OPEN ACCESS

molecules

Review

Investigation of Carbohydrate Recognition via Computer Simulation

Quentin R. Johnson [1], **Richard J. Lindsay** [2], **Loukas Petridis** [3] **and Tongye Shen** [2,*]

[1] UT-ORNL Graduate School of Genome Science and Technology, Knoxville, TN 37996, USA;
E-Mail: qjohnso3@utk.edu

[2] Department of Biochemistry and Cellular & Molecular Biology, University of Tennessee, Knoxville,
TN 37996, USA; E-Mail: rlindsa2@utk.edu

[3] Center for Molecular Biophysics, Oak Ridge National Lab, Oak Ridge, TN 37830, USA;
E-Mail: petridisl@ornl.gov

* Author to whom correspondence should be addressed; E-Mail: tshen@utk.edu;
Tel.: +1-865-9744088; Fax: +1-865-9746306.

Academic Editor: Els Van Damme

Abstract: Carbohydrate recognition by proteins, such as lectins and other (bio)molecules, can be essential for many biological functions. Recently, interest has arisen due to potential protein and drug design and future bioengineering applications. A quantitative measurement of carbohydrate-protein interaction is thus important for the full characterization of sugar recognition. We focus on the aspect of utilizing computer simulations and biophysical models to evaluate the strength and specificity of carbohydrate recognition in this review. With increasing computational resources, better algorithms and refined modeling parameters, using state-of-the-art supercomputers to calculate the strength of the interaction between molecules has become increasingly mainstream. We review the current state of this technique and its successful applications for studying protein-sugar interactions in recent years.

Keywords: protein-carbohydrate interaction; binding free energy; molecular dynamics simulation; lectin

---

1. Introduction

Carbohydrates are an important class of macromolecules that serve various roles in biological systems [1,2]. Besides the commonly-known functions for metabolic energy (saccharides, such as starch and glycogen, provide fuel for biological systems) and for mechanical support (saccharides, such as cellulose, are the essential building block of cell walls), carbohydrates can play roles in signal transduction, immune response, cell trafficking, cell adhesion and cell-cell interaction. Many of these roles involve protein-carbohydrate interaction and selective protein recognition in the presence of carbohydrates.

The nature of the structure and physical properties of carbohydrates presents unique challenges for designing carbohydrate recognition sites. First of all, at the monosaccharide level, sugar residues are largely similar. There exists a large set of common monosaccharides; many of them are isomers with subtle differences, and they lack distinct chemical groups. Secondly, beyond the monomer level, the potential connection of these monomers via glycosidic bonds is even more complex. Thirdly, glycomics are not directly coded in genomics, unlike proteomics. Thus, the primary structures of sugars (sequence information) are a lot more susceptible to the stochastic nature of chemical reaction events [3]. Such randomness makes the sensing of the consensus of sugar structures difficult. This feature can even be exploited by viruses to evade antibodies [4].

These features of carbohydrates also make the quantitative study of carbohydrate recognition using many traditional methods difficult in several aspects. (1) Impurity: One can imagine that the heterogeneity of primary structures makes purification difficult, which is the necessary first step of any *in vitro* examination. (2) Specificity: The word "recognition" means more than a strong binding affinity between a sugar and the corresponding sensing protein. Strong binding means little if the protein cannot selectively target a specific sugar. Thus, the selectivity of a particular protein-carbohydrate interaction can only be evaluated by examining that protein interaction with large sets of "background noise" carbohydrates (sugars with similar structures to the target). (3) Cooperativity: Since the sugar-protein interactions are largely weak, often multiple binding sites can get involved to enhance the signal. Cooperative binding can be involved in sugar recognition. Cooperativity makes the binding affinity of sites difficult to measure, as the properties of one site depend on the binding status of the others. We argue that these difficulties can be circumvented using the computational study of sugar recognition [5–8]. Indeed, in the last three decades, there have been extensive applications of computer simulations of biomolecular systems to characterize their biophysical features. Computational methods can be used to directly model a desired sugar species and avoid the purification problem completely. Large-scale parallel computing can also be used to scan a large set of sugar structures and binding scenarios with great efficiency. Furthermore, perturbative binding affinity calculations can provide high-resolution results in a cost-effective fashion. Moving forward, we will review the current progress made in the field of the computational study of protein-sugar interactions, the successes achieved and the challenges faced. Since there are many successful results achieved in this field, this work is not meant to be an extensive review of the whole field. Rather, we focus on raising awareness that computational methods can be a powerful tool in the field of sugar recognition.

With a better understanding of nature's design of proteins that recognize sugars through experimental and computational studies, we may further enhance our ability to engineer drugs and proteins. For example, oseltamivir and other drugs were designed to block the flu virus [9], while a set of novel proteins block HIV [10]. We may also further design molecular sensors and even pathways involving the desired carbohydrate-protein interactions. Beyond recognition by proteins, the specific interaction between carbohydrates and RNA is another important area that can benefit from the study of sugar recognition [11]. Likely, many of the computational methods developed for studying protein-sugar interaction will also be applicable to the RNA-sugar interaction.

2. Structural Features of Carbohydrate-Recognizing Proteins

Obtaining the structural information of protein-carbohydrate complexes is often the first step for evaluating the strength and specificity of carbohydrate recognition via computer simulations. Since there is a variety of carbohydrates, there are many carbohydrate-binding proteins and carbohydrate-active enzymes that target these sugar molecules. Thus, there are various ways of classifying these interactions. One is to classify the interactions according to the size of the carbohydrate. Some proteins are mono- and oligo-saccharide sensors, while others interact with oligosaccharides and polysaccharides. Another classification is from the protein perspective, whether the protein is purely a binding protein or contains a catalytic module (of the same polypeptide chain). The former group, those that only bind specifically to carbohydrates, are called lectins, while the latter are carbohydrate-active enzymes (CAZymes) that modify the sugars with which they interact. We will review the structural features of the sugar recognition components of both lectins and CAZymes. Both CAZymes and lectins use β-strands in their secondary structure to construct recognition motifs and to design the binding site in general. Examples of sugar-interacting structures from a lectin and a CAZyme are shown in Figure 1. It is interesting to point out that many lectins and CAZymes may share similar folds, such as the β-trefoil.

Lectins are a group of proteins known for their ability to recognize saccharides, but unlike CAZymes, they usually lack enzymatic activity. While lectins have a relatively weak affinity for monosaccharides due to the shallow nature of these binding pockets, they show an increased binding affinity for more complex oligo- and poly-saccharides [12,13]. They are commonly observed as oligomer aggregates, and each lectin monomer typically contains from one to three carbohydrate-recognition domains (CRDs) [12,14]. This quality allows multivalent interaction with complex carbohydrates, contributing to the specificity and the increasing binding strength [13,14]. Lectins can be classified according to the primary and tertiary structural similarities of their CRDs into L-type, C-type, R-type and several others [15,16]. The most common of these, C-type lectins, are characterized by their calcium dependence and a fold, called the C-type lectin-like domain (CTLD) [17]. This motif consists of a double loop joined by two disulfide bridges, with conserved hydrophobic and hydrophilic residues [18]. All CTLDs of carbohydrate-interacting proteins include a calcium ion, which mediates substrate binding [18,19]. R-type lectins contain CRDs with structures similar to the fold of the ricin B chain, a structure called a β-trefoil fold [2,20]. On the other hand, L-type lectin CRDs share a jelly-roll fold, a β-barrel structure formed by multiple pairs of antiparallel β-sheets [2].
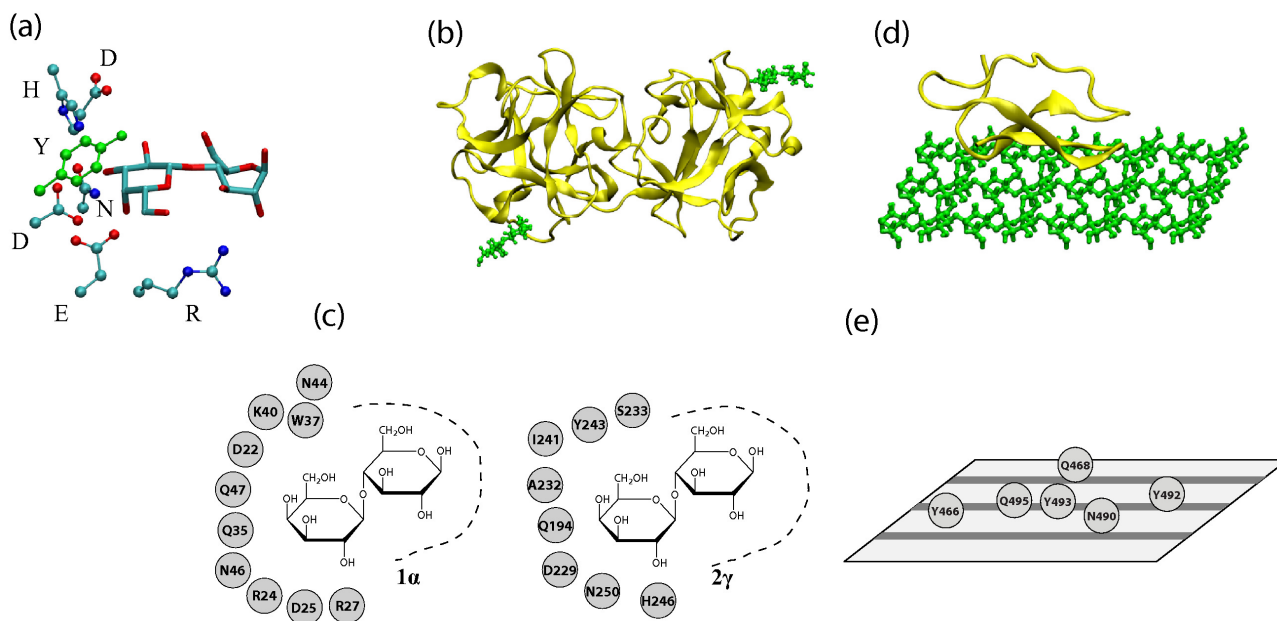
Figure 1. (**a**) A three-dimensional representation of a binding pocket of a lectin (2γ site of the ricin B chain). Only the sugar residues and amino acid residues involved in ring-stacking and hydrogen bonding are shown. Hydrogen and backbone atoms are excluded for clarity. (**b**) A three-dimensional representation of a lectin (ricin chain B, PDB Code 2AAI) with lactoses bound at sites 1α (lower-left) and 2γ (upper-right). (**c**) A depiction of binding sites 1α and 2γ with possible hydrogen bond and ring stacking residues shown. (**d**) A three-dimensional representation of the cellulose binding module (CBM) of a CAZyme (Cel7A with cellulose modeled by M. Crowley's group) (**e**) Residues of Cel7A's CBM that interact with cellulose. Cellulose chains are represented by gray stripes on the plane.

Carbohydrate-active enzymes are proteins that break down, construct or otherwise modify carbohydrates [21,22]. In addition to a catalytic module, CAZymes contain carbohydrate-binding modules (CBMs) responsible for substrate recognition. One role of CBMs is to maintain a high concentration of enzyme near the substrate polysaccharide. Keeping the catalytic domain in proximity to its substrate increases the efficiency of the enzyme. Some CBMs have been found to target not only particular glycans, but particular segments of polysaccharide chains. Some CBMs can also serve to disrupt carbohydrate surfaces [22]. Upon binding to a substrate surface, the crystalline organization is disrupted and the surface area of the substrate is increased, making it more accessible to the catalytic domain.

CBMs can be classified in multiple ways. Particularly, they may be classified into three "types" from the carbohydrate perspective. Type A modules interact with insoluble glycan surfaces (such as the surface of cellulose), Type B binds individual glycan chains, and Type C binds small sugars, mono-, di- and tri-saccharides. CBMs can also be classified from the protein perspective in detail. While CBMs show a huge amount of variation in amino acid sequence (70 families), their tertiary structures can be described in terms of just seven "fold families" according to the similarities of their tertiary structure [21,23]. While these fold families share structural motifs, due to varying sequences, it is not always possible to predict specific sugar recognition based on the fold family alone.

The most commonly-observed fold family is the so-called "β-sandwich", composed of two β-sheets, each with three to six antiparallel β-strands and always containing a metal ion [24]. The next most common fold is the "β-trefoil" fold, which is consistent with that of the R-type of ricin. This fold consists of three subunits with the resultant structure being a β-barrel capped by three hairpin turns, capable of binding up to three carbohydrate molecules [21,24]. While the β-sandwich and β-trefoil fold structures are adaptable to many different substrates, some fold families are more specific. For example, the cellulose-binding and oligosaccharide/oligonucleotide-binding (OB) motifs appear to be specialized for cellulose and chitin recognition. These are small planar CBMs, composed of a β-sheet and coil, whose substrate specificity is mediated primarily by ring interactions with aromatic amino acid residues. Another example is the hevein fold family, comprised of mostly a coil with two small β-sheets and a short helical region and specific to chitin. A variation of the hevein fold with an additional β-sheet is found in one CBM family and makes up another fold family [21].

All of these classifications of protein-sugar interactions come from the collections of the structural information of the complexes. Databases exist to help classify carbohydrate-binding proteins. These databases and associated tools also serve as the starting point of computer simulation and modeling [25–29]. For example, the Glyco3D database is comprised of six databases containing three-dimensional structures and other information for glycans and proteins that interact with them, including lectins, glycosaminoglycan binding proteins, monoclonal antibodies with carbohydrate ligands and glycosyltransferases [26]. The Carbohydrate-Active Enzymes database (CAZy) contains present knowledge of experimentally-characterized CAZymes [23]. General algorithms exist to classify three-dimensional carbohydrate structures in PDBs and to describe how they interact with the protein environment [27]. These structural databases can provide valuable bioinformatics on protein-carbohydrate interaction and can serve as the starting point of high-resolution computer simulation study of carbohydrate recognition.

3. Physical Interactions at the Recognition Sites

The actual recognition by proteins is realized by the complicated, but direct physical interactions between the protein and carbohydrate [30]. There are five types of physical interactions often involved in recognition: (1) hydrogen bonding; (2) sugar ring-aromatic ring stacking; (3) salt bridge interaction of charged carbohydrate moieties; (4) metal ion-mediated interaction; and (5) solvent-mediated interaction.

Hydrogen bonding is the most important interaction between carbohydrates and proteins. Each carbohydrate residue generally has multiple hydroxyl groups and a ring oxygen. These chemical groups act as hydrogen donors and/or acceptors. Often, the recognition site of the protein has either the side chains of polar amino acid residues or protein backbone components that form hydrogen bonds with the carbohydrate. Since most hydrogen bonds formed are of the type O-H⋯O or N-H⋯O, the strength of the typical H-bond interaction involved in recognition is about 5–6 kcal/mol in a vacuum and 1.5–2 kcal/mol in aqueous solution [31,32].

Besides the well-studied hydrogen bonding, another important, but yet to be further explored, interaction is the stacking interaction [30,33,34] between the pyranose ring of carbohydrates and aromatic amino acids, Trp, Phe and Tyr. Having a similar nature to that of ring-ring stacking interactions

($\pi - \pi$ bond) between the base pairs of DNA, this sugar ring-aromatic ring interaction (CH$-\pi$ bond) has been reported to be essential for the recognition of many monosaccharides. Through quantum chemistry calculations, this ring-ring interaction has been reported to be about 3–6 kcal/mol in the gas phase [33]. However, the experimental estimation of this interaction in solution is much weaker, reported to be about 0.5–0.8 kcal /mol [34].

A third type of physical interaction is direct Coulombic interaction. Although most sugar residues are neutral, a few important monosaccharides are negatively charged, such as the sialic acid family [1,2,35]. In such cases, the electrostatic interactions might be crucial for recognition. For example, positively-charged arginine is known to be essential for a group of sialic acid-recognizing lectins [36–38]. Thus, the salt-bridge interactions were suspected to play an important role in the recognition of certain sugars. However, since conservative mutation of Arg to Lys can abolish the binding affinity of a subset of these lectins, this indicates that interactions other than electrostatic could be more important for those specific cases [39].

Metal ions, such as calcium, can also play a role in protein-carbohydrate interaction [18]. There is a class of ion-mediated carbohydrate-protein interactions. Besides stabilizing the recognition site of the protein and, thus, indirectly enhancing the interaction, metal ions can form direct contacts with both the protein and sugar [18,40]. Besides ions explicitly coordinating the interactions, the solvation effect can also play a role in carbohydrate-protein binding. Although saccharides in general are largely hydrophilic, one would believe that desolvation effects play a smaller role in carbohydrate-protein interaction compared to that in protein-protein interaction [41]. Nevertheless, desolvation is important for some carbohydrate-binding proteins and their target polysaccharides. One well-known example is the interaction between the hydrophobic face of cellulose and the CBM of cellulase [42–44].

4. Computer Simulation of Protein-Carbohydrate Complexes

Computer simulation has become an important tool in molecular biology in general, since most research subjects in this field are microscopic and sensitive to environmental perturbation, which can make direct observation difficult [45]. The development of *in silico* methods [46] can provide a complementary approach to traditional approaches, such as nuclear magnetic resonance, crystallography and scattering techniques [47,48]. Atomistic simulation methods, after careful validation by comparison to experiment, are unmatched in efficiently presenting detailed and direct quantitative information with nanoscale resolution [49–51]. These simulation techniques directly track the three-dimensional movement of all of the atoms of the molecular complex according to physical laws and have been successfully applied to study the structure, dynamics and thermodynamics of biomolecules [52]. Simulations require two types of input information: (1) the initial structure of the system, which can be obtained from experimental data or modeling; and (2) the so-called "force field", empirical potential energy functions (derived from quantum chemistry calculations and/or fitted from experimental observations) that determine how atoms interact. Based on this information, a computer simulation can be performed for a complicated biomolecular system using supercomputers. The evolution of structures of the systems are recorded as time-stamped snapshots (trajectories) for analysis.

The early biological applications of atomistic molecular dynamics (MD) simulation focused on proteins, with the first simulation performed on the protein BPTI [53] almost 40 years ago. Thus, the force-fields of proteins are quite mature. Later, the application of simulations was expanded to include other important biomolecules, such as nucleic acids, lipids and carbohydrates, and corresponding force fields have been developed [54]. Particularly, carbohydrate force fields at all-atom resolutions, such as CHARMM [55–57], GLYCAM [5,58], GROMOS [59] and OPLS [60], have been used to study the behavior of polysaccharides [61,62]. MARTINI-based coarse-grained force fields, which provide a useful methodology to study large systems on long time scales, have also been developed for polysaccharides [63,64]. Besides the generic advanced methods used in sampling and free energy calculation that we will discuss below, advanced simulation algorithms that target the features of carbohydrate force fields have been developed, as well [51,65]. Indeed, the dramatic energy landscape of torsional angles, unique puckered ring conformations of carbohydrates and many other factors make the sampling of carbohydrates demanding [66]. These enhanced sampling methods ease the difficulty of crossing over high energy states of carbohydrates.

There are still several challenges for carbohydrate force field development. Polysaccharides, with their abundant hydroxyl groups, are known to form various intra- and inter-molecular hydrogen bonds. For example, modeling of the highly directional hydrogen bonds of polysaccharides in a crystal environment poses significant challenges and is hypothesized to lead to overestimation of carbohydrate crystal volumes [57,67]. In contrast, proteins have a well-defined secondary structural arrangement of hydrogen bonding, and nucleic acid hydrogen bonds can be easily tracked by base pairing. Furthermore, sugar residues are flexible biomolecules that adopt multiple stereoisomers, and care must be taken to optimize force fields that accurately treat their inherent conformational complexity [5,55]. For example, six-membered pyranose rings are puckered and adopt stable chair or boat conformations. Furthermore, when covalently linked to form polysaccharides, monosaccharides are linked with various types of glycosidic linkages, each of which can exist in either an α or β anomer [56,68]. Accurately quantifying the conformational properties of oligosaccharides based on the values of their glycosidic torsion angles is essential when predicting carbohydrate-protein structures [68].

Another potential challenge is a classical description of biomolecules and ions. For one, the calcium ions and metallic bonds in MD simulations [69] are traditionally difficult. Clearly, this may affect the small class of metal ion-mediated protein-sugar interactions. This is a general issue of metal ions in simulations and is not specific to our current topic. Furthermore, simulation of classic force fields provides fast calculations by only tracking the motions of nuclei. However, there are certain effects that cannot be described using such force fields. One well-known example is the cooperativity of hydrogen bonds [70]. Indeed, polarizable force fields may be necessary to accurately describe carbohydrate recognition and have been developed to incorporate such many-body interactions in carbohydrate simulations [71–73].

A further technical challenge exists when combining atomistic carbohydrate and protein force fields, which arises from inconsistency with the treatment of 1–4 interactions. Empirical force fields normally disregard the 1–2 (between two covalently bonded atoms) and 1–3 (between two atoms separated by two covalent bonds) nonbonded interactions. However, 1–4 nonbonded interactions (atom separated by three covalent bonds) are considered in addition to internal parameters (dihedral term) in the

determination of conformational energies. A 1–4 scaling factor of the non-bonded term is used to avoid potential exaggeration of short-range repulsion caused by the 6–12 Lennard–Jones interaction potential. For example, the AMBER [74] force field for proteins has chosen the scaling factor to be 1/1.2, whereas CHARMM force fields [75] do not scale this interaction, *i.e.*, the scaling factor is 1.0. Most carbohydrate force fields also do not scale the 1–4 interactions [5,59,75]. However, the OPLS carbohydrate force field includes 1–5 and 1–6 interaction scaling to improve the agreement with experiments [60]. In practical terms, in a simulation of a protein-carbohydrate non-covalent complex, one can simply keep a different 1–4 scaling factor for the protein and for the carbohydrate. However, in a simulation of a glycoprotein, in which a carbohydrate is covalently bonded to a protein, it is difficult to keep two separate scaling factors [76]. Nevertheless, simulations of glycosylated proteins often provide insight into the intramolecular interactions between amino acid and carbohydrate residues [77–80].

Despite these challenges, simulations of carbohydrates, glycoproteins and protein-carbohydrate complexes are helping researchers understand the mechanisms of carbohydrate recognition in various systems [8,66,81–84].

5. Free Energy Calculation of Protein-Carbohydrate Interactions

With the quantitative spatial data obtained by molecular dynamics simulations in hand, analytical methods are required to reveal the binding mechanism between protein and carbohydrate. As stated in the Introduction, ultimately, the recognition ability is measured as the selective and strong binding between the recognition site and the target carbohydrate. Thus, an important quantitative property that determines the strength of binding is free energy [85]. Although quantitative spatial data can be obtained directly by molecular dynamics simulations, further analysis methods are often required to obtain the thermodynamics of carbohydrate binding.

In general, the binding thermodynamics between two (bio)molecules in solution can be loosely divided into four terms: (1) a direct enthalpic term of binding energy, which is described directly by molecular mechanics (MM); (2) an entropic term due to the loss of overall translational and rotational degrees of freedom (DOFs) upon binding [85]; (3) an entropic term due to the change of the internal degrees of freedom [86]; and (4) a solvent term, including desolvation effects (such as the solvent accessible (SA) method [87]) of the binding and the polar solvent effects (such as Poisson–Boltzmann (PB) or generalized Born (GB) methods [88–90]). The overall rotational DOFs can be coupled with internal DOFs, and the clear separation for large and flexible molecules is difficult.

There are a large number of methods that derive free energy changes ($\Delta G = G_{cplx} - G_{protein} - G_{sugar}$) based on information obtained directly from simulation results. In this work, we do not intend to provide a comprehensive review on free energy calculation and the thermodynamics of binding in general. Rather, we present an overview of this subject with the focus on applications to carbohydrate-protein systems. Roughly speaking, two groups of free energy methods are commonly applied for studying sugar recognition. One group of methods, including the widely applied MM-PBSA ("MM" + "PB" + "SA") and MM-GBSA methods [41,87], focuses on the "end-point" calculations and can obtain the relative binding free energy. They often ignore the effect of the concentrations of sugar and protein. Thus, these methods do not provide the absolute binding free energy. However, they are demonstrated

to be suitable for the calculation of binding free energy of relatively large carbohydrate systems (beyond mono- and di-saccharides). On the other hand, the second group of methods, including the well-known double decoupling method [91,92] and other free energy perturbation/thermodynamics integration methods [52,93], applies rigorous statistical mechanics and can obtain the absolute binding free energy or accurate free energy changes between different states. This group of "alchemical transformation" methods is more suitable for the recognition of small sugars, as the "path" calculation for larger systems can be more computationally intensive. Thus, the two groups of methods can complement one another.

As it stands, there is no universally efficient method for free energy calculation for all types of systems; the suitable method depends on the nature of the question and the resolution of the answer desired. Therefore, depending on the size of the sugar and whether the absolute binding strength or just a relative binding free energy (specificity) is desired, different methods can be selected. For example, Nyholm *et al.* successfully estimated the binding energies of hevein with a mono- and di-saccharide using the double decoupling method [94]. Their simulations correctly predict the enthalpy-dominant nature of the hevein-carbohydrate interactions and correctly estimate the binding free energy of the hevein-GlcNAc complex as -2.0 kcal/mol (experimental value -2.0 kcal/mol); however, they overestimated the hevein-(GlcNAc)2 as -5.2 kcal/mol (experimental value of -3.8 kcal/mol). On the other hand, many studies did not provide the accurate absolute values of binding, but provide the relative correct ranking. For example, Naismith *et al.* performed nanosecond-scale molecular dynamics simulations followed by MM-PBSA analysis to explain why a pentasaccharide binds to concanavalin a (lectin) with the same affinity as a trimannoside [95]. They found that MM-PBSA correctly ranks the free energies of binding of a set of protein-carbohydrate complexes [95]. Liang *et al.* studied the different binding free energies of mannose and glucose to the C-type lectin rat mannose binding protein [96]. Their result was qualitatively accurate, but overestimated the experimentally-determined relative binding free energy by a factor of two [96]. Obtaining a correct ranking order, but overestimating the strength of the binding by about a couple fold seems to be a common occurrence in many simulation studies. Pathiaseril and Woods modeled the binding free energies of several mono- and tri-saccharides with monoclonal anti-Salmonella antibody Se155-4 relative to the affinity of the wild-type carbohydrate ligand [7]. The simulations were shown to reasonably reproduce the known geometries of the ligands and the ligand-protein complexes [7]. The predicted values agree closely with experimental values; the difference is less than 1 kcal/mol in most cases. Typically, the error bars of measurement are about 0.2 to 1.0 kcal/mol for both simulation and experimental sides. Kuntz *et al.* explored how structurally similar compounds yield vastly different affinities with neuraminidase [97]. They found that improved positioning of ligand atoms in the active site due to polar and hydrophobic functionalities may be a major factor for increasing ligand selectivity. They also found that more potent ligands make less use of water-mediated interactions in the active site [97]. Though these studies have different focuses, they all provided new insights on sugar recognition by proteins. There is a vast variety of potential applications and carbohydrate recognition systems ready for researchers to perform computer simulation studies. With more systems being studied, the methods used will be further calibrated and automated. They will provide better simulation and calculation ability as carbohydrate, protein and solvent force fields are refined and simulation and analysis algorithms are improved.

6. Cross-Talk Between Recognition Sites

As stated, carbohydrate-interacting proteins, such as lectins, are known to have multiple recognition sites. One can speculate the advantage of such designs, a simple answer is that multiple binding sites increase the chances of binding and, thus, strengthen the interaction. Besides the simple additive effects of having multiple binding sites, the concept of cooperative (allosteric) binding has been raised in various sensing systems [98–101]. The so-called allosteric binding refers to the concept that the binding strength of one binding site can be positively (or negatively) influenced by the binding status of another distant binding site. Indeed, it is well known that such a cooperative effect can make the signal sharper, *i.e.*, the dosage-response curves show a sigmoidal shape and have a shorter transition area between the "on" and "off" states. This is especially important for the sugar-lectin-type of interaction, as the binding affinity is usually relatively weak at 2–4 kcal/mol [102].

Despite its importance in sensing and signaling systems, allostery has not been emphasized by traditional methods. There are two potential challenges in detecting the level of allostery. For one, researchers traditionally judge whether there is allostery by determining whether there are two conformations, such as the R versus T states of proteins. However, capturing two or even more different conformations can be difficult, as the structural changes that most allosteric effects involve are quite subtle. Furthermore, the ensemble view of allosteric changes was only presented in recent years, which emphasizes that fluctuation effects alone can induce allosteric changes without the need to alter the ground state or mean structure level [103]. Another problem is model (parameter) fitting. For example, for a simple two-binding site system, researchers often assume one of the two following scenarios to be correct: (1) two binding sites have intrinsically identical binding properties, but have a (positive or negative) cooperative binding between them; or (2) the two binding events are independent with respect to each other, and one site is intrinsically stronger than the other. Both Model (1) and (2) require only two parameters [104,105]. However, a more complete model is that two sites have intrinsically different binding strengths, but also have a certain level of cooperativity. Thus, a total of four parameters are required to fully characterize such situations. Often, traditional models opt to fit a simple model with fewer parameters due to the limited availability of data, while computational methods can calculate the binding free energy of various combinations of binding states [106].

Here, we use the binding between a lectin (the chain B of ricin) and lactose (the terminal disaccharide of a cell-surface polysaccharide) as an example to illustrate how computer simulation is able to capture the cooperativity between different binding sites. The ricin toxin B chain folds into two β-trefoil structures and contains potentially six sugar recognition sites. Practically, only two of them, sites 1α and 2γ, have retained binding ability. There have been contradictory experimental observations regarding the binding of these two sites, as summarized in our previous report [107]. Thus, we studied this protein-carbohydrate complex through all-atom simulation. We calculated binding free energy between the ricin B chain (RTB) and lactose using the MM-GBSA method and reported that specific residues interact with the saccharide via hydrogen bonding [106]. However, beyond the standard free energy calculation, we also found that there is a level of positive allosteric effect between these two binding sites. Indeed, when both ligands are bound, the interactions between carbohydrates and RTB are stronger than the case with only one ligand bound. To go beyond qualitatively reporting the allosteric binding

and to reveal the actual allosteric interactions between the two binding sites, we have analyzed our early simulation data with a new method called CAMERRA (computation of allosteric mechanism by evaluating residue-residue associations), which we have previously applied to study another allosteric system, the nuclear receptor complex of RXR:TR [107]. As displayed in Figure 2, we show the positive cooperativity between the binding events of lactose to sites 1α and 2γ via CAMERRA. We displayed the top two principal components, which can be viewed as a vibrational mode for residue-residue contacts. Here, contact dynamics (breaking and forming of the interactions between amino acid residues) are shown. A positive value indicates contact forming, while a negative value indicates breaking. One can see that PC1 does not show much cross talk between the two binding sites, while PC2 shows overall positive allostery between the contact events, *i.e.*, when lactose increases contact with site 1α, the other lactose increases the contact with site 2γ. Particularly, we found that the interaction between Gln199 and lactose at 2γ was correlated with stronger binding at site 1α by residues Arg27, Asn44, Gly26 and Gly115. The network of residues involved in the crosstalk between the binding sites is visualized with colored cylinders between amino acid residues. Future engineering and design of lectins with multiple binding sites may explore such cross-talk features and enhance the ability of sugar recognition.
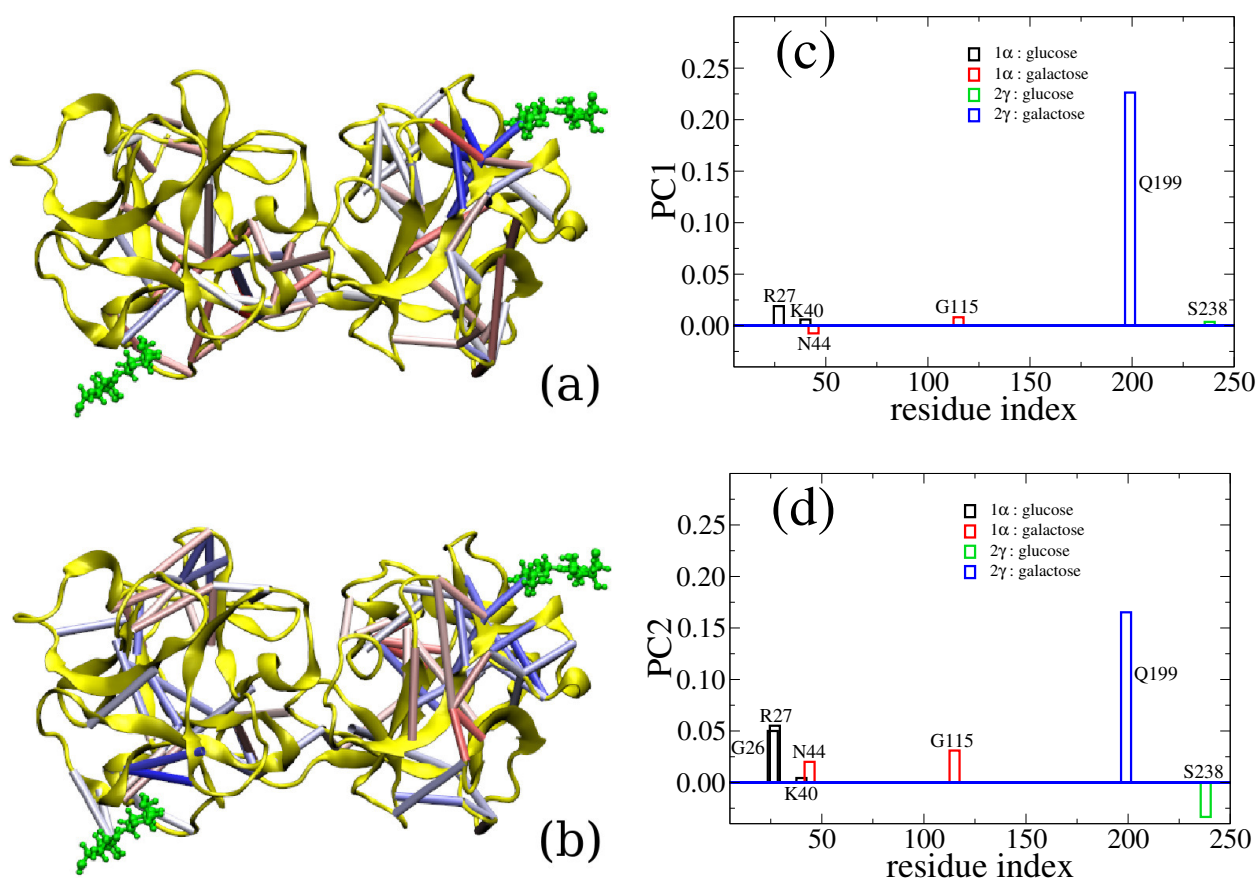

Figure 2. Three-dimensional representations of PC1 and PC2 values are shown in (**a**) and (**b**), respectively. For clarity, only those contacts with absolute values of >0.05 are explicitly displayed as colored cylinders. Elements in eigenvectors PC1 and PC2 that contain the contact interaction between ligands and the protein complex are show in (**c**) and (**d**), respectively.

7. Conclusions

In this work, we reviewed the common schemes used by proteins for carbohydrate recognition and the potential challenges faced in studying the recognition mechanism. We presented how computational methods can assist in the quantification of sugar recognition and, especially, how those methods can circumvent the traditional problems (purification, specificity, and cooperativity) faced. Furthermore, different methods used to calculate the binding free energy between protein and carbohydrate were compared. Successful examples of a binding study via computer simulation were given to demonstrate that this has become a mature technique, despite the existing imperfections, which are being improved.

Author Contributions

The simulations of ricin-lactose interaction, subsequent analysis of trajectories and interpretation of data were performed by Q.R.J., R.J.L., and T.S. Survey of previous computer simulations on protein-carbohydrate interactions and preparation of the manuscript was performed by Q.R.J., R.J.L., L.P., and T.S.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Taylor, M.E.; Drickamer, K. *Introduction to Glycobiology*, 3rd ed.; Oxford University Press: New York, NY, USA, 2011.
2. Varki, A.; Cummings, R.D.; Esko, J.D.; Freeze, H.H.; Stanley, P.; Bertozzi, C.R.; Hart, G.W.; Etzler, M.E. *Essentials of Glycobiology*, 2nd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2009.
3. Sphyris, N.; Lord, J.M.; Wales, R.; Roberts, L.M. Mutational analysis of the Ricinus lectin B-chains galactose-binding ability of the 2γ subdomain of Ricinus communis agglutinin B-chain. *J. Biol. Chem.* **1995**, *270*, 20292–20297.

4. Kwong, P.D.; Doyle, M.L.; Casper, J.D.; Cicala, C.; Leavitt, S.A.; Majeed, S.; Steenbeke, T.D.; Venturi, M.; Chaiken, I.; Fung, M.; *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **2002**, *420*, 678–682.

5. Kirschner, K.N.; Yongye, A.B.; Tschampel, S.M.; Gonzalez-outeirino, J.; Daniels, C.R.; Foley, B.L.; Woods, R.J. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2007**, *29*, 622–655.

6. DeMarco, M.L.; Woods, R.J. Structural Glycobiology: A Game of Snakes and Ladders. *Glycobiology* **2008**, *18*, 426–440.

7. Pathiaseril, A.; Woods, R.J. Relative Energies of Binding for Antibody-Carbohydrate-Antigen Complexes Computed from Free-Energy Simulations. *J. Am. Chem. Soc.* **2000**, *122*, 331–338.

8. Zhang, Q. Molecular Dynamics Simulaitons of Biomimetic Carbohydrate Materials. Ph.D. Thesis, Royal Institute of Technology, Stockholm, Sweden, 2011.

9. Moscona, A. Neuraminidase Inhibitors for Influenza. *N. Engl. J. Med.* **2005**, *353*, 1363–1373.

10. Akkouh, O.; Ng, T.B.; Singh, S.S.; Yin, C.; Dan, X.; Chan, Y.S.; Pan, W.; Cheung, R.C.F. Lectins with Anti-HIV Activity: A Review. *Molecules* **2015**, *20*, 648–668.

11. Francois, B.; Russell, R.J.M.; Murray, J.B.; Aboul-ela, F.; Masquida, B.; Vicens, Q.; Westhof, E. Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding. *Nucleic Acids Res.* **2005**, *33*, 5677–5690.

12. Michel, M.; Mayer, R.; Roche, A. Sugar-lectin interactions: Sugar clusters, lectin multivalency and avidity. *Carbohydr. Lett.* **2000**, *4*, 35–52.

13. Drickamer, K.; Taylor, M.E. Biology of Animal Lectins. *Annu. Rev. Cell Biol.* **1993**, *9*, 237–264.

14. Lee, R.T.; Lee, Y.C. Affinity enhancement by multivalent lectin-carbohydrate interaction. *Glycoconj. J.* **2000**, *17*, 543–551.

15. Rini, J.M. Lectin Structure. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 551–577.

16. Weis, W.I.; Drickamer, K. Structural Basis of Lectin-Carbohydrate Recognition. *Annu. Rev. Biochem.* **1996**, *65*, 441–473.

17. Sharon, N.; Lis, H. History of lectins: From hemagglutinins to biological recognition molecules. *Glycobiology* **2004**, *14*, 53R–62R.

18. Zelensky, A.N.; Gready, J.E. The C-type lectin-like domain superfamily. *FEBS J.* **2005**, *272*, 6179–6217.

19. Drickamer, K. C-type lectin-like domains. *Curr. Opin. Struct. Biol.* **1999**, *9*, 585–590.

20. Loris, R. Principles of structures of animal and plant lectins. *Biochim. Biophys. Acta* **2002**, *1572*, 198–208.

21. Boraston, A.B.; Bolam, D.N.; Gilbert, H.J.; Davies, G.J. Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochem. J.* **2004**, *382*, 769–781.

22. Shoseyov, O.; Shani, Z.; Levy, I. Carbohydrate Binding Modules: Biochemical Properties and Novel Applications. *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 283–295.

23. Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Res.* **2009**, *37*, D233–D238.

24. Fujimoto, Z. Structure and Function of Carbohydrate-Binding Module Families 13 and 42 of Glycoside Hydrolases, Comprising a β-Trefoil Fold. *Biosci. Biotechnol. Biochem.* **2013**, *77*, 1363–1371.

25. Lutteke, T.; Bohne-Lang, A.; Loss, A.; Goetz, T.; Frank, M.; von der Lieth, C.W. GLYCOSCIENCES.de: An Internet portal to support glycomics and glycobiology research. *Glycobiology* **2006**, *16*, 71R–81R.

26. Perez, S.; Sarkar, A.; Breton, C.; Drouillard, S.; Rivet, A.; Imberty, A. Glyco3D: A Portal for Structural Glycoscience; 2013. *Methods Mol. Biol.* **2015**, *1273*, 241–258.

27. Lutteke, T.; Frank, M.; von der Lieth, C. Carbohydrate Structure Suite (CSS): Analysis of carbohydrate 3D structures derived from the PDB. *Nucleic Acids Res.* **2005**, *33*, D242–D246.

28. Jo, S.; Im, W. Glycan fragment database: A database of PDB-based glycan 3D structures. *Nucleic Acids Res.* **2013**, *41*, D470–D474.

29. Kuttel, M.; Mao, Y.; Widmalm, G.; Lundborg, M. CarbBuilder: An Adjustable Tool for Building 3D Molecular Structures of Carbohydrates for Molecular Simulation. In Proceedings of the 2011 IEEE 7th International Conference on E-Science (e-Science), Stockholm, Sweden, 5–8 December 2011; pp. 395–402.

30. Quiocho, F.A. Carbohydrate-Binding Proteins: Tertiary Structures and Protein-Sugar Interactions. *Annu. Rev. Biochem.* **1986**, *55*, 287–315.

31. Marechal, Y. *The Hydrogen Bond and the Water Molecule: The Physics and Chemistry of Water, Aqueous and Bio-Media*; Elsevier: Amsterdam, The Netherlands, 2007.

32. Sheu, S.Y.; Yang, D.Y.; Selzle, H.L.; Schlag, E.W. Energetics of hydrogen bonds in peptides. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12683–12687.

33. Asensio, J.L.; Arda, A.; Canada, F.J.; Jimenez-Barbero, J. Carbohydrate-Aromatic Interactions. *Acc. Chem. Res.* **2013**, *46*, 946–954.

34. Laughrey, Z.R.; Kiehna, S.E.; Riemen, A.J.; Waters, M.L. Carbohydrate-pi interactions: What are they worth? *J. Am. Chem. Soc.* **2008**, *130*, 14625–14633.

35. Faller, C.E.; Guvench, O. Terminal sialic acids on CD44 N-glycans can block hyaluronan binding by forming competing intramolecular contacts with arginine side chains. *Proteins Struct. Funct. Bioinform.* **2014**, *82*, 3079–3089.

36. Angata, T.; Hayakawa, T.; Yamanaka, M.; Varki, A.; Nakamura, M. Discovery of Siglec-14, a novel sialic acid receptor undergoing concerted evolution with Siglec-5 in primates. *FASEB J.* **2006**, *20*, 1964–1973.

37. Varki, A. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature* **2007**, *446*, 1023–1029.

38. Stencel-Baerenwald, J.E.; Reiss, K.; Reiter, D.M.; Stehle, T.; Dermody, T.S. The sweet spot: Defining virus-sialic acid interactions. *Natl. Rev. Microbiol.* **2014**, *12*, 739–749.

39. Vinson, M.; van der Merwe, P.A.; Kelm, S.; May, A.; Jones, E.Y.; Crocker, P.R. Characterization of the Sialic Acid-binding Site in Sialoadhesin by Site-directed Mutagenesis. *J. Biol. Chem.* **1996**, *271*, 9267–9272.

40. Jamal-Talabani, S.; Boraston, A.B.; Turkenburg, J.P.; Tarbouriech, N.; Ducros, V.M.A.; Davies, G.J. Ab Initio Structure Determination and Functional Characterization Of CBM36: A New Family of Calcium-Dependent Carbohydrate Binding Modules. *Structure* **2004**, *12*, 1177–1187.

41. Elcock, A.H.; Sept, D.; McCammon, J.A. Computer Simulation of Protein-Protein Interactions. *J. Phys. Chem. B* **2001**, *105*, 1504–1518.

42. Lehtio, J.; Sugiyama, J.; Gustavsson, M.; Fransson, L.; Linder, M.; Teeri, T.T. The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 484–489.

43. Beckham, G.T.; Matthews, J.F.; Bomble, Y.J.; Bu, L.; Adney, W.S.; Himmel, M.E.; Nimlos, M.R.; Crowley, M.F. Identification of Amino Acids Responsible for Processivity in a Family 1 Carbohydrate-Binding Module from a Fungal Cellulase. *J. Phys. Chem. B* **2010**, *114*, 1447–1453.

44. Nimlos, M.R.; Beckham, G.T.; Matthews, J.F.; Bu, L.; Himmel, M.E.; Crowley, M.F. Binding Preferences, Surface Attachment, Diffusivity, and Orientation of a Family 1 Carbohydrate-binding Module on Cellulose. *J. Biol. Chem.* **2012**, *287*, 20603–20612.

45. Karplus, M.; Petsko, G.A. Molecular-dynamics simulations in biology. *Nature* **1990**, *347*, 631–639.

46. Hong, L.; Petridis, L.; Smith, J.C. Biomolecular Structure and Dynamics with Neutrons: The View from Simulation. *Isr. J. Chem.* **2014**, *54*, 1264–1273.

47. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W.H. Freeman: New York, NY, USA, 1999.

48. Tinoco, I.; Sauer, K.; Wang, J.C.; Puglisi, J.D.; Harbison, G.; Rovnyak, D. *Physical Chemistry: Principles and Applications in Biological Sciences*; Prentice Hall: Upper Saddle River, NJ, USA, 2013.

49. Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Shaw, D.E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520.

50. Knott, B.; Crowley, M.; Himmel, M.; Stahlberg, J.; Beckham, G. Carbohydrate-Protein Interactions That Drive Processive Polysaccharide Translocation in Enzymes Revealed from a Computational Study of Cellobiohydrolase Processivity. *J. Am. Chem. Soc.* **2014**, *136*, 8810–8819.

51. Yang, M.; Huang, J.; MacKerell, A. Enhanced Conformational Sampling of Carbohydrates using Biasing Potential and Solute Tempering Replica Exchange: Application to the N-glycan on the HIV gp120 Envelope Protein. *Biophys. J.* **2015**, *108*, 157a.

52. McCammon, J.A.; Harvey, S.C. *Dynamics of Proteins and Nucleic Acids*; Press Syndicate of the University of Cambridge: New York, NY, USA, 1987.

53. McCammon, J.A.; Gelin, B.R.; Karplus, M. Dynamics of folded proteins. *Nature* **1977**, *267*, 585–590.

54. Mackerell, A.D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584–1604.

55. Guvench, O.; Greene, S.N.; Kamath, G.; Brady, J.W.; Venable, R.M.; Pastor, R.W.; Mackerell, A.D., Jr. Additive Empirical Force Field for Hexopyranose Monosaccharides. *J. Comput. Chem.* **2008**, *29*, 2543–2564.

56. Guvench, O.; Hatcher, E.; Venable, R.M.; Pastor, R.W.; Mackerell, A.D., Jr. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5*, 2353–2370.

57. Guvench, O.; Mallajosyula, S.S.; Raman, E.P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T.J.; Jamison, F.W.; MacKerell, A.D. CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate-Protein Modeling. *J. Chem. Theory Comput.* **2011**, *7*, 3162–3180.

58. Kirschner, K.N.; Lins, R.D.; Maass, A.; Soares, T.A. A Glycam-Based Force Field for Simulations of Lipopolysaccharide Membranes: Parametrization and Validation. *J. Chem. Theory Comput.* **2012**, *8*, 4719–4731.

59. Scott, W.R.P.; Hunenberger, P.H.; Tironi, I.G.; Mark, A.E.; Billeter, S.R.; Fennen, J.; Torda, A.E.; Huber, T.; Kruger, P.; van Gunsteren, W.F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.

60. Jorgensen, W.L.; Maxwell, D.S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

61. Xiong, X.; Chen, Z.; Cossins, B.; Shi, J. Force fields and scoring functions for carbohydrate simulation. *Carbohydr. Res.* **2015**, *401*, 73–81.

62. Foley, B.L.; Tessier, M.B.; Woods, R.J. Carbohydrate force fields. *WIRES Comput. Mol. Sci.* **2012**, *2*, 652–697.

63. Lopez, C.A.; Rzepiela, A.J.; de Vries, A.H.; Dijkhuizen, L.; Hunenberger, P.H.; Marrink, S.J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210.

64. Lopez, C.A.; Bellesia, G.; Redondo, A.; Langan, P.; Chundawat, S.P.S.; Dale, B.E.; Marrink, S.J.; Gnanakaran, S. MARTINI Coarse-Grained Model for Crystalline Cellulose Microfibers. *J. Phys. Chem. B* **2015**, *119*, 465–473.

65. Mishra, S.K.; Kara, M.; Zacharias, M.; Koca, J. Enhanced conformational sampling of carbohydrates by Hamiltonian replica-exchange simulation. *Glycobiology* **2014**, *24*, 70–84.

66. Shen, T.; Langan, P.; French, A.; Johnson, G.P.; Gnanakaran, S. Conformational Flexibility of Soluble Cellulose Oligomers: Chain Length and Temperature Dependence. *J. Am. Chem. Soc.* **2009**, *131*, 14786–14794.

67. Hatcher, E.R.; Guvench, O.; MacKerell, A.D. CHARMM Additive All-Atom Force Field for Acyclic Polyalcohols, Acyclic Carbohydrates, and Inositol. *J. Chem. Theory Comput.* **2009**, *5*, 1315–1327.

68. Nivedha, A.K.; Makeneni, S.; Foley, B.L.; Tessier, M.B.; Woods, R.J. Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *J. Comput. Chem.* **2014**, *35*, 526–539.

69. Hoops, S.C.; Anderson, K.W.; Merz, K.M. Force field design for metalloproteins. *J. Am. Chem. Soc.* **1991**, *113*, 8262–8270.

70. Xantheas, S. Cooperativity and hydrogen bonding network in water clusters. *Chem. Phys.* **2000**, *258*, 225–231.

71. Vanommeslaeghe, K.; MacKerell, A.D., Jr. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta* **2015**, *1850*, 861–871.

72. Lamoureux, G.; Roux, B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003**, *119*, 3025–3039.

73. Patel, D.; He, X.; MacKerell, A.J. Polarizable Empirical Force Field for Hexopyranose Monosaccharides Based on the Classical Drude Oscillator. *J. Phys. Chem. B* **2015**, *119*, 637–652.

74. Case, D.A.; Darden, T.A.; Cheatham, T.E., III; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Crowly, M.; Walker, R.C.; Zhang, W.; *et al*. *AMBER Molecular Dymnamics Package*; AMBER 10 University of California: San Francisco, CA, USA, 2008.

75. Brooks, B.R.; Brooks, C.L., III; Mackerell, A.D., Jr.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; *et al*. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

76. Johnson, Q.R.; Lindsay, R.J.; Raval, S.R.; Dobbs, J.S.; Nellas, R.B.; Shen, T. Effects of Branched O-Glycosylation on a Semiflexible Peptide Linker. *J. Phys. Chem. B* **2014**, *118*, 2050–2057.

77. Meynier, C.; Guerlesquin, F.; Roche, P. Computational Studies of Human Galectin-1: Role of Conserved Tryptophan Residue in Stacking Interaction with Carbohydrate Ligands. *J. Biomol. Struct. Dyn.* **2009**, *27*, 49–57.

78. Ford, M.G.; Weimar, T.; Kohli, T.; Woods, R.J. Molecular dynamics simulations of galectin-1-oligosaccharide complexes reveal the molecular basis for ligand diversity. *Proteins Struct. Funct. Bioinf.* **2003**, *53*, 229–240.

79. Butenhof, K.J.; Gerken, T.A. Structure and dynamics of mucin-like glycopeptides. Examination of peptide chain expansion and peptide-carbohydrate interactions by stochastic dynamics simulations. *Biochemistry* **1993**, *32*, 2650–2663.

80. Mandal, T.K.; Mukhopadhyay, C. Effect of glycosylation on structure and dynamics of MHC class I glycoprotein: A molecular dynamics study. *Biopolymers* **2001**, *59*, 11–23.

81. Mark, P.; Zhang, Q.; Czjzek, M.; Brumer, H.; Agren, H. Molecular dynamics simulations of a branched tetradecasaccharide substrate in the active site of a xyloglucan endo-transglycosylase. *Mol. Simul.* **2011**, *37*, 1001–1013.

82. Favreau, A.; Faller, C.E.; Guvench, O. CD44 Receptor Unfolding Enhances Binding by Freeing Basic Amino Acids to Contact Carbohydrate Ligand. *Biophys. J.* **2013**, *105*, 1217–1226.

83. Plazinski, W.; Knys-Dzieciuch, A. The 'order-to-disorder' conformational transition in CD44 protein: An umbrella sampling analysis. *J. Mol. Graph. Model.* **2013**, *45*, 122–127.

84. Jamison, F.W. II; Foster, T.J.; Barker, J.A.; Hills, R.D., Jr.; Guvench, O. Mechanism of Binding Site Conformational Switching in the CD44-Hyaluronan Protein-Carbohydrate Binding Interaction. *J. Mol. Biol.* **2011**, *406*, 631–647.

85. Chang, C.E.; Gilson, M.K. Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *J. Am. Chem. Soc.* **2004**, *126*, 13156–13164.

86. Jana, M.; Bandyopadhyay, S. Conformational flexibility of a protein-carbohydrate complex and the structure and ordering of surrounding water. *Phys. Chem. Chem. Phys.* **2012**, *14*, 6628–6638.

87. Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; *et al.* Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

88. Qiu, D.; Shenkin, P.S.; Hollinger, F.P.; Still, W.C. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.

89. David, L.; Luo, R.; Gilson, M. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comput. Chem.* **2000**, *21*, 295–309.

90. Shen, T.; Wong, C.F.; McCammon, J.A. Brownian dynamics simulation of helix-capping motif. *Biopolymers* **2003**, *70*, 252–259.

91. Gilson, M.K.; Given, J.A.; Bush, B.L.; McCammon, J.A. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.* **1997**, *72*, 1047–1069.

92. Hamelberg, D.; McCammon, A. Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.

93. McCammon, J.A.; Lybrand, T.P.; Allison, S.A.; Northrup, S.H. Ligand binding: New theoretical approaches to molecular recognition. In *Biomolecular Stereodynamics*; Adenine Press: New York, NY, USA, 1986; Volume 3.

94. Koppisetty, C.; Frank, M.; Lyubartsev, A.; Nyholm, P. Binding energy calculations for hevein-carbohydrate interactions using. *J. Comput. Aided Mol. Des.* **2015**, *29*, 13–21.

95. Bryce, R.; Hillier, I.; Naismith, J. Carbohydrate-Protein Recognition: Molecular Dynamics Simulations and Free Energy Analysis of Oligosaccharide Binding to Concanavalin A. *Biophys. J.* **2001**, *81*, 1373–1388.

96. Liang, G.; Schmidt, R.; Yu, H.; Cumming, D.; Brady, J. Free Energy Simulation Studies of the Binding Specificity of Mannose-Binding Protein. *J. Phys. Chem.* **1996**, *100*, 2528–2534.

97. Masukawa, K.; Kollman, P.; Kuntz, I. Investigation of Neuraminidase-Substrate Recognition Using Molecular Dynamics and Free Energy Calculations. *J. Med. Chem.* **2003**, *46*, 5628–5637.

98. Pauling, L. The Oxygen Equilibrium of Hemoglobin and Its Structural Interpretation. *Proc. Natl. Acad. Sci. USA* **1935**, *21*, 186–191.

99. Hill, T.L. *Cooperativity Theory in Biochemistry: Steady-State and Equilibrium Systems*, 1st ed.; Springer: New York, NY, USA, 2011.

100. Changeux, J.P.; Edelstein, S.J. Allosteric Mechanisms of Signal Transduction. *Science* **2005**, *308*, 1424–1428.

101. Cui, Q.; Karplus, M. Allostery and cooperativity revisited. *Protein Sci.* **2008**, *17*, 1295–1307.

102. Williams, B.A.; Chervenak, M.C.; Toone, E.J. Energetics of lectin-carbohydrate binding. A microcalorimetric investigation of concanavalin A-oligomannoside complexation. *J. Biol. Chem.* **1992**, *267*, 22907–22911.

103. Popovych, N.; Sun, S.; Ebright, R.H.; Kalodimos, C.G. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* **2006**, *13*, 831–838.

104. Zentz, C.; Frenoy, J.P.; Bourrillon, R. Binding of galactose and lactose to ricin. *Biochem. Biophys. Acta* **1978**, *536*, 18–26.

105. Adair, W.L.; Kornfeld, S. Isolation of the Receptors for Wheat Germ Agglutinin and the Ricinus communis Lectins from Human Erythrocytes Using Affinity Chromatography. *J. Biol. Chem.* **1974**, *249*, 4696–4704.

106. Yao, J.; Nellas, R.B.; Glover, M.M.; Shen, T. Stability and Sugar Recognition Ability of Ricin-like Carbohydrate Binding Domains. *Biochemistry* **2011**, *50*, 4097–4104.

107. Johnson, Q.R.; Lindsay, R.J.; Nellas, R.B.; Fernandez, E.J.; Shen, T. Mapping Allostery through Computational Glycine Scanning and Correlation Analysis of Residue-Residue Contacts. *Biochemistry* **2015**, *54*, 1534–1541.