glucose interacts with many residues by forming multiple hydrogen bonds, which contribute to a favorable enthalpic interaction. Meanwhile, the glucose undergoes numerous rotations and axial movements, which might result in a favorable entropic contribution. Overall, the free energy of glucose binding in the protein cavity might be lower compared to other ligand positions and slow down the process of the glucose transfer to the intracellular medium.

Like the general diffusion transport mechanism of porin, the motions of maltotriose in the maltoporin membrane channel involve a series of continuous conformational changes.[167] There is a continuous stretch of aromatic residues in the channel arranged a left-handed helical path that forms a “greasy slide.” The first event is the binding of sugar to the first residue of the “greasy slide” which occurs via van der Waals interactions to the hydrophobic face of the glucosyl ring. Deeper penetration into the channel occurs throughout guided diffusion of the oligosaccharide along the “greasy slide.” Gradual dehydration of the malto-oligosaccharide favors the establishment of short hydrogen bonding interactions between the sugars’ hydroxyl groups and the surrounding amino-acids. This is due to the conformational flexibility at the glycosidic linkages and the primary hydroxyl groups. The presence of the charged side chains (referred to as

"polar tracks") mimics the first hydration shell to the sugar by providing hydrogen bonds to its hydroxyl groups. The polar tracks are divided into donor and acceptor lanes all along the greasy slide. The motion of the glucose residues to the next binding site of the ≪ greasy slide ≫ occurs in combination with a rearrangement of hydrogen bonds. Such an arrangement is referred to as the "register shift." The continuous making and breaking of hydrogen bonds induce the oligosaccharide motion through the porin in a capillary-like fashion (**Fig. 18**).

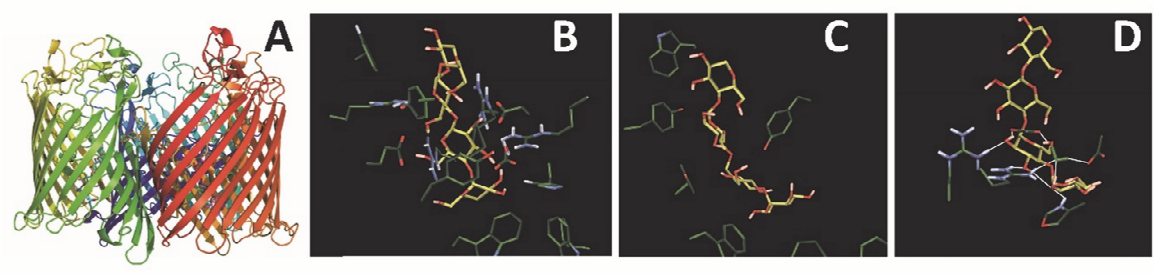

Fig. 18. Three-dimensional structure of maltoporin(A) along with snapshots of the interaction of malto-oligosaccharides within the channel (from (B) to (D)). The first event is the binding of sugar to the first residue of the "greasy slide" which occurs via van der Waals interactions to the hydrophobic face of the glucosyl ring. Deeper penetration into the channel occurs throughout guided diffusion of the oligosaccharide along the"greasy slide." Gradual dehydration of the malto-oligosaccharide favors the establishment of short hydrogen bonding interactions between thesugars' hydroxyl groups and the surrounding amino-acids. This is due to the conformational flexibility at the glycosidic linkages and the primaryhydroxyl groups. The presence of the charged side chains (referred to as "polar tracks") mimics the first hydration shell to the sugar by providing hydrogen bonds to its hydroxyl groups. The polar tracks are divided into donor and acceptor lanes all along the greasy slide. The motion of the glucose residues to the next binding site of the greasy slide occurs in combination with a rearrangement of hydrogen bonds. Such an arrangement is referred to as the "register shift."

6.6 Glycosaminoglycans: The GAGs

The molecular modeling of the structure, dynamic and interactions of the GAGs is the concatenation of most of the difficulties in glycoscience as they combine the challenges of both glycans and polyelectrolyte polysaccharides. The GAGs comprise a family of complex anionic polysaccharides including (1) glucosaminoglycans (heparin and heparan sulfate), (2) galactosylaminoglycans (chondroitin sulfate and dermatan sulfate), and (3) hyaluronic acid and keratan sulfate. GAGs display a variety of sulfation patterns which contribute to the range of their sequence and conformations, which impact on their molecular recognition and biological activities. Their size and heterogeneity necessitate multiscale modeling of glycosaminoglycans from disaccharide fragments to polysaccharide.[114] The characterization of optimized conformation of a heparin disaccharide by Quantum methods, the application of molecular modeling to hyaluronan decasaccharide in water and the multi-microsecond aqueous simulation of heparan and proteoglycans and heterogeneous glycosaminoglycans [168] provide some illustrations of the Spatio-temporal achievements. The lower resolution Coarse-Grained simulations of larger molecules and on longer timescale is particularly suited for investigating the dynamics of GAGs.[112]

In addition to their participation in the physicochemical properties of the extracellular matrix, GAG fragments are recognized explicitly by protein receptors.[169] Being part of proteoglycans or in case of HA not linked covalently to any protein, GAGs interact with their protein targets in the Extracellular Matrix and at the cell surface such growth factors and chemokines. They play a role in the regulation of many processes, such as hemostasis, growth factor control, anticoagulation, and cell adhesion. As such, GAGs are very promising targets for the development of novel bioactive molecules of therapeutic value as well as for the design of innovative functional materials to control and promote the processes of application in the field of bones and skin regeneration. More than 20 protein-GAGs systems were characterized over the last 15 years (**Fig. 19**).

The development of new force fields,[171] scoring functions, databases contributed to the understanding of GAGs in conjunction with data resulting from experimental techniques. These proved to be relevant and useful for the detailed characterization of structure-functions and structure-properties relationships.[172] A flow chart describing the use of computational approaches to be addressed key questions on GAG–protein interactions has been established to understand such fundamental questions as: (i) Does my protein bind to GAGs; (ii) Where does the GAG bind; (iii) What is the most optimal GAG sequence; (iv) Is my GAG-protein complex stable?[173]

 https://doi.org/10.1016/B978-0-12-819475-1.00004-3

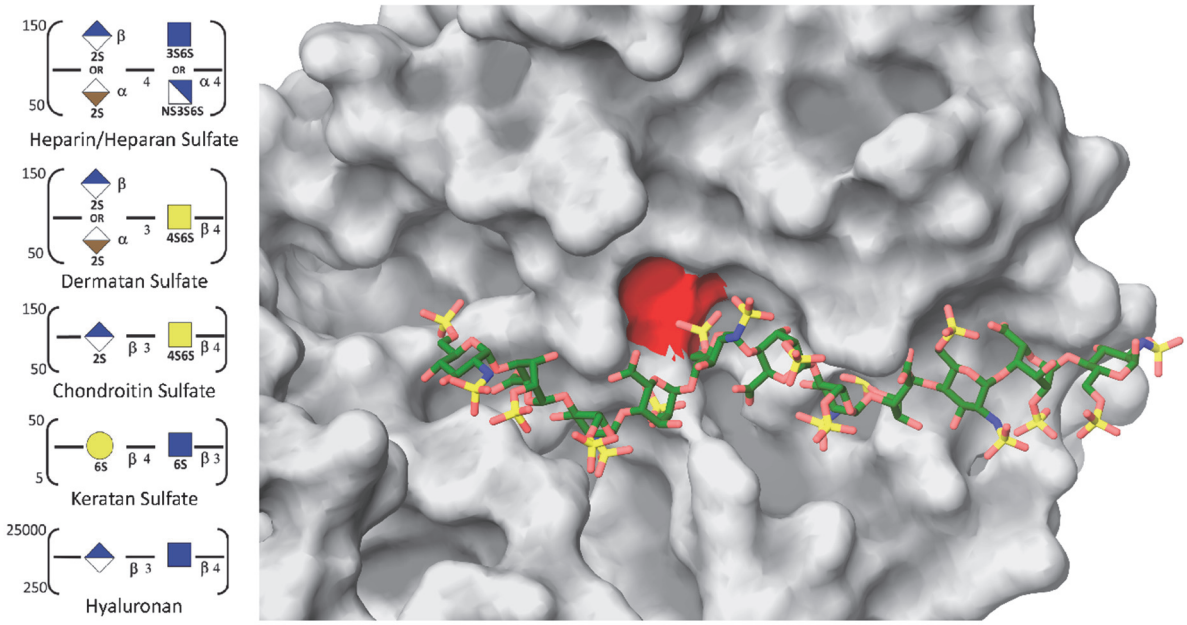

Fig. 19. Schematic representation of the repeating disaccharide segments found in GAGs which at the exception of Hyaluronic Acid display a variety of sulfation patterns. A model structure of the catalytic domain of human sulfatase-2, an extracellular endosulfatase editing the pattern of 6-O-sulfation in Heparan sulfate, with a short fragment of its ligand, GlcNS(6*S*)-[IdoA(2*S*)-GlcNS(6S)]5. Red spot indicates the position ofcatalytic cysteine modified FGly residue. The 3-D structure was constructed on the base of homology modeling, the position of GAG chain wasderived from the molecular docking procedure.[170]

Nevertheless, many challenges in computational modeling remain: (i) to investigate the role of solvation and the contribution of the polyelectrolyte nature of GAGs. (ii) To characterize GAGs chains having a degree of polymerization greater than ten, which display a high degree of conformational flexibility. New docking protocols need to be developed, such as Coarse-Grained techniques and fragment-based approach for docking. (iii) To elucidate the specificity of their interactions with proteins and the role of GAG length and sulfation patterns. (iv) To investigate how GAG molecules induce allosteric effects on their target proteins. (v) To assess the thermodynamics and kinetics features of protein-GAGs systems. (vi) To decipher how GAGs–protein interactions take place in the context of the multicomponent nature of the *peri* and extracellular matrix.

7 Structural glycobioinformatics

The upcoming scientific production in Glycoscience amounts to the yearly publication of about 70,000 research articles. Amplified by the availability of sophisticated and powerful high-performance computing and searching capacities, the space of accessible information has substantially increased such that data mining opens new prospects of discovery. This new view not only emphasizes the worth of analyzing raw data from published work but also points at the untapped wealth that may be harvested in collected data sets from which extracted information can translate into knowledge.

The field of structural glycoscience has partially benefited from such advances with the development of tools and databases for structural analysis of glycan and polysaccharides, and their interactions. A variety of online resources mainly in the form of databases covering glycan and glycoproteins structures, enzymes responsible for their biosynthesis and degradation, glycan-binding to human pathogens, glyco-epitope and their antibodies, etc. has been developed by independent research groups worldwide.

The accumulation of information, resulting from the development of enabling technologies, has laid the foundation of a rich computational toolbox tailored for the detection and high-resolution determination of complex glycans. In parallel, a variety of online resources in the form of tools and databases covering glycan and glycoproteins structures have been developed by independent research groups worldwide. At present, more than 150 entries are freely available on the internet, yet these often produced independently of one another.[174]

Parallel with the development of methods in molecular modeling, there has been a revival in the

number and scope of databases, websites, and both real and virtual glycan libraries that address the information needs in glycosciences. These efforts intend to (1) assess“primary data” (covalent and 3-D structures of glycans and glycoconjugates) and (2) organize these primary data into databases, which can be used for (a) speeding up the production of primary data, (b) predicting new features, and (c) characterizing structure-activity or structure-function relationships, throughout their integration into metadatabases.

- Glyco3D: A portal for structural glycosciences[175] includes a family of databases covering the 3-D features of monosaccharides, disaccharides, oligosaccharides, polysaccharides, lectins, glycosyltransferases, monoclonal antibodies and glycosaminoglycan binding proteins that have been developed with nonproprietary software and are freely available to the scientific community (http://glyco3d.cermav.cnrs.fr).

- POLYS-GLYCAN BUILDER: An intuitive application to build 3-D structures of polysaccharides (algae, bacteria, GAG, plants). (http://glyco3d.cermav.cnrs.fr/builder.php)

- GFDB (http://www.glycanstructure.org): A glycan fragment database: a database of PDB-based glycan 3-D structures.

- GLYCAM (http://glycam.org): Primary sequences for some common glycans have been pre-built, and their predicted 3-D structures are available. High Mannose, Hybrid N-Glycan, Complex Type, Sialyl/Fucose Complexes.

- GlycoMapsDB (http://www.glycosciences.de/modeling/glycomapsdb/): A database containing more than 2500 calculated conformational maps for a variety of di- to pentasaccharide fragment.

- CHARMM GUI (http://charmm-gui.org/input/glycan).

- EPS Database The EPS Database provides access to detailed structural (1D–3D) taxonomic and bibliographic information on bacterial EPS. (http://www.epsdatabase.com/):

- POLYSAC3DB PolySac3DB is an annotated database that contains the 3D structural information and original fiber diffraction data of 157 polysaccharide entries that result from an extensive screening of scientific literature. (http://www.polysac3db.cermav.cnrs.fr/):

- MATRIX-DB MatrixDB is a biological database focused on molecular interactions between extracellular proteins and polysaccharides. It contains protein-protein interactions (PPIs) and also protein–glycosaminoglycan interactions and 3D structures of GlycosAminoGlycans. (http://www.matrixdb.univ-lyon1.fr/):

- UniLectin3D: UniLectin platform is a dedicated portal of databases and tools to study the lectins. (https://unilectin.eu/)

8 Conclusions

Since the publication of the first edition of Comprehensive Glycoscience, there has been a paramount increase in the development and applications of computational methods aimed at establishing the structural and dynamic features of complex carbohydrates, either in isolation or complexed to other bio-molecules. Thanks to quantum chemical methods, the developments and implementations of force fields capable of taking into account the specificity of carbohydrates (stereo-electronic effect, gauche effect, etc.) constitute significant land-marks. The integration of these force field in several “generic” software provides many users with a comprehensive way to carry on computational explorations of their endeavor in associations with their experiments. To a certain extent, molecular simulation techniques have reached the sampling power and a sufficient level of sophistication to provide not only the missing structural insight necessary to interpret or support experiments but also to be a primary tool of scientific discovery. Complementary to such developments, advancements in high-performance computing have allowed molecular simulation methods not only to play a more substantial role in supporting experiments; but to transcend such mandate to guide experimental design and to lead autonomously scientific discovery. Within an affordable time of computation, new dimensions, both spatial and temporal, can be assessed. For example, atomistic MD simulations provide unique insight, giving an accurate description of the 3-D structure and real dynamics motifs at the actual timescale when molecular events take place. Not only structure-function relationships

can be proposed in some instances, but the characterization of physicochemical and mechanical properties open the road to establishing new structure-properties relationships. The development of Coarse-Grained simulations provides such applications to most systems of glycoscience, as oligosaccharides, polysaccharides, plant cell-wall polysaccharides, chitin and chitosan, glycosaminoglycans, glycoconjugates, N- and O-linked glycans, glycolipids and lipopolysaccharide membranes. Because of the reduced complexity inherent to the model, Coarse-Grained computer simulations have the potential to bridge the gap between experiment and all-atom calculations for systems of high complexity, more specifically they are well-suited to study the dynamic nano-clustering of glycolipids in which the carbohydrate–carbohydrate interactions, the formation of ≪ rafts ≫ play a crucial role. One may envisage that such studies, will help to decipher the intriguing multivalency effect that governs essential protein–carbohydrate interaction. Indeed, since most carbohydrate-binding proteins, particularly lectins and adhesins display a rather low affinity and generally narrow carbohydrate recognition domains involving fewer than a tetrasaccharide residue, their intrinsic specificities often reside in their valence together with their various topologies. The simultaneous presentation of several proper and identical glycoside units converts relatively weak interactions into specific recognition effects. Therefore, one needs to consider some physicochemical principles that underline such associations, for example, patches of glycolipids and glyco-surfaces, in other terms the ≪ glyco landscape ≫ or ≪ glycotope ≫. This concept may apply to the field of research dealing with the solid-state degradation of crystalline or semi-crystalline polysaccharides by enzymes. For the time being, the computational exploration of such systems is far from being complete, despite the significant fundamental contribution that computer simulations brought to the understanding of the chemistry underlying the mechanism of Glycosyl Hydrolases. They allowed for the identification of the catalytic residues, the validation of complex conformational itineraries while capturing mechanistic details that escape experimental probes. This field of research is profiting from the wealth of crystal structures of proteins, along with their carbohydrate complexes. The derived tools and constituting databases for structural analysis of glycan and polysaccharides and their interactions provide an untapped wealth of sometimes unexplored data sets from which new knowledge may result. The structural data gathered by X-ray crystallography, offer the opportunity to organize well-focused databases using PDB information, with an appropriate curation of the glycan topology. The development of high-quality glycomics databases counteracts the lack of precision reflected in the abundance of unreviewed and incorrect information regarding both glycoconjugates and glycan-binding proteins in genome and protein databases, and yet, such curated database offers a high predictive power.[176]

In contrast, the field of molecular modeling, despite its massive utilization of computer resources, does not have a place where any deposited and documented results could be stored and made available publicly. It is, therefore, tempting to suggest the creation and the organization of such a repository of 3-D data. Ideally, the data would correspond to the highest populated conformers identified through the simulation analysis with detailed information on the relative populations and energetics. They would constitute a new structural database. Besides offering the possibility to reproduce published results, such a database would contribute to the wealth of publicly available resources in glycoscience. Its integration into metadatabases would complement the set of both experimental and computational data, which with application based on machine learning will allow the rapid advancement of glycoscience and its contribution to the understanding of the many processes and architectures involving these complex biomolecules.

References

1.. S. Perez, C. Gautier, A. Imberty, in: B. Ernst, G. Hart, P. Sinay (Eds.), Oligosaccharides in Chemistry and Biology: A Comprehensive Handbook, Weinheim, Wiley/VCH, 2000, pp. 969–1001.

2.. S. Perez, Comprehensive Glycoscience, 2.1: Analysis of Glycans, 2007193–220.

3.. T. Lutteke, A. Bohne-Lang, A. Loss, T. Goetz, M. Frank, C.W. von der Lieth, Glycobiology 16 (5) (2006) 71R–81R.

4.. A. Varki, R.D. Cummings, M. Aebi, N.H. Packer, P.H. Seeberger, J.D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, J.J. Prestegard, R.L. Schnaar, H.H. Freeze, J.D. Marth, C.R. Bertozzi, M.E. Etzler, M. Frank, F.G. Vliegenthart, T. Lutteke, S. Perez, E. Bolton, P. Rudd, J.

Paulson, M. Kanehisa, P. Toukach, K.F. Aoki-Kinoshita, A. Dell, H. Narimatsu, W. York, N. Taniguchi, S. Kornfeld, Glycobiology 25 (2015) 1323–1324.
5.. S. Perez, Glycopedia, Available at:. https://glycopedia.eu/e-chapters/symbolic-representation-of-monosaccharides-in-the-age-of-glycobiology/article/abstract, 2014.
6.. D.B. Werz, R. Ranzinger, S. Herget, A. Adibekian, C.-W. von der Lieth, P.H. Seeberger, ACS Chem. Biol. 2 (10) (2007) 685–691.
7.. S. Perez, I. Tvaroska, Adv. Carbohydr. Chem. Biochem. 71 (2014) 9–136.
8.. O.C. Grant, R.J. Woods, Curr. Opin. Struct. Biol. 28 (2014) 47–55.
9.. R. Buckingham, Proc. R. Soc. A. 168 (933) (1938) 264–283.
10.. M.P. Allen, D.J. Tildesley, Computer Simulations of Liquids, Oxford University Press, Oxford, UK, 1987.
11.. K.N. Kirschner, A.B. Yongye, S.M. Tschampel, J. Gonzalez-Outeirino, C.R. Daniels, B.L. Foley, R.J. Woods, J. Comput. Chem. 29 (4) (2008) 622–655.
12.. O. Guvench, S. Mallajosyula, E. Raman, E. Hatcher, K. Vanommeslaeghe, T. Foster, F. Jamison, A. MacKerell, J. Chem. Theory Comput. 7 (10) (2011)3162–3180.
13.. E. Fadda, R.J. Woods, Drug Discov. Today 15 (15–16) (2010) 596–609.
14.. K. Nester, K. Gaweda, W. Plazinski, J. Chem. Theory Comput. 15 (2) (2019) 1168–1186.
15.. L. Pol-Fachin, V. Rusu, H. Verli, R. Lins, J. Chem. Theory Comput. 8 (11) (2012) 4681–4690.
16.. L. Pol-Fachin, H. Verli, R. Lins, J. Comput. Chem. 35 (29) (2014) 2087–2095.
17.. S.S. Mallajosyula, S. Jo, W. Im, A.D. MacKerell, Methods Mol. Biol. 1273 (2015) 407–429.
18.. A. Aytenfisu, M. Yang, A. MacKerell, J. Chem. Theory Comput. 14 (6) (2018) 3132–3143.
19.. P. Pandey, A. Aytenfisu, A. MacKerell, S. Mallajosyula, J. Chem. Theory Comput. 15 (9) (2019) 4982–5000.
20.. R.H. Marchessault, S. Perez, Biopolymers 18 (1979) 2369–2374.
21.. H.B. Mayes, L.J. Broadbelt, G.T. Beckham, J. Am. Chem. Soc. 136 (2014) 1008–1022.
22.. V. Spiwok, B. Kralova, I. Tvaroška, Carbohydr. Polym. 345 (4) (2010) 530–537.
23.. E. Braun, J. Gilmer, H.B. Mayes, D.L. Mobley, J.I. Monroe, S. Prasad, D.M. Zuckerman, Best practices for foundations in molecular simulations [Article v1.0], Living J. Comput. Mol. Sci. 1 (1) (2018) 5957.
24.. J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, J. Comput. Phys. 23 (3) (1977) 327–341.
25.. B. Hess, H. Bekker, H. Berendsen, J. Fraaije, J. Comput. Chem. 18 (12) (1997) 1463–1472.
26.. Y. Sugita, Y. Okamoto, Chem. Phys. Lett 314 (1–2) (1999) 141–151.
27.. L. Wang, R. Friesner, B. Berne, J. Phys. Chem. B. 115 (30) (2011) 9431–9438.
28.. K. Burusco, N. Bruce, I. Alibay, R. Bryce, ChemPhysChem 16 (15) (2015) 3233–3241.
29.. A. Laio, M. Parrinello, Proc. Natl. Acad. Sci. U. S. A. 99 (20) (2002) 12562–12566.
30.. I. Ohmura, G. Morimoto, Y. Ohno, A. Hasegawa, M. Taiji, Philos. Trans. A Math. Phys. Eng. Sci. 372 (2021) (2014) 20130387.
31.. D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, Bank, J. A, J.M. Jumper, J.K. Salmon, Y. Shan, W. Wriggers,Science 330 (6002) (2010) 341–346.
32.. J. Jung, A. Naurse, C. Kobayashi, Y. Sugita, J. Chem. Theory Comput. 12 (10) (2016) 4947–4958.
33.. Lenovo, The Convergence of HPC, AI, Why it's Great for Supercomputing and the Enterprise, In Inside HPC, 2009. https://insidehpc.com/2019/06/ hpc-and-ai-convergence-supercomputing/.
34.. S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. Dawid, A. Kolinski, Chem. Rev. 116 (14) (2016) 7898–7936.
35.. F. Muller-Plathe, ChemPhysChem 3 (9) (2002) 754–769.
36.. S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, A.H. de Vries, J. Phys. Chem. B. 111 (27) (2007) 7812–7824.
37.. T.A. Wassenaar, H.I. Ingolfsson, M. Priess, S.J. Marrink, L.V. Schafer, J. Phys. Chem. B. 117 (13) (2013) 3516–3530.
38.. P. Liu, S. Izvekov, G.A. Voth, J. Phys. Chem. B. 111 (39) (2007) 11566–11575.
39.. J. Gu, F. Bai, H. Li, X. Wang, Int. J. Mol. Sci. 13 (11) (2012) 14451–14469.
40.. Z. Yu, D. Lau, J. Mol. Model. 21 (5) (2015) 128.
41.. C. Scherer, D. Andrienko, Phys. Chem. Chem. Phys. 20 (34) (2018) 22387–22394.
42.. H.I. Ingolfsson, C.A. Lopez, J.J. Uusitalo, D.H. de Jong, S.M. Gopal, X. Periole, S.J. Marrink, Wiley Interdiscip. Rev. Comput. Mol. Sci. 4 (3) (2014) 225–248.
43.. P. Kar, M. Feig, Adv. Protein Chem. Struct. Biol. 96 (2014) 143–180.
44.. J. Kleinjung, F. Fraternali, Curr. Opin. Struct. Biol. 25 (2014) 126–134.
45.. K. Hadley, C. McCabe, Mol. Simulat. 38 (8–9) (2012) 671–681.
46.. S.J. Marrink, D.P. Tieleman, Chem. Soc. Rev. 42 (16) (2013) 6801–6822.
47.. C.A. Lopez, A.J. Rzepiela, A.H. de Vries, L. Dijkhuizen, P.H. Hunenberger, S.J. Marrink, J. Chem. Theory Comput. 5 (12) (2009) 3195–3210.
48.. S.O. Yesylevskyy, L.V. Schafer, D. Sengupta, S.J. Marrink, PLoS Comput. Biol. 6 (6) (2010), e1000810.
49.. H. Ma, F.J. Irudayanathan, W. Jiang, S. Nangia, J. Phys. Chem. B. 119 (46) (2015) 14668–14682.

50.. P.S. Schmalhorst, F. Deluweit, R. Scherrers, C.P. Heisenberg, M. Sikora, J. Chem. Theory Comput. 13 (10) (2017) 5039–5053.
51.. M.E. Johnson, T. Head-Gordon, A. Louis, J. Chem. Phys. 126 (14) (2007) 144509.
52.. J.W. Wagner, J.F. Dama, A.E. Durumeric, G.A. Voth, J. Chem. Phys. 145 (4) (2016), 044108.
53.. J. Koca, J. Mol. Struct. (THEOCHEM) 308 (1994) 13–24.
54.. J. Koca, Prog. Biophys. Mol. Biol. 70 (1998) 137–173.
55.. S.B. Engelsen, J. Koca, I. Braccini, C. Herve du Penhoat, S. Perez, Carbohydr. Res. 276 (1995) 1–29.
56.. J. Koca, S. Perez, A. Imberty, J. Comput. Chem. 16 (3) (1995) 296–310.
57.. T. Peters, B. Meyer, R. Stuike-Prill, R. Somorjai, J.R. Brisson, Carbohydr. Res. 238 (1993) 49–73.
58.. T. Weimar, B. Meyer, T. Peters, J. Biomol. NMR 3 (1993) 399–414.
59.. A. Nahmany, F. Strino, J. Rosen, G.J. Kemp, P.G. Nyholm, Carbohydr. Res. 340 (5) (2005) 1059–1064.
60.. J. Rosen, L. Miguet, S.J. Perez, J. Chem. 1 (2009) 16.
61.. J.B. Lamarck, Philosophie Zoologique, Librarie F. Savy, Paris, 1809.
62.. T. Michell, Machine Learning, McGraw Hill, New York, 1997259–283.
63.. J. Rosen, A. Robobi, P.G. Nyholm, Carbohydr. Res. 337 (18) (2002) 1633–1640.
64.. J. Rosen, A. Robobi, P.G. Nyholm, Carbohydr. Res. 339 (5) (2004) 961–966.
65.. F. Strino, A. Nahmany, J. Rosen, G.J. Kemp, I. Sa-correia, P.G. Nyholm, Carbohydr. Res. 340 (5) (2005) 1019–1024.
66.. S.B. Engelsen, Water in Biomaterials Surface Science, John Wiley & Sons Ltd, 200153–89.
67.. S.B. Engelsen, C. Monteiro, C. Herve du Penhoat, S. Perez, Biophys. Chem. 93 (2001) 103–127.
68.. A. Singh, D. Montgomery, X. Xue, B.L. Foley, R.J. Woods, Glycobiology 29 (7) (2019) 515–518.
69.. S.B. Engelsen, P.I. Hansen, S. Perez, Biopolymers 101 (7) (2014) 733–743.
70.. M. Kuttel, J. Stahle, G. Widmalm, J. Comput. Chem. 37 (22) (2016) 2098–2105.
71.. S. Jo, T. Kim, V.G. Iyer, W. Im, J. Comput. Chem. 29 (2008) 1859–1865.
72.. P.-C. Hsu, B.M.H. Bruininks, D. Jefferies, P.C.T. Souza, J. Lee, D.S. Patel, S.J. Marrink, Y. Qi, S. Khalid, W. Im, J. Comput. Chem. 38 (27) (2017) 2354–2363.
73.. R. Danne, C. Poojari, H. Martinez-Seara, S. Rissanen, F. Lolicato, T. Rog, I. Vattulainen, J. Chem. Inf. Model. 57 (10) (2017) 2401–2406.
74.. T.A. Wassenaar, H.I. Ingolfsson, R.A. Bockmann, D.P. Tieleman, S.J. Marrink, J. Chem. Theory Comput. 11 (5) (2015) 2144–2155.
75.. S.R. Stowell, T. Ju, R.D. Cummings, Annu. Rev. Pathol. 10 (2015) 473–510.
76.. S.S. Pinho, C.A. Reis, Nat. Rev. Cancer 15 (9) (2015) 540–555.
77.. C. Xu, D.T. Ng, Nat. Rev. Mol. Cell Biol. 16 (12) (2015) 742–752.
78.. A.M. Harbison, L.P. Brosnan, K. Fenlon, E. Fadda, Glycobiology 29 (1) (2019) 94–103.
79.. S. Battella, M.C. Cox, A. Santoni, G. Palmieri, J. Leukoc. Biol. 99 (1) (2016) 87–96.
80.. J.M. Hayes, A. Frostell, R. Karlsson, S. Muller, S.M. Martin, M. Pauers, F. Reuss, E.F. Cosgrave, C. Anneren, G.P. Davey, P.M. Rudd, Mol. Cell. Proteomics 16 (10) (2017) 1770–1788.
81.. S. Iida, H. Misaka, M. Inoue, M. Shibata, R. Nakano, N. Yamane-Ohnuki, M. Wakitani, K. Yano, K. Shitara, M. Satoh, Clin. Cancer Res. 12 (2006) 2879–2887.
82.. H. Xu, H. Guo, I.Y. Cheung, N.K. Cheung, Cancer Immunol. Res. 4 (7) (2016) 631–638.
83.. A. Harbison, E. Fadda, bio Rxiv (2019).
84.. B. Echeverria, S. Serna, S. Achilli, C. Vives, J. Pham, M. Thepaut, C.H. Hokke, F. Fieschi, N.C. Reichardt, ACS Chem. Biol. 13 (8) (2018) 2269–2279.
85.. C. Ferrara, F. Stuart, P. Sondermann, P. Brunker, P. Umana, J. Biol. Chem. 281 (8) (2006) 5032–5036.
86.. X. Wei, J. Decker, S. Wang, H. Hui, J. Kappes, X. Wu, J. Salazar-Gonzalez, M. Salazar, J. Kilby, M. Saag, N. Komarova, M. Nowak, B. Hahn, P.
Kwong, G. Shaw, Nature 422 (6929) (2003) 307–312.
87.. G. Stewart-Jones, C. Soto, T. Lemmin, G. Chuang, A. Druz, R. Kong, P. Thomas, K. Wagh, T. Zhou, A. Behrens, T. Bylund, C. Choi, J. Davison, I.
Georgiev, M. Joyce, Y. Do Kwon, M. Pancera, J. Taft, Y. Yang, B. Zhang, S. Shivatare, V. Shivatare, C. Lee, C. Wu, C. Bewley, D. Burton, W. Koff, M. Connors, M. Crispin, U. Baxa, B. Korber, C. Wong, J. Mascola, P. Kwong, Cell 165 (4) (2016) 813–826.
88.. R. Kong, H. Duan, Z. Sheng, K. Xu, P. Acharya, X. Chen, C. Cheng, A.S. Dingens, J. Gorman, M. Sastry, C.-H. Shen, B. Zhang, T. Zhou, G.-Y. Chuang,
C.W. Chao, Y. Gu, A.J. Jafari, M.K. Louder, S. O'Dell, A.P. Rowshan, E.G. Viox, Y. Wang, C.W. Choi, M.M. Corcoran, A.R. Corrigan, V.P. Dandey,
E.T. Eng, H. Geng, K.E. Foulds, Y. Guo, Y.D. Kwon, B. Lin, K. Liu, R.D. Mason, M.C. Nason, T.Y. Ohr, L. Ou, R. Rawi, E.K. Sarfo, A. Schon, J.P.
Todd, S. Wang, H. Wei, W. Wu, NISC Comparative Sequencing Program, J.C. Mullikin, R.T. Bailer, N.A. Doria-Rose, G.B. Karlsson Hedestam, D.G. Scorpio, J. Overbaugh, J.D. Bloom, B. Carragher, C.S. Potter, L. Shapiro, P.D. Kwong, J.R. Mascola, Cell 178 (2019) 567–584.

89.. T. Lemmin, C. Soto, J. Stuckey, P. Kwong, Structure 25 (10) (2017) 1631–1639.
90.. M. Yang, J. Huang, R. Simon, L.X. Wang, A.D. MacKerell, Sci. Rep. 7 (1) (2017) 4435.
91.. H. Koldso, D. Shorthouse, J. Helie, M.S.P. Sansom, PLoS Comput. Biol. 10 (10) (2014), e1003911.
92.. D. Shorthouse, G. Hedger, H. Koldso, M. Sansom, Biochimie 120 (2016) 105–109.
93.. T. Shen, S. Gnanakaran, Biophys. J. 96 (8) (2009) 3032–3040.
94.. G. Srinivas, X. Cheng, J.C. Smith, J. Chem. Theory Comput. 7 (2011) 2539–2548.
95.. G. Srinivas, X. Cheng, J.C. Smith, J. Phys. Chem. B. 118 (2014) 3026–3034.
96.. G. Bellesia, S.P.S. Chundawat, P. Langan, A. Redondo, B.E. Dale, S. Gnanakaran, J. Phys. Chem. B. 116 (28) (2012) 8031–8037.
97.. D.C. Glass, K. Moritsugu, X. Cheng, J.C. Smith, Biomacromolecules 139 (2012) 2634–2644.
98.. S. Markutsya, A. Devarajan, J.Y. Baluyut, T.L. Windus, M.S. Gordon, M.A. Lamm, J. Chem. Phys. 138 (2013) 214108.
99.. C.A. Lopez, G. Bellesia, A. Redondo, P. Langan, S.P.S. Chundawat, B.E. Dale, S.J. Marrink, S. Gnanakaran, J. Phys. Chem. B. 119 (2015) 465–473.
100..F. Bingxin, J.K. Maranas, Cellul. 22 (2015) 31–44.
101..A.B. Poma, M. Chwastyk, M. Cieplak, Cellul. 23 (2016) 1573–1591.
102..A.B. Poma, M. Chwastyk, M. Cieplak, Phys. Chem. Chem. Phys. 19 (41) (2017) 28195–28206.
103..R. Mirzaeifar, Z. Qin, M.J. Buehler, Nanoscale 7 (2015) 5435–5445.
104..M. Shishehbor, P.D. Zavattieri, J. Mech. Phys. Solids 124 (2019) 871–896.
105..M.G. Ramezani, B. Golchinfar, J. Compos. Sci. 3 (2019) 57.
106..J. Sauter, A. Grafmuller, J. Chem. Theory Comput. 13 (1) (2017) 223–236.
107..L. Li, P. Perre, X. Frank, K. Mazeau, Carbohydr. Polym. 127 (2015) 438–450.
108..S.W. Benner, C.K. Hall, J. Phys. Chem. B. 120 (29) (2016) 7253–7264.
109..L. Tsereteli, A. Grafmuller, PLoS One 12 (7) (2017), e0180938.
110..M. Bathe, G.C. Rutledge, A.J. Grodzinsky, B. Tidor, Biophys. J. 88 (2005) 3870–3887.
111..B.M. Sattelle, J. Shakeri, A. Almond, Biomacromolecules 14 (2013) 1149–1159.
112..S.A. Samsonov, L. Bichmann, M.T. Pisabarro, J. Chem. Inf. Model. 55 (2015) 114–124.
113..S. Samsonov, E.A. Lubecka, K.K. Bojarski, R. Ganzynkowicz, A. Liwo, Biopolymers 110 (2019), e23269.
114..A. Almond, Curr. Opin. Struct. Biol. 50 (2018) 58–64.
115..D.J. Beltran-Villegas, D. Intriago, K.H.C. Kim, N. Behabtu, J.D. Londono, A. Jayaraman, Soft Matter 15 (23) (2019) 4669–4681.
116..A.F. Ghobadi, A. Jayaraman, Soft Matter 12 (8) (2016) 2276–2287.
117..D.A. Brant, Q. Rev. Biophys. 9 (1976) 527–596.
118..O.N. Makshakova, T.A. Gorshkova, P.V. Mikshina, Y.F. Zuev, S. Perez, Carbohydr. Polym. 158 (2017) 93–101.
119..D. Jefferies, J. Shearer, S. Khalid, J. Phys. Chem. B. 123 (17) (2019) 3567–3575.
120..S. Khalid, T.J. Piggot, F. Samsudin, Acc. Chem. Res. 52 (1) (2019) 180–188.
121..Yaniv, O.; Reddy, Y. H. K.; Bayer, E. A.; Lamed, R.; Frolow, F. (to be published)
122..T. Nomura, H. Iwase, N. Saka, N. Takahashi, B. Mikami, K. Mizutani, Acta Crystallogr. D 75 (2019) 426–436.
123..T. Matsuzawa, T. Jo, T. Uchiyama, J.A. Manninen, T. Arakawa, K. Miyazaki, S. Fushinobu, K. Yaoi, FEBS J. 283 (2016) 2340–2353.
124..A.S. Borisova, E.V. Eneyskaya, S. Jana, S.F. Badino, J. Kari, A. Amore, M. Karlsson, H. Hansson, M. Sandgren, M.E. Himmel, P. Westh, C.M. Payne,
A.A. Kulminskaya, J. Stahlberg, Biotechnol. Biofuels 11 (2018) 5.
125..C.A. Fowler, F. Sabbadin, L. Ciano, G.R. Hemsworth, L. Elias, N. Bruce, S. McQueen-Mason, G.J. Davies, P.H. Walton, Biotechnol. Biofuels 12 (2019) 232.
126..L. Petersen, A. Ardevol, C. Rovira, P.J. Reilly, J. Phys. Chem. B. 113 (2009) 7331–7339.
127..B.C. Knott, M. Haddad Momeni, M.F. Crowley, L.F. Mackenzie, A.W. Gotz, M. Sandgren, S.G. Withers, J. Stahlberg, G.T. Beckham, J. Am. Chem. Soc. 136 (2014) 321–329.
128..G. Vaaje-Kolstad, B. Westereng, S.J. Horn, Z. Liu, H. Zhai, M. Sorlie, V.G. Eijsink, Science 330 (2010) 219–222.
129..B. Wang, E.M. Johnston, P. Li, S. Shaik, G.J. Davies, P.H. Walton, C. Rovira, ACS Catalysis 8 (2018) 1346–1351.
130..B. Wang, P.H. Walton, C. Rovira, ACS Catalysis 9 (2019) 4958–4969.
131..J. Coines, L. Raich, C. Rovira, Curr. Opin. Chem. Biol. 53 (2019) 183–191.
132..A.B. Poma, M. Chwastyk, M. Cieplak, J. Phys. Chem. B. 119 (2015) 12028–12041.
133..J. Wohlert, L.A. Berglund, J. Chem. Theory Comput. 7 (2011) 753–760.
134..A. Asztalos, M. Daniels, A. Sethi, T. Shen, P. Langan, A. Redondo, S. Gnanakaran, Biotechnol. Biofuels 5 (1) (2012) 55.
135..M. Reynolds, S. Perez, C. R. Chim. 14 (1) (2011) 74–95.
136..R. Woods, M. Tessier, Curr. Opin. Struct. Biol. 20 (5) (2010) 575–583.
137..D. de Sanctis, S. Perez, Beilstein J. Org. Chem. 13 (2017) 1145–1167.

138..I. Tvaroška, in: T. Lutteke, M. Frank (Eds.), Glycoinformatics, Methods in Molecular Biology, Springer Science + Business Media, New York, 2015, pp. 489–499.
139..M. Kumari, S. Kozmon, P. Kulhanek, J. Stepan, I. Tvaroska, J. Koca, J. Phys. Chem. B. 119 (12) (2015) 4371–4381.
140..L.L. Lairson, B. Henrissat, G.F.J. Davies, S.G. Withers, Annu. Rev. Biochem. 77 (2008) 521–555.
141..D. Albesa-Jove, M.E. Guerin, Curr. Opin. Struct. Biol. 40 (2016) 23–32.
142..Makshakova, O.; Breton, C.; Perez, S., Sci. Rep. 10 (2020) 13514
143..M. Nitenberg, O. Makshakova, J. Rocha, S. Perez, E. Marechal, M.A. Block, A. Girard-Egrot, C. Breton, Glycobiology 30 ,(2020), 396-406 https://doi.org/10.1093/ glycob/cwz106.
144..A. Warshel, M. Levitt, J. Mol. Biol. 103 (1976) 227–249.
145..I. Andre, G. Potocki-Veronese, S. Barbe, C. Moulis, M. Remaud-Simeon, Curr. Opin. Chem. Biol. 19 (2014) 17–24.
146..A. Ardevol, C. Rovira, J. Am. Chem. Soc. 137 (2015) 7528–7547.
147..S.A. Jongkees, S.G. Withers, Acc. Chem. Res. 47 (2014) 226–235.
148..M.K. Bilyard, H.J. Bailey, L. Raich, M.A. Gafitescu, T. Machida, J. Iglesias-Fernandez, S.S. Lee, C.D. Spicer, C. Rovira, W.W. Yue, B.G. Davis, Nature 563 (7730) (2018) 235–240.
149..V.A. Streltsov, S. Luang, A. Peisley, J.N. Varghese, J.R. Ketudat Cairns, S. Fort, M. Hijnen, I. Tvaroška, A. Arda, J. Jimenez-Barbero, M.Alfonso-Prieto, C. Rovira, F. Mendoza, L. Tiessler-Sala, J.E. Sanchez-Aparicio, J. Rodriguez-Guerra, J.M. Lluch, J.D. Marechal, L. Masgrau, M. Hrmova, Nat. Commun. 2222 (2019) 1–17.
150..P. Csermely, R. Palotai, R. Nussinov, Trends Biochem. Sci. 35 (10) (2010) 539–546.
151..A. Mohan, C. Oldfield, P. Radivojac, V. Vacic, M. Cortese, A. Dunker, V. Uversky, J. Mol. Biol. 362 (5) (2006) 1043–1059.
152..J. Topin, M. Lelimousin, J. Arnaud, A. Audfray, S. Perez, A. Varrot, A. Imberty, ACS Chem. Biol. 11 (7) (2016) 2011–2020.
153..S. Perez, N. Mouhous-Riou, N. Nifantev, Y. Tsvetkov, B. Bachet, A. Imberty, Glycobiology 6 (5) (1996) 537–542.
154..H. Azurmendi, M. Martin-Pastor, C. Bush, Biopolymers 63 (2) (2002) 89–98.
155..G.M. Torrie, J.P. Valleau, J. Comput. Phys. 23 (2) (1977) 187–199.
156..I. Alibay, K. Burusco, N. Bruce, R. Bryce, J. Phys. Chem. B 122 (9) (2018) 2462–2474.
157..D. Hamelberg, J. Mongan, J.A. McCammon, J. Chem. Phys. 120 (24) (2004) 11919–11929.
158..M. Agostino, P.A. Ramsland, E. Yuriev, in: E. Yuriev, P.A. Ramsland (Eds.), Structural Glycobiology, 1st ed., CRC Press, Boca Raton, FL, 2012, pp. 111–138.
159..A.S. Sarkar, S. Perez, in: E. Yuriev, P.A. Ramsland (Eds.), Structural Glycobiology, 1st ed., CRC Press, Boca Raton, FL, 2012, pp. 71–110.
160..T.J.A. Ewing, S. Makino, A.G. Skillman, I.D. Kinz, J. Comput. Aided Mol. Des. 15 (2001) 411–428.
161..R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, J. Med. Chem. 47 (2004) 1739–1749.
162..K.L. Hudson, G.J. Bartlett, R.C. Diehl, J. Agirre, T. Gallagher, L.L. Kiessling, D.N. Woolfson, J. Am. Chem. Soc. 137 (2015) 15152–15160.
163..R. Marchetti, S. Perez, A. Arda, A. Imberty, J. Jimenez-Barbero, A. Silipo, A. Molinaro, ChemistryOpen 5 (4) (2016) 274–296.
164..R.E. Forgione, C. Di Carluccio, M. Kubota, Y. Manabe, R. Fukase, A. Molinaro, T. Hashiguchi, R. Marchetti, A. Silipo, Sci. Rep. 10 (2020) 1589.
165..M.-S. Park, PLoS One 10 (4) (2015), e0125361.
166..T. Galochkina, M. Ng Fuk Chong, L. Challali, S. Abbar, C. Etchebest, Sci. Rep. 9 (2019) 998.
167..R. Dutzler, T. Schirmer, M. Karplus, S. Fischer, Structure 10 (9) (2002) 1273–1284.
168..B.M. Sattelle, J. Shakeri, M.J. Cliff, A. Almond, Biomacromolecules 16 (2015) 951–961.
169..A. Imberty, H. Lortat-Jacob, S. Perez, Carbohydr. Res. 342 (3–4) (2011) 430–439.
170..A. Seffouh, R. El Masri, O. Makshakova, E. Gout, H. ZEO, J.P. Andrieu, H. Lortat-Jacob, R.R. Vives, Cell. Mol. Life Sci. 76 (9) (2019) 1807–1819.
171..A. Singh, M.B. Tessier, K. Pederson, X. Wang, A.P. Venot, G.J. Boons, J.H. Prestegard, R.J. Woods, Can. J. Chem. 94 (2016) 927–935.
172..L. Foley, M. Tessier, R. Woods, Wiley Int. Rev. Comp. Mol. Sci. 2 (2012) 652–697.
173..N.V. Sankaranarayanan, B. Nagarajan, U.R. Desai, Curr. Opin. Struct. Biol. 50 (2018) 91–108.
174..Alocci, D.; Lisacek, F.; Perez, S. 2019 A Traveler's Guide to Complex Carbohydrates in the Cyber Space. Available at: http://www.glycopedia.eu
175..S. Perez, A. Sarkar, A. Rivet, C. Breton, A. Imberty, Methods Mol. Biol. 1273 (2015) 241–258.
176..F. Bonnardel, A. Kumar, M. Wimmerova, M. Lahmann, S. Perez, A. Varrot, F. Lisacek, A. Imberty, Structure (2019) 764–775.
177..S. Perez, T. Tubiana, A. Imberty, M. Baaden, Glycobiology 25 (5) (2015) 483–491.
178..M. Battistel, H. Azurmendi, M. Frank, D. Freedberg, J. Am. Chem. Soc. 137 (42) (2015) 13444–13447.
179..G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, J. Comput. Chem. 19 (1998) 1639–1662.

Glossary

ADCC antibody-dependent cellular cytotoxicity
CG coarse grain
CNC cellulose nano crystals
CORCEMA-ST complete relaxation and conformational exchange matrix—saturation transfer
GA genetic algorithm
GAG glycosaminoglycans
GH glycosidic hydrolase
GT glycosyl transferase
IgG1 immunoglobulin G1
MD molecular dynamics
MM molecular mechanics
QM quantum mechanics
SNFG symbol nomenclature for glycans

Biography

Serge Perez was born in Perigueux, France, and graduated, in crystallography, from the University of Bordeaux, France. He then spent one year at the Institute of Molecular Biology at the University of Oregon, Eugene, The USA and subsequently 3 years at the Department of Chemistry at the University of Montreal, Canada. He returned to France upon accepting a permanent position as a Junior Scientist of the Centre National de le Recherche Scientifique (CNRS) at Centre de Recherches sur les Macromolecules Vegetales (CERMAV), Grenoble, and he spent a sabbatical year working for Eastman Kodak in Rochester, USA. Later on, he moved on to Nantes, France, at the Institut de la Recherche Agronomique, where he was awarded a position as Senior Research Scientist to start a research group in Molecular Engineering. In January 1996 he moved back to Grenoble to become Chairman of the CERMAV, and hold this position for 11 years. He left this place to the European Synchrotron Radiation Facility to take up the position of Scientific Director in charge of the Life Sciences, Chemistry and X-ray Imaging programs. After completing his term, he returned to the CNRS in 2012. His research interests span across the whole area of structural and conformational analysis of oligosaccharides, polysaccharides, glycoconjugates and protein–carbohydrate interactions in solution and in the solid-state. This includes interests in computational chemistry and molecular modeling, crystallography, NMR spectroscopy, along with the structure-function and structure-properties relationship. He was also involved in management duties at the doctoral levels and in the creation of glyco-biotechnology companies. Besides his broad range interest in the field of structural glycoscience, Serge Perez is involved in international and international committees related to the evaluation of scientific research and the structuring of multilateral collaborative schemes. He has a keen interest in startup companies, and the economy of glycoscience, and *e*-learning as exemplified throughout the glycopedia experience. Most recently he founded the internet site: glycopedia.eu, an initiative aimed at promoting the field of glycosciences and provide material for educational purposes.

Elisa Fadda got her PhD in Theoretical and Computational Chemistry at the Universite de Montreal in 2004 with a study on the determination of NMR shieldings ab initio, from time-dependent density-functional theory (TD-DFT) under the supervision of Prof Dennis R. Salahub. She then joined Dr. Regis Pomes' group in Molecular Structure and Function at the Hospital for Sick Children (Sickkids) Research Institute as a postdoctoral fellow, where she strengthened her background in computer simulation techniques specializing in statistical mechanics-based approaches to study the mechanisms and energetics of ion translocation events through membrane proteins. In 2008 she joined Prof Robert J. Woods' group in the School of Chemistry at NUI Galway as a research fellow, where she applied and further developed her expertise in computational biophysics for the study of glycans, glycan-binding proteins and glycan-processing enzymes for applications in glycoengineering. In 2013 she joined the Department of Chemistry at Maynooth University as an Assistant Lecturer and in 2014 as an Assistant Professor (Lecturer), with tenure in 2018. Elisa Fadda is author of 23 peer-reviewed articles, two reviews and three book chapters and she is the senior author (first, last and/or corresponding) of 17 of these contributions. She is currently based at the Hamilton Institute at Maynooth University, where her research interests and work are focused in understanding sequence to structure and function relationships, structural disorder and molecular recognition of complex carbohydrates by computer simulation methods and in developing new computational tools to accelerate glycoanalytics and advance glycomics.

Olga Makshakova graduated in physics from Kazan State University, Russian Federation. Then she proceeded the application of both experimental and theoretical methods of physics to study the spatial structure and interactions of biological macromolecules in the Laboratory of Molecular Biophysics (under the supervision of Prof. Vladimir D. Fedotov) of Kazan Institute of Biochemistry and Biophysics of Russian Academy of Sciences. She got her PhD in Biophysics in 2010 with the thesis on the joint application of FTIR-spectroscopy and quantum chemical (DFT) calculations to biopolymer structure and their hydrated state. After the defence of PhD, she got the permanent position in Kazan Institute of Biochemistry and Biophysics of Russian Academy of Sciences. In the field of her interest there remain structure and dynamics of biopolymers studied with optical and NMR-spectroscopy, computational modeling of bio-macromolecular interactions by classical MD in All-Atom and Coarse-Grained presentation. Since 2015 thanks to the collaboration with Serge Perez, she has been involved in the field of glycoscience with a particular focus on polysaccharide structure and glycoconjugate assembly on the lipid membrane interface. She is an author of 20 peer-reviewed articles, out of them to 11 she contributed as a principal co-author (first or last corresponding author). Now she is a senior researcher in the laboratory of biophysical chemistry of nanosystems (Prof. Zuev) of Kazan Institute of Biochemistry and Biophysics of Russian Academy of Sciences. Currently, her work resides in the field of bio-macromolecular assembly and structure induced-fit under protein–carbohydrate interactions for innovative bio-inspired materials.

Comprehensive Glycoscience, 2nd edition, https://doi.org/10.1016/B978-0-12-819475-1.00004-3