Communications in Information and Systems
Volume 21, Number 1, 147–163, 2021

Carbohydrate-protein interactions: advances and challenges

Shuang Zhang, Kyle Yu Chen, and Xiaoqin Zou*

A carbohydrate, also called saccharide in biochemistry, is a biomolecule consisting of carbon (C), hydrogen (H) and oxygen (O) atoms. For example, sugars are low molecular-weight carbohydrates, and starches are high molecular-weight carbohydrates. Carbohydrates are the most abundant organic substances in nature and essential constituents of all living things. Protein-carbohydrate interactions play important roles in many biological processes, such as cell growth, differentiation, and aggregation. They also have broad applications in pharmaceutical drug design. In this review, we will summarize the characteristic features of protein-carbohydrate interactions and review the computational methods for structure prediction, energy calculations, and kinetic studies of protein-carbohydrate complexes. Finally, we will discuss the challenges in this field.

Keywords and phrases: Carbohydrate-protein interactions, saccharide-protein interactions, sugar-protein binding, structure prediction, molecular docking, drug discovery.

1. Introduction

Protein-carbohydrate interactions, also called protein-saccharide interactions, are gaining increasing popularity due to their fundamental roles in numerous aspects of biology and food industry. Saccharide is a synonym of carbohydrate, which is a biomolecule consisting of carbon, hydrogen, and oxygen atoms. Carbohydrates are one of the most abundant materials on earth. Based on the different lengths of repeating units, carbohydrates (i.e., saccharides) are divided into four chemical groups: monosaccharides, disaccharides, oligosaccharides and polysaccharides, as shown in Figure 1.

Carbohydrates have a wide variety of functions, one of which is to serve as energy storage molecules [1]. For example, starch functions as a storage for the energy from the photosynthetic process in plants. In animals, glycogen is the counterpart form of energy storage [2]. Carbohydrates also function as

*Correspondence to: Xiaoqin Zou, Email: zoux@missouri.edu.

structural components. Cellulose, considered the most common organic compound in the world, provides the cell with structural support. The second most common polysaccharide in the world after cellulose, chitin, also provides animals with similar functions. Carbohydrates serve as a recognition element as well. Glyconjugates, which are carbohydrates on cell surfaces that are bonded to other chemical species such as proteins, peptides and lipids, play a role in cell-cell interactions. One class of proteins called lectins, which are ubiquitous in nature and are found in animals, plants and microorganisms, can bind carbohydrates with high specificity. Such an interaction constitutes the basis for lectin-mediated drug targeting [3, 4] and drug delivery [5]. In addition, carbohydrates mediate biological functions by interacting with enzymes and antibodies. Therefore, protein-carbohydrate interactions are essential to immune systems, fertilization, and blood clotting [6].

This review will present the features of protein-carbohydrate interactions and summarize the computational methods for predicting protein-carbohydrate interactions. Challenges in this field will be discussed at the end.

2. Characterization of protein-carbohydrate interactions

Experimentally, X-ray diffraction and nuclear magnetic resonance (NMR) spectroscopy have been employed to determine the geometry of protein-carbohydrate complexes at the atomic level. X-ray crystallography provides static pictures of stabilized complexes, whereas NMR in combination with computational studies reveals dynamic structural behaviors of the carbohydrate binding process [7].Unlike a protein, whose primary structure is linear, carbohydrates vary from linear to highly branched structures, as illustrated in Figure 1c. High flexibility is thus the major obstacle to obtain high-resolution crystal structures of carbohydrates. Therefore, despite the huge successes of X-ray crystallography in solving protein structures, it is difficult to determine the structures of carbohydrates with X-ray crystallography directly. NMR techniques are often used to study carbohydrates in solutions. Particularly, solid-state NMR (ssNMR) is a useful tool to observe the behaviors of different saccharides in their intact chemical forms bound to receptors. For instance, the structure and dynamics of Escherichia coli peptidoglycan, an essential component of the bacterial cell wall, was investigated by using ssNMR [8]. Because NMR data are the average of a conformational ensemble, which may lead to a virtual conformation, NMR should be used with caution when determining 3D carbohydrate structures

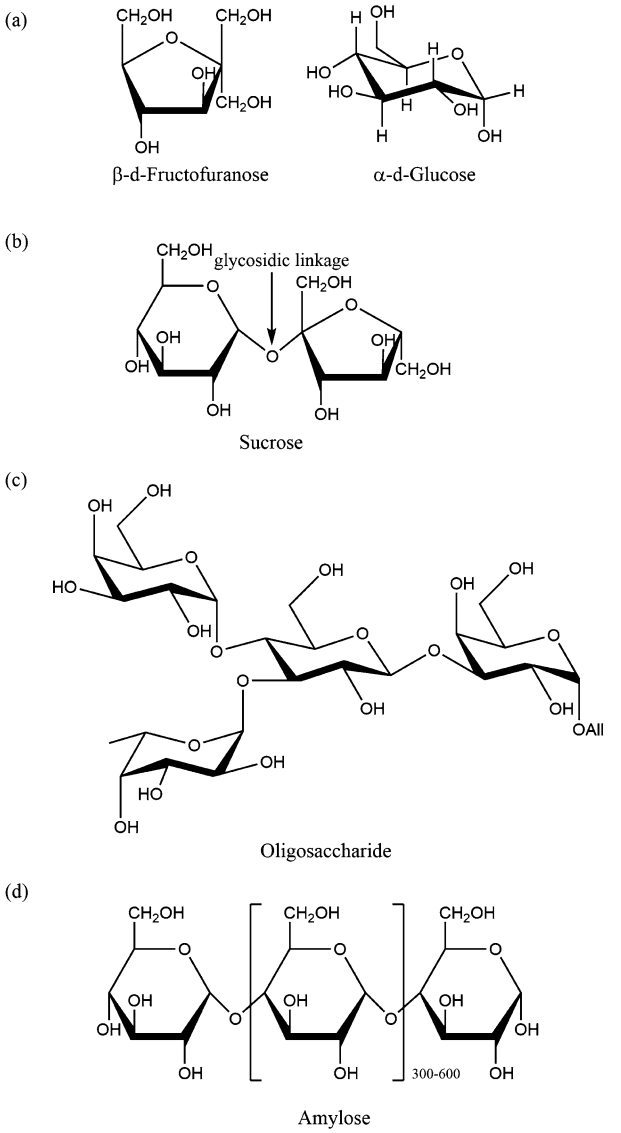

Figure 1: Four types of saccharides. (a) Two typical monosaccharides, β-d-Fructofuranose and α-d-Glucose. Monosaccharides with five-membered rings are called furanose, and monosaccharides six-membered rings are called pyranose. (b) Disaccharide is formed when two monosaccharides are joined by a carbon-oxygen-carbon linkage known as glycosidic linkage. Sucrose is a common disaccharide composed of glucose and fructose. (c) An oligosaccharide is a saccharide polymer having three to ten monosaccharides. (d) Polysaccharides consist of (usually many) monosaccharides. Amylose is a polysaccharide made of many α-d-Glucose units.

[9]. Recently, cryo-EM has been used to solve the structures of carbohydrate-binding proteins [10, 11, 12, 13]. However, it remains a challenge for cryo-EM to determine carbohydrates structures. Although it is possible to determine the structure of a monosaccharide and its ring conformation at a high resolution, neither the monosaccharide nor its ring conformation can be identified at a resolution lower than 2.4 Å[13].

To understand the formation of protein-carbohydrate structures, much attention has been drawn to studying protein-carbohydrate interactions. Similar to other protein-ligand interactions, protein-carbohydrate interactions are characterized by several biophysical properties such as thermodynamic parameters, dissociation constants and kinetics (e.g., k_{on} and k_{off} rate constants). A number of techniques including isothermal titration calorimetry (ITC) [14], frontal affinity chromatography (FAC) [15], and surface plasmon resonance (SPR) [16] have been used to measure the dissociation constants for protein–sugar recognition. SPR is one of the most widely used techniques for the determination of kinetic parameters such as association/dissociation constants and on/off rates (k_{on}/k_{off}) This method has been successfully applied to exploring the kinetics and mechanisms of lectin-carbohydrate recognition [17]. It should be noted that SPR analysis requires immobilization of the protein and the carbohydrates. Recently, other approaches have been developed to investigate carbohydrate-protein interactions. For example, lectin-carbohydrate interactions were measured using a glycan microarray [18]. Kinetic analysis of the conformational changes in the mutants of maltose-binding proteins (MBPs) was performed using single-molecule fluorescence resonance energy transfer (smFRET) [19]. The binding kinetics of carbohydrate-protein interactions on unfixed cancer cell surfaces was studied using a quartz crystal microbalance (QCM) biosensor [20].

It has been reported that protein-carbohydrate interactions are relatively weak [7, 21, 22, 23]. For example, the dissociation constants for most lectin-monosaccharide interactions are in the mM range [24]. The protein-carbohydrate binding affinity increases with increasing sizes of oligosaccharide [25], except in the assessment of oligomannosyls (consisting 1 to 6 monosaccharide units) for their abilities to inhibit yeast cell adherence [26]. However, the dissociation constants are usually greater than μM [27, 28]. Such weak interactions facilitate the formation of transient states and therefore help partner proteins to reach their appropriate locations [22].

Generally, the carbohydrate-binding sites on the hydrophilic surfaces of lectins and other proteins are shallow [29]. Therefore, the binding affinities between lectins and their carbohydrate ligands are lower than many other

protein-ligand binding affinities. However, like other protein-ligand interactions, the binding specificity of protein-carbohydrate interactions originates from the combination of a variety of interactions, including hydrogen bonding, hydrophobic interactions and electrostatic interactions. Due to the existence of numerous hydroxyl groups in every carbohydrate, hydrogen bonds probably contribute the most to the protein-carbohydrates interactions, especially in lectin-carbohydrate complexes. Hydroxyl groups may act as donors or acceptors. Sometimes, the hydroxyl groups participate in hydrogen bonding as both donors and as acceptors, known as "cooperative hydrogen bonding". The formation of the hydrogen bonds often involves carbohydrate hydroxyl groups contacting with aspartates, glutamates, asparagines, glutamines, arginines, and serines, as well as with backbone amines and carbonyl groups. Bidentate hydrogen bonds are formed when two adjacent hydroxyl groups of a carbohydrate molecule establish hydrogen bonds with both carboxylate oxygens of either aspartates or glutamates. This feature is observed in enzymes, as well as in lectins [30]. Hydrogen bonding interactions are found to be important for selectivity, by using engineered sugars [31] and fluorodeoxy derivatives of natural lectin ligands [32, 33].

In some carbohydrates, the clustering of several adjacent C-H groups arising from the steric disposition of sugar hydroxyl groups creates a hydrophobic surface that can form nonpolar interactions with the aromatic rings of Trp, Tyr, and Phe residues in proteins. Such CH-π stackings between saccharides and aromatic rings in carbohydrate-binding sites have been demonstrated experimentally [34]. Efforts have been made to better understand CH-π interactions [35, 36, 37, 38]. The attractive forces are thought to result from 1) an entropic contribution arising from the mutual shielding of both nonpolar surfaces from bulk water, and 2) the enthalpic contribution of non-conventional hydrogen bonds between the partially positively charged C-H groups and the quadrupole of the π-system of the aromatic ring [39].

Electrostatic interaction is also crucial for carbohydrate-protein recognition. Charged residues and ions are often found in active sites of carbohydrate-binding proteins [40, 41]. Salt bridges are formed between the charged residues of saccharides and protein residues with opposite charges. In some instances, divalent cations, such as calcium and magnesium, bridge the hydroxyls of carbohydrates and negatively charged aspartates/glutamates, facilitating proteins to recognize their carbohydrate ligands [7].

Apart from noncovalent interactions, glycosidic linkages also play a vital role in protein-carbohydrate complexes, as shown in Figure 1 and Figure 2. Although glycosidic torsion angles govern the flexibility of carbohydrates,

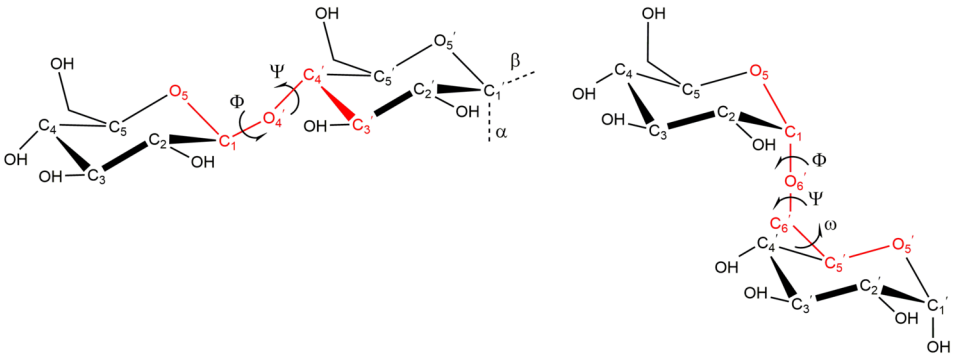

Figure 2: Glycosidic torsion angles of disaccharides. ϕ (O_5 - C_1 - O'_x - C'_x), ψ (C_1 - O'_x - C'_x - C'_{x-1}), ω (O'_6 - C'_6 - C'_5 - O'_5). x is denoted as the index of a carbon atom of the neighboring monosaccharide. The C_1 atom is the anomeric carbon. In a cyclic carbohydrate, the anomeric carbon is the carbonyl carbon in acyclic form. The orientation of the connecting oxygen determines whether the anomer is α or β [43]. If the oxygen is below the plane of ring, the anomer is α. Otherwise, the anomer is β.

it has been reported that torsion angles in glycosidic linkages have angular preferences instead of random distributions [42]. Hence, taking into account glycosidic torsion angles provides further insight into the assessment of protein-carbohydrate complex structures [42, 43].

3. Predicting protein-carbohydrate interactions

Due to the limited number of experimentally solved protein-carbohydrate structures, computational methods provide valuable complementary tools to investigate protein-carbohydrate interactions. One of these methods is molecular docking. Molecular docking is a powerful computational approach that is widely used to predict receptor-ligand complexes [44, 45]. Here, a receptor refers to a protein or a protein oligomer, and a ligand is either a small molecule or another protein. Scoring functions for docking are designed to rank different ligand poses through the evaluation of the strength of the interactions. Over the last two decades, more than 60 different docking programs have been developed and successfully used in predicting protein-ligand structures [46], such as DOCK [47], AutoDock [48], ZDock [49], HADDOCK [50], and MDock [51, 52]. However, most docking programs often lead to unnatural glycosidic angles when applied to carbohydrates [42]. Due to the peculiarity of protein-carbohydrate interactions, a few programs have been

developed for protein-carbohydrate docking [43, 53] to address this challenge.

As aforementioned, the CH-π interaction is important for protein-carbohydrate complex formation. By adding an energy term for the CH-π stacking interactions to the SLICK scoring function [54], BALLDock [53] has been successfully applied to a set of carbohydrate-lectin complexes. The SLICK scoring function is composed of four terms, including a CH-π interaction term, a hydrogen bonding term, a softened van der Waals term, and a Coulombic electrostatic term. The score S is given by

$$S = s_0 + s_{CH\pi}S_{CH\pi} + s_{hb}S_{hb} + s_{vdw}\Delta E_{vdw} + s_{es}\Delta E_{es} \quad (1)$$

while the weighting coefficients s are manually optimized. BALLDock and the SLICK scoring function successfully redocked 17 out of 18 plant lectin complex structures in a calibration set with high accuracy, much better than another docking program FlexX [55]. In a larger test set, 14 out of 22 complexes consisting of plant lectins and non-plant lectins were obtained, with the root-mean-square deviations (rmsd) of the best ranked ligand orientations less than 1.5 Å.

Inclusion of conformational preference provides another opportunity to improve the efficiency and accuracy of carbohydrate docking. Recently, carbohydrate Intrinsic (CHI) energy functions [42]. were developed to estimate the energy arising from distortion of the glycosidic linkages. The CHI energy functions were derived from fitting the Gaussian expansions to average energy curves obtained from quantum mechanical calculations:

$$f(x) = \sum_{i=1}^{N} a_i e^{-(x-b_i)^2/c_i} \quad (2)$$

where f is the CHI energy function, N is the number of individual Gaussian functions, x is the torsion angle (ϕ, ψ, or ω shown in Figure 2), and a_i, b_i, and c_i are the magnitude, width and midpoint of the distribution, respectively. By implementing the CHI-energy functions into the AutoDock Vina scoring function [56], Vina-Carb [43], increased the docking accuracy compared to AutoDock Vina. When Vina-Carb was applied to antibodies, lectins and carbohydrate binding modules (CBM), the success rates of predicting accurate binding modes reached 86%, 50%, 42%, respectively, in comparison with 70%, 50%, and 0% for AutoDock Vina. In addition, inclusion of explicit water during carbohydrate docking can improve docking accuracy [57, 58].

Molecular dynamics (MD) simulation [59] and hybrid method (quantum mechanics/molecular mechanics or QM/MM) are also important computational tools to explore protein-carbohydrate interactions. For example, QM/MM combined with MD simulation was conducted on a norovirus capsid protein bound with a Lewis antigen [60]. In a different report [61], comparative studies on MM/Poisson–Boltzmann surface area (PBSA), MM/generalized Born surface area (GBSA) and QM-MM/GBSA were performed on monosaccharides bound to *Ralstonia solanacearum* lectin (RSL), a six-bladed β-propeller trimeric lectin with 90 amino acid residues in each monomeric chain. The limitations to these approaches are short timescale and inaccuracies in force fields.

It should be noted that PBSA and GBSA are two typical implicit solvent models for computing theoretical binding free energies of carbohydrate-protein complexes. Accurate prediction of binding free energies is essential to understanding biomolecular recognition. Generally, the binding free energy consists of enthalpy-related and entropy-dependent components [62]. Some of these terms can be directly obtained from MD simulations, such as the values for van der Waals and electrostatic energies. Desolvation free energy is usually estimated by using implicit solvent approximation, such as PBSA and GBSA. Comparing to explicit solvent models, implicit models represent solvent molecules as a continuum dielectric media instead of individual "explicit" solvent molecules. This method improves the computational speed significantly. PBSA is based on the Poisson continuum dielectric model and GBSA is an approximation to the PBSA with more computational efficiency but less precision. These two methods have been widely used in the study of lectin binding [63, 64, 65]. Although implicit solvent models simplify the calculation and reproduce the behavior of bulk water, not all water molecules behave like bulk water. Some tightly bound waters play as a part of the protein or ligand [62, 66], and hydrogen-bond networks are important to stabilize protein-carbohydrate complexes, such as *Polyporus squamosus* lectin (PSL) bound with a trisaccharide [67]. These additional effects are omitted when applying implicit solvent models to binding free energy calculations.

Finally, kinetics of carbohydrate-protein binding has also been studied with MD simulations. Recently, Negami et al. [68] employed coarse-grained molecular dynamics (CGMD) simulations with the MARTINI force field [69] to study the levansucrase–carbohydrate system. The binding rate constants were estimated using a Bayesian approach. Solanke et al. studied the hevein domain (HEV32) bound with N-acetylglucosamine mono-, di- or trisaccharides by using unbiased MD simulation [70]. Three of the six

simulations successfully reproduced the correct binding modes in agreement with experimental results. Unbinding events were observed in one simulation (monosaccharide) as well. The dissociation constant was estimated based on these binding/unbinding events.

4. Discussion

In this review, we have briefly summarized the recent advances in the field of protein-carbohydrate interactions. Experimentally, it remains a challenge to determine the structure of a protein-carbohydrate complex, despite the achievements of X-ray diffraction, NMR and cryo-EM. New refinement protocols need to be developed for X-ray and cryo-EM methods [71]. The main drawback of NMR-based studies is its relatively intrinsic low sensitivity. Efforts have been made to overcome this problem, such as using higher magnetic fields [72]. Pertaining to computational methods, molecular docking is most efficiently used to predict the structures of protein-carbohydrate complexes. Vina-Carb is the most successful carbohydrate docking program to date. However, due to missing the CH-π interaction, it fails when predicting the binding modes of carbohydrates that stack against aromatic amino acids. Additionally, pyranose (i.e., saccharides that have a six-membered ring) and furanose (i.e., saccharides with a five-membered ring) commonly have distortions in presence of carbohydrate-active enzymes [73]. Unlike pyranose, furanose rings can adopt different conformations with little difference in their free energies [74], resulting in more ring flexibility [75]. Thus, improvement in the scoring functions for carbohydrate-protein docking is urgently needed.

Acknowledgement

This work was supported by NIH R01GM109980 and R35GM136409 to XZ.

References

[1] W. Pigman, The carbohydrates: chemistry and biochemistry. *Academic Press* (1970).

[2] E. Sadava, M. Hillis and H.C. Heller, Life (9th, International ed.), *WH Freeman* (2011).

[3] M. Wirth, A. Fuchs, M. Wolf, B. Ertl and F. Gabor, Lectin-mediated drug targeting: preparation, binding characteristics, and antiproliferative activity of wheat germ agglutinin conjugated doxorubicin on Caco-2 cells. *Pharm Res* **15** (1998), 1031–1037.

[4] M. Wirth, G. Hamilton and F. Gabora, Lectin-mediated drug targeting: quantification of binding and internalization of wheat germ agglutinin and Solatium tuberosum lectin using Caco-2 and HT-29 cells. *J Drug Target* **6** (1998), 95–104.

[5] M.A. Clark, B.H. Hirst and M.A. Jepson, Lectin-mediated mucosal delivery of drugs and microparticles. *Adv Drug Deliver Rev* **43** (2000), 207–223.

[6] A. Maton, Human biology and health. *Prentice Hall: Englewood Cliffs, N.J.* (1993).

[7] M. del Carmen Fernández-Alonso, D. Díaz, M.Á. Berbis, F. Marcelo, J. Cañada and J. Jiménez-Barbero, Protein-carbohydrate interactions studied by NMR: from molecular recognition to drug design. *Curr Protein Pept Sci* **13** (2012), 816–830.

[8] T. Kern, S. Hediger, P. Müller, C. Giustini, B. Joris, C. Bougault, W. Vollmer and J.-P. Simorre, Toward the characterization of peptidoglycan structure and protein-peptidoglycan interactions by solid-state NMR spectroscopy. *J Am Chem Soc* **130** (2008), 5618–5619.

[9] M.L. DeMarco and R.J. Woods, Structural glycobiology: a game of snakes and ladders. *Glycobiology* **18** (2008), 426–440.

[10] D. Lyumkis, J.-P. Julien, N. de Val, A. Cupo, C.S. Potter, P.-J. Klasse, D.R. Burton, R.W. Sanders, J.P. Moore and B. Carragher, Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* **342** (2013), 1484–1490.

[11] A.C. Walls, M.A. Tortorici, B.-J. Bosch, B. Frenz, P.J. Rottier, F. DiMaio, F.A. Rey and D. Veesler, Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* **531** (2016), 114–117.

[12] A.C. Walls, M.A. Tortorici, B. Frenz, J. Snijder, W. Li, F.A. Rey, F. DiMaio, B.-J. Bosch and D. Veesler, Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat Struct Mol Biol* **23** (2016), 899.

[13] M. Atanasova, H. Bagdonas and J. Agirre, Structural glycobiology in the age of electron cryo-microscopy. *Curr Opin Struct Biol* **62** (2020), 70–78.

[14] M.L. Doyle, Characterization of binding interactions by isothermal titration calorimetry. *Curr Opin Struct Biol* **8** (1997), 31–35.

[15] J. Hirabayashi, Y. Arata and K.-i. Kasai, Reinforcement of frontal affinity chromatography for effective analysis of lectin–oligosaccharide interactions. *J Chromatogr A* **890** (2000), 261–271.

[16] D.A. Mann, M. Kanai, D.J. Maly and L.L. Kiessling, Probing low affinity and multivalent interactions with surface plasmon resonance: ligands for concanavalin A. *J Am Chem Soc* **120** (1998), 10575–10582.

[17] M. Kapoor, C.J. Thomas, K. Bachhawat-Sikder, S. Sharma and A. Surolia, Exploring kinetics and mechanism of protein–sugar recognition by surface plasmon resonance. *Method Enzymol* **362** (2003), 312–329.

[18] P.-H. Liang, S.-K. Wang and C.-H. Wong, Quantitative analysis of carbohydrate-protein interactions using glycan microarrays: determination of surface and solution dissociation constants. *J Am Chem Soc* **129** (2007), 11177–11184.

[19] M.H. Seo, J. Park, E. Kim, S. Hohng and H.S. Kim, Protein conformational dynamics dictate the binding affinity for a ligand. *Nat Commun* **5** (2014), 3724.

[20] X. Li, S. Song, Q. Shuai, Y. Pei, T. Aastrup, Y. Pei and Z. Pei, Real-time and label-free analysis of binding thermodynamics of carbohydrate-protein interactions on unfixed cancer cell surfaces using a QCM biosensor. *Sci Rep* **5** (2015), 14066.

[21] A. Laederach and P.J. Reilly, Modeling protein recognition of carbohydrates. *Proteins Struct Funct Bioinf* **60** (2005), 591–597.

[22] W.S. Somers, J. Tang, G.D. Shaw and R.T. Camphausen, Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P-and E-selectin bound to SLeX and PSGL-1. *Cell* **103** (2000), 467–479.

[23] R. Marchetti, S. Perez, A. Arda, A. Imberty, J. Jimenez-Barbero, A. Silipo and A. Molinaro, "Rules of Engagement" of Protein-Glycoconjugate Interactions: A Molecular View Achievable by using NMR Spectroscopy and Molecular Modeling. *ChemistryOpen* **5** (2016), 274–296.

[24] F.P. Schwarz, K.D. Puri, R. Bhat and A. Surolia, Thermodynamics of monosaccharide binding to concanavalin A, pea (Pisum sativum) lectin, and lentil (Lens culinaris) lectin. *J Biol Chem* **268** (1993), 7668–7677.

[25] W. Zou, R. Mackenzie, L. Thérien, T. Hirama, Q. Yang, M.A. Gidney and H.J. Jennings, Conformational epitope of the type III group B

Streptococcus capsular polysaccharide. *J Immunol* **163** (1999), 820–825.

[26] T. Kanbe and J.E. Cutler, Minimum chemical requirements for adhesin activity of the acid-stable part of Candida albicans cell wall phosphomannoprotein complex. *Infect Immun* **66** (1998), 5812–5818.

[27] R. Ramkumar, A. Surolia and S.K. Podder, Energetics of carbohydrate binding by a 14 kDa S-type mammalian lectin. *Biochem J* **308** (1995), 237–241.

[28] R. Ramkumar and S.K. Podder, Elucidation of the mechanism of interaction of sheep spleen galectin-1 with splenocytes and its role in cell–matrix adhesion. *J Mol Recognit* **13** (2000), 299–309.

[29] C. Taroni, S. Jones and J.M. Thornton, Analysis and prediction of carbohydrate binding sites. *Protein Eng* **13** (2000), 89–98.

[30] K. Piotukh, V. Serra, R. Borriss and A. Planas, Protein-Carbohydrate Interactions Defining Substrate Specificity in Bacillus 1, 3-1, 4-β-d-Glucan 4-Glucanohydrolases as Dissected by Mutatioal Analysis. *Biochemistry* **38** (1999), 16092–16104.

[31] A. Rivera-Sagredo, D. Solís, T. Díaz-Mauriño, J. Jiménez-Barbero and M. Martín-Lomas, Studies on the molecular recognition of synthetic methyl β-lactoside analogs by ricin, a cytotoxic plant lectin. *Eur J Biochem* **197** (1991), 217–228.

[32] D. Solís, P. Fernández, T. Díaz-Mauriño, J. Jiménez-Barbero and M. Martín-Lomas, Hydrogen-bonding pattern of methyl β-lactoside binding to the Ricinus communis lectins. *Eur J Biochem* **214** (1993), 677–683.

[33] D. Solís, J. Jiménez-Barbero, M. Martín-Lomas and T. Díaz-Mauriño, Probing hydrogen-bonding interactions of bovine heart galectin-1 and methyl β-lactoside by use of engineered ligands. *Eur J Biochem* **223** (1994), 107–114.

[34] M. Muraki, M. Morikawa, Y. Jigami and H. Tanaka, The roles of conserved aromatic amino-acid residues in the active site of human lysozyme: a site-specific mutagenesis study. *Bba-Protein Struct M* **916** (1987), 66–75.

[35] M.d.C. Fernández-Alonso, F.J. Cañada, J. Jiménez-Barbero and G. Cuevas, Molecular Recognition of Saccharides by Proteins. Insights on

the Origin of the Carbohydrate-Aromatic Interactions. *J Am Chem Soc* **127** (2005), 7379–7386.

[36] S. Vandenbussche, D. Díaz, M.C. Fernández-Alonso, W. Pan, S.P. Vincent, G. Cuevas, F.J. Cañada, J. Jiménez-Barbero and K. Bartik, Aromatic–carbohydrate interactions: an NMR and computational study of model systems. *Chem Eur J* **14** (2008), 7570–7578.

[37] K. Ramirez-Gualito, R. Alonso-Rios, B. Quiroz-Garcia, A. Rojas-Aguilar, D. Díaz, J. Jiménez-Barbero and G. Cuevas, Enthalpic nature of the CH/π interaction involved in the recognition of carbohydrates by aromatic compounds, confirmed by a novel interplay of NMR, calorimetry, and theoretical calculations. *J Am Chem Soc* **131** (2009), 18129–18138.

[38] K.L. Hudson, G.J. Bartlett, R.C. Diehl, J. Agirre, T. Gallagher, L.L. Kiessling and D.N. Woolfson, Carbohydrate-Aromatic Interactions in Proteins. *J Am Chem Soc* **137** (2015), 15152–15160.

[39] J. Jiménez-Barbero, F. Javier Cañada, J.L. Asensio, N. Aboitiz, P. Vidal, A. Canales, P. Groves, H.-J. Gabius and H.-C. Siebert, Hevein Domains: An Attractive Model to Study Carbohydrate–Protein Interactions at Atomic Resolution. *Advances in Carbohydrate Chemistry and Biochemistry* **60** (2006), 303–354.

[40] F.A. Quiocho, Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. *Annu Rev Biochem* **55** (1986), 287–315.

[41] A.B. Boraston, D.N. Bolam, H.J. Gilbert and G.J. Davies, Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382** (2004), 769–781.

[42] A.K. Nivedha, S. Makeneni, B.L. Foley, M.B. Tessier and R.J. Woods, Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *J Comput Chem* **35** (2014), 529–539.

[43] A.K. Nivedha, D.F. Thieker, S. Makeneni, H. Hu and R.J. Woods, Vina-Carb: improving glycosidic angles during carbohydrate docking. *J Chem Theory Comput* **12** (2016), 892–901.

[44] D.B. Kitchen, H. Decornez, J.R. Furr and J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* **3** (2004), 935–949.

[45] I.A. Guedes, C.S. de Magalhaes and L.E. Dardenne, Receptor-ligand molecular docking. *Biophys Rev* **6** (2014), 75–87.

[46] N.S. Pagadala, K. Syed and J. Tuszynski, Software for molecular docking: a review. *Biophys Rev* **9** (2017), 91–102.

[47] C.M. Venkatachalam, X. Jiang, T. Oldfield and M. Waldman, LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J Mol Graph Model* **21** (203), 289–307.

[48] F. Österberg, G.M. Morris, M.F. Sanner, A.J. Olson and D.S. Goodsell, Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins Struct Funct Bioinf* **46** (2002), 34–40.

[49] R. Chen, L. Li and Z. Weng, ZDOCK: an initial-stage protein-docking algorithm. *Proteins Struct Funct Bioinf* **52** (2003), 80–87.

[50] C. Dominguez, R. Boelens and A.M. Bonvin, HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125** (2003), 1731–1737.

[51] S.Y. Huang and X. Zou, An iterative knowledge-based scoring function to predict protein–ligand interactions: I. Derivation of interaction potentials. *J Comput Chem* **27** (2006), 1866–1875.

[52] S.Y. Huang and X. Zou, An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. *J Comput Chem* **27** (2006), 1876–1882.

[53] A. Kerzmann, J. Fuhrmann, O. Kohlbacher and D. Neumann, BALL-Dock/SLICK: a new method for protein-carbohydrate docking. *J Chem Inf Model* **48** (2008), 1616–1625.

[54] A. Kerzmann, D. Neumann and O. Kohlbacher, SLICK-Scoring and Energy Functions for Protein-Carbohydrate Interactions. *J Chem Inf Model* **46** (2006), 1635–1642.

[55] M. Rarey, B. Kramer, T. Lengauer and G. Klebe, A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **261** (1996), 470–489.

[56] O. Trott and A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput chem* **31** (2010), 455–461.

[57] S.A. Samsonov, J. Teyra and M.T. Pisabarro, Docking glycosaminoglycans to proteins: analysis of solvent inclusion. *J Comput Aid Mol Des* **25** (2011), 477–489.

[58] D.F. Gauto, A.A. Petruk, C.P. Modenutti, J.I. Blanco, S. Di Lella and M.A. Martí, Solvent structure improves docking prediction in lectin–carbohydrate complexes. *Glycobiology* **23** (2013), 241–258.

[59] E. Fadda and R.J. Woods, Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov Today* **15** (2010), 596–609.

[60] T. Ishida, Computational analysis of carbohydrate recognition based on hybrid QM/MM modeling: a case study of norovirus capsid protein in complex with Lewis antigen. *Phy Chem Chem Phy* **20** (2018), 4652–4665.

[61] S.K. Mishra and J. Koča, Assessing the Performance of MM/PBSA, MM/GBSA, and QM-MM/GBSA Approaches on Protein/Carbohydrate Complexes: Effect of Implicit Solvent Models, QM Methods, and Entropic Contributions. *Phy Chem Chem Phy* **122** (2018), 8113–8121.

[62] J.A. Hadden, M.B. Tessier, E. Fadda and R.J. Woods, Calculating binding free energies for protein–carbohydrate complexes. *Glycoinformatics* **1273** (2015), 431–465.

[63] R. Bryce, I. Hillier and J. Naismith, Carbohydrate-protein recognition: molecular dynamics simulations and free energy analysis of oligosaccharide binding to concanavalin A. *Biophys J* **81** (2001), 1373–1388.

[64] M.G. Ford, T. Weimar, T. Köhli and R.J. Woods, Molecular dynamics simulations of galectin-1-oligosaccharide complexes reveal the molecular basis for ligand diversity. *Proteins Struct Funct Bioinf* **53** (2003), 229–240.

[65] Z. Liu and Y. Zhang, Molecular dynamics simulations and MM–PBSA calculations of the lectin from snowdrop (Galanthus nivalis). *J Mol Model* **15** (2009), 1501.

[66] J. Wang, C. Tan, Y.-H. Tan, Q. Lu and R. Luo, Poisson-Boltzmann solvents in molecular dynamics simulations. *Commun Comput Phys* **3** (2008), 1010–1031.

[67] R. Kadirvelraj, O.C. Grant, I.J. Goldstein, H.C. Winter, H. Tateno, E. Fadda and R.J. Woods, Structure and binding analysis of Polyporus squamosus lectin in complex with the Neu5Acα2-6Galβ1-4GlcNAc human-type influenza receptor. *Glycobiology* **21** (2011), 973–984.

[68] T. Negami, K. Shimizu and T. Terada, Coarse-grained molecular dynamics simulations of protein-ligand binding. *J Comput Chem* **35** (2014), 1835–1845.

[69] G.A. Voth, Coarse-graining of condensed phase and biomolecular systems. *CRC press* (2008).

[70] C.O. Solanke, D. Trapl, Z. Sucur, V. Mareska, I. Tvaroska and V. Spiwok, Atomistic simulation of carbohydrate-protein complex formation: Hevein-32 domain. *Sci Rep* **9** (2019), 18918.

[71] P. Valverde, J.I. Quintana, J.I. Santos, A. Ardá, J. Jimeénez-Barbero, Novel NMR avenues to explore the conformation and interactions of glycans. *ACS omega* **4** (2019), 13618–13630.

[72] J.H. Ardenkjaer-Larsen, G.S. Boebinger, A. Comment, S. Duckett, A.S. Edison, F. Engelke, C. Griesinger, R.G. Griffin, C. Hilty and H. Maeda, Facing and overcoming sensitivity challenges in biomolecular NMR spectroscopy. *Angew Chem Int Ed Engl* **54** (2015), 9162–9185.

[73] T.D. Heightman and A.T. Vasella, Recent insights into inhibition, structure, and mechanism of configuration-retaining glycosidases. *Angew Chem Int Ed Engl* **38** (1999), 750–770.

[74] J. Agirre, Strategies for carbohydrate model building, refinement and validation. *Acta Crystallogr D Struct Biol* **73** (2017), 171–186.

[75] R.J. Woods, Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. *Chem Rev* **118** (2018), 8005–8024.

Shuang Zhang
Dalton Cardiovascular Research Center
Department of Physics and Astronomy
Department of Biochemistry
Institute for Data Science and Informatics
University of Missouri
Columbia, MO 65211
USA
E-mail address: sz325@missouri.edu

Kyle Yu Chen
Rock Bridge High School
4303 South Providence Rd
Columbia, MO 65211
USA
E-mail address: kylec9596@gmail.com

Xiaoqin Zou
Dalton Cardiovascular Research Center
Department of Physics and Astronomy
Department of Biochemistry
Institute for Data Science and Informatics
University of Missouri
Columbia, MO 65211
USA
E-mail address: zoux@missouri.edu

Received December 15, 2020