Review

CH/π Interactions in Carbohydrate Recognition

Vojtěch Spiwok [ID]

Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague, Technická 3, Prague 6, 166 28, Czech Republic; spiwokv@vscht.cz; Tel.: +420-22044-3028; Fax: +420-22044-5140

Abstract: Many carbohydrate-binding proteins contain aromatic amino acid residues in their binding sites. These residues interact with carbohydrates in a stacking geometry via CH/π interactions. These interactions can be found in carbohydrate-binding proteins, including lectins, enzymes and carbohydrate transporters. Besides this, many non-protein aromatic molecules (natural as well as artificial) can bind saccharides using these interactions. Recent computational and experimental studies have shown that carbohydrate–aromatic CH/π interactions are dispersion interactions, tuned by electrostatics and partially stabilized by a hydrophobic effect in solvated systems.

Keywords: carbohydrate-protein interactions; CH/π interactions; lectins; interaction energy; non-canonical hydrogen bond

1. Carbohydrate–Aromatic Interactions

Our first contact with interactions between carbohydrates and aromatic systems was in the project focused on cold-active enzymes. We isolated the gene [1], cloned [1], recombinantly produced [1], and in collaboration with crystallographers, determined the three-dimensional structure [2,3] of a cold-active β-galactosidase from Antarctic bacterium *Arthrobacter* sp. C2-2. The project was focused on structural features of cold-active enzymes, thus we compared its structure with a mesophilic (normal temperature loving) enzyme. β-Galactosidase from *E. coli* [4] was a natural choice as the mesophilic counterpart. The presence of the residue Trp999 in *E. coli* enzyme and the absence of the corresponding tryptophan residue in the cold-active enzyme (Figure 1) was the most striking difference in active sites of both enzymes. This residue in *E. coli* β-galactosidase forms a parallel ("stacking") interaction with bound substrate [3,4]. This residue had also been previously shown by site-directed mutagenesis experiments to be important in catalysis and transglycosylation reactions (alternative transferase reactions found in many glycosidases) [5]. Later, it turned out that unique catalytic properties of the cold-active enzyme could not be attributed to this single residue; nevertheless, the residue Trp999 in *E. coli* β-galactosidase inspired our research for several years.

A closer look at this type of interaction has shown that these are widespread in carbohydrate recognition [6–8]. To illustrate this, it is possible to statistically evaluate the presence of residues in the vicinity of carbohydrate moieties in experimentally determined three-dimensional protein structures. The Glyvicinity tool from the glycosciences.de package [9] can be used for this purpose. It can be shown that tryptophan is the most abundant amino acid residue (12.1%) in the vicinity of β-D-glucopyranose in experimentally determined protein structures (distance < 5 Å, resolution < 3 Å, 9 April 2017). Furthermore, this number is likely to be negatively biased by non-specific complexes, with glycosidic detergents used to facilitate protein crystallization. The fact that tryptophan is the most abundant residue in the vicinity of β-D-glucopyranose is in contrast with the fact that this residue is the least abundant in human proteome (1.3%, based on codon usage [10]). More detailed analysis can be found in the work of Hudson et al. [11].
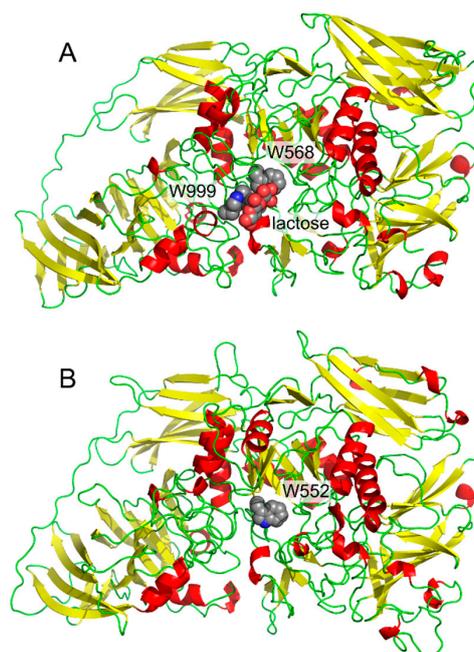
Figure 1. Comparison of β-galactosidases monomers from *E. coli* (**A**, with bound lactose, PDB ID: 1JZ8) and *Arthrobacter* sp. C2-2 (**B**, PDB ID: 1YQ2).

Tyrosine is the second most abundant residue in the vicinity of β-D-glucopyranose with 9.0%, despite also being a rather non-frequent amino acid in proteins. A similar, but not so strong, trend can be observed for other monosaccharide residues found in common glycoconjugates.

β-D-Glucopyranose is structurally predisposed for carbohydrate-aromatic interactions thanks to the fact that all of its ring C–H hydrogens are oriented axially. This makes it possible to interact with an aromatic system (indole of Trp or benzene of Phe or Tyr) in a parallel stacking geometry (Figure 2A), either at the *A* (top) or *B* (bottom) face of the carbohydrate ring (see the work of Rose et al. [12] for definition of faces; the *B* face corresponds to the α face in the anomer-based terminology in the D-hexopyranose series). α-D-Glucopyranose favors only the *A* face due to the fact that the anomeric hydroxyl group blocks the *B* face. Another common monosaccharide readily interacting with aromatic residues is α- or β-D-galactopyranose or α-L-fucopyranose. These carbohydrates interact via a patch of C–H groups at carbons *C3*, *C4*, *C5* and *C6* at the *B* face (Figure 2B). Similarly, β-D-mannopyranose can form a similar patch by C–H groups of carbons *C1*, *C2* and *C3* (Figure 2C). We can speculate that hexopyranoses found in common glycoconjugates were evolutionary selected from 16 possible stereoisomers for their ability to form carbohydrate–aromatic interactions. However, it must be mentioned that other carbohydrate-related factors [13] may be involved in the evolutionary selection of monosaccharides [14].

Figure 3 shows a series of examples of carbohydrate–aromatic interactions in natural systems. They show that carbohydrate–aromatic interactions are involved in a wide range of processes including carbohydrate binding, catalytic processing and transport. This overview can start with a carbohydrate-binding protein, in particular a lectin. As mentioned above, β-D-glucopyranose can form carbohydrate–aromatic interactions by either *A* or *B* face. Occasionally it forms a complex by both faces, as illustrated by carbohydrate binding module CBM4 from *Cellulomonas fimi* [15] (Figure 3A). Carbohydrate processing enzymes are represented by endoglucanase from *Clostridium thermocellum* [16]. The structure of an inactive mutant of this enzyme depicted in Figure 3B shows a hexasaccharide chain bound to the enzyme. Five of these six monosaccharide residues interact by carbohydrate–aromatic residues; two with tryptophan and three with tyrosine residues. The only exception is the catalytic-1 site (other sites are binding sites). Other examples of enzymes cleaving

β-D-glucopyranosides indicate that carbohydrate–aromatic interactions (in the orientation depicted in Figure 2A) are not compatible with catalysis. On the other hand, the orientation of D-galacto (and also D/L-fuco) depicted in Figure 2B is compatible with catalysis (possibly due to a longer distance from the cleaved bonds) as illustrated by *E. coli* and *Arthrobacter* β-galactosidase (Figure 1). A similar geometry to β-D-galactosidase can be found in a complex of cholera toxin with its carbohydrate ligand, which is one of strongest carbohydrate–protein complexes known [17].
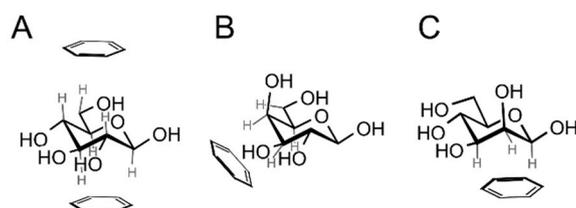

Figure 2. Prevalent geometries of carbohydrate–aromatic CH/π interactions: (**A**) β-D-glucopyranose can interact via both faces (only the *A* face is available for α-D-glucopyranose); (**B**) α- or β-D-galacto- or D-fucopyranose interacts via hydrogens on atoms *C3*, *C4*, *C5* and *C6* (a mirror assembly is typical for L-fucopyranose); (**C**) β-D-mannopyranose interacts via hydrogen atoms on *C1*, *C2* and *C3*.
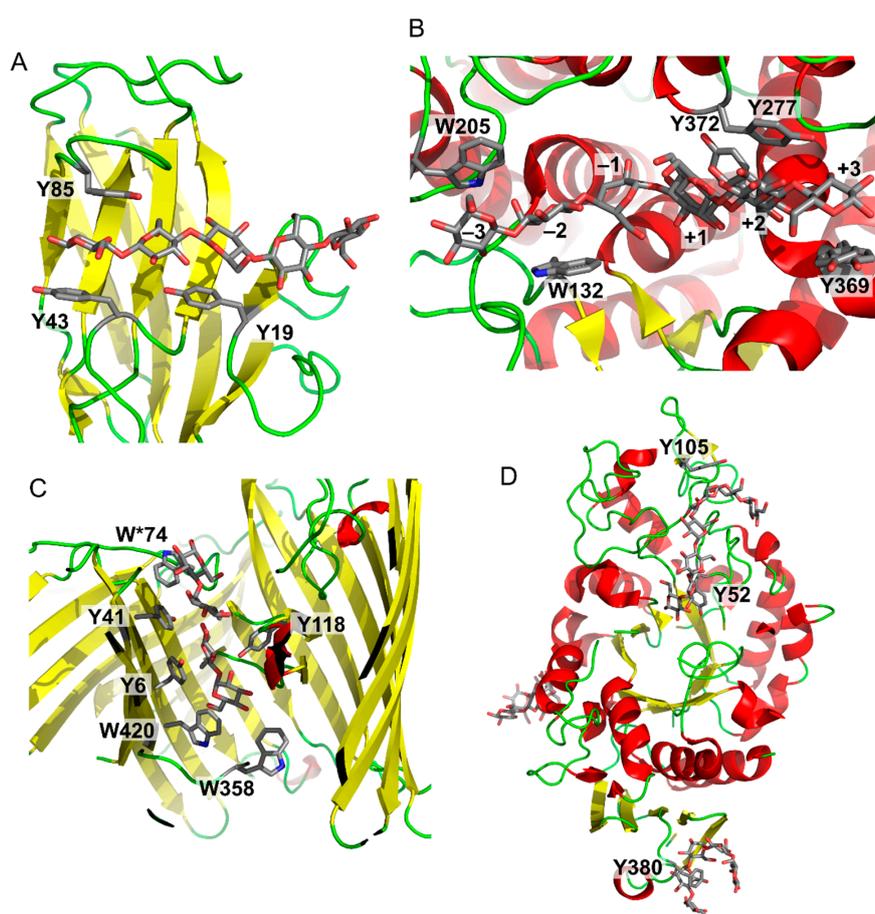

Figure 3. Examples of carbohydrate–aromatic complexes. (**A**) *Cellulomonas fimi* carbohydrate binding module CMB4 with (β-D-Glcp)$_5$ (PDB ID: 1GU3); (**B**) *Clostridium thermocellum* endoglucanase CelA with (β-D-Glcp)$_6$ (PDB ID: 1KWF); (**C**) "Greasy slide" of *E. coli* LamB maltoporin with two maltose molecules (PDB ID: 1MPM, residue W*74 belongs to another protein chain); (**D**) "Sugar tongs" in barley α-amylase (Y105 and Y52 form the active site; Y380 forms the "sugar tongs" site).

An interesting aromatic system involved in carbohydrate recognition and transport is "greasy slide" found in maltoporin [18] (Figure 3C). This carbohydrate transporter contains an aromatic patch formed by two tryptophans and two tyrosines on one side, and another tyrosine at the opposite side of the transport tunnel. The transported maltose can "slide" on these aromatic surfaces into the cell. Mutagenesis studies demonstrated an important role of these residues for the transport function [19].

Another interesting example of carbohydrate–aromatic interactions are "sugar tongs" [20] (Figure 3D). It was found that barley α-amylase 1 contains a surface tyrosine residue (Y380) distant from the active site. Despite its high distance from the active site its mutation to a non-aromatic residue drastically decreases affinity towards its substrate and its catalytic activity [21]. It is supposed that it helps the enzyme to interact with starch substrate as supported by a site directed mutagenesis and structural biology studies [20,21].

A cellulose–lignin assembly represents an example of natural carbohydrate–aromatic interaction between a carbohydrate and a non-protein system [22]. Carbohydrate–aromatic interactions can be found also in fully artificial systems. Artificial carbohydrate receptors are being developed for various applications, including future sugar sensors or pathogen scavengers. Many artificial carbohydrate receptors with different chemotypes have been developed [23]. Notably, receptors involving aromatic systems in carbohydrate–aromatic stacking geometries are highly successful. Davies et al. synthesized a series of carbohydrate receptors with an example shown in Figure 4A [24]. In water it binds cellobiose to its cavity with very high affinity in comparison with other artificial receptors (Kd = 1–2 mM) and with remarkable specificity.
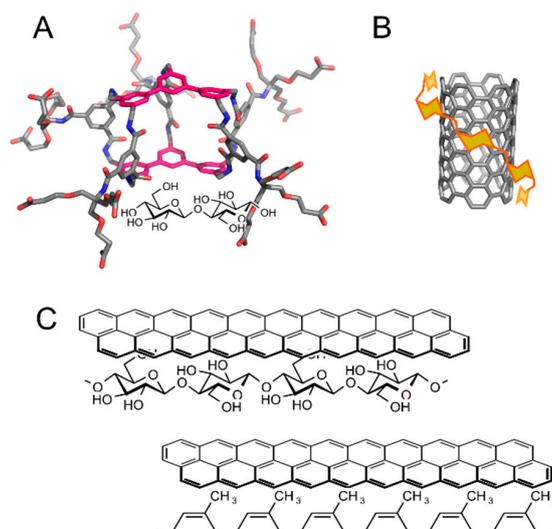

Figure 4. Examples of artificial carbohydrate–aromatic complexes: (**A**) An artificial carbohydrate receptor with ligand cellobiose (illustrative 3D structure generated by Avogadro [25]); (**B**) Schematic view of carbon nanotube solubilization by amylose; (**C**) Writing by pencil on a paper (top) and erasing it by a pencil eraser (bottom).

There are many other examples of artificial systems indicating an important role of carbohydrate–aromatic interactions in molecular recognition. For example, porous graphitic carbon columns can be successfully applied in the chromatography of carbohydrates, despite their high polarity [26]. Carbohydrates were also successfully used in the solubilisation of carbon-based nanomaterials such as carbon nanotubes [27,28] (Figure 4B). In addition, carbohydrate–aromatic interactions can be exploited in formation of gels from low-molecular weight compounds, such as FMOC-(fluorenylmethylcarbonyl-) or napthylmethylcarbonyl-functionalised gluco- or galactosamine [29].

The most lovely example of carbohydrate–aromatic interaction is writing with a pencil on paper, first coined by Motohiro Nishio in his book about CH/π interactions [30] (Figure 4C). A pencil contains aromatic molecules of graphite. Paper is composed of cellulose, which is a β-D-glucan with highly exposed C–H bonds (Figure 4C, top). A trace made by a pencil on a paper can hardly be washed away by polar as well as non-polar solvents, but it can be erased by a rubber eraser (Figure 4C, bottom), which is a competitive donor of C–H bonds.

2. Physical Nature of Carbohydrate–Aromatic Interactions—CH/π Bond

Noncovalent interactions can be classified based on geometry (e.g., stacking interactions), interacting moieties (e.g., hydrogen bonds, cation–π or halogen bonds) or based on their physical nature (e.g., electrostatic or dispersion). Carbohydrate–aromatic interactions are often seen as stacking interactions due to parallel orientation of the interacting carbohydrate and aromatic rings and also as CH/π interaction due to the fact that C–H bonds typically point toward the aromatic system.

The physical nature of carbohydrate–aromatic interactions was not clear when we started to study them. There were three different explanations found in the literature, all of them driven by chemical intuition rather than by experimental or computational results. The first group of authors explained carbohydrate–aromatic interactions as mostly electrostatic interactions between a partially negatively charged center of the aromatic ring and a partially positively charged hydrogen atom on a C–H bond (or more precisely, an aromatic quadrupole and a polarized C–H bond). The second group of authors explained these interactions as an example of CH/π interactions of mostly disperse nature. The third group of authors saw these interactions as purely hydrophobic, i.e., the effect of solvation and desolvation processes.

In order to elucidate the physical nature of these interactions we [31–34] and other authors [35–40] carried out a series of quantum chemical calculations. In principle, energy stabilizing in an intermolecular complex can be calculated as a difference between energy of the complex and energies of both interacting molecules. The quantum chemistry method must be carefully chosen and corrections to possible artifacts must be applied [41]. The value of energy stabilizing certain a noncovalent interaction can be further dissected into various components, most notably electrostatic, polarization and dispersion. Precise dissection into these terms require application of special computational methods (reviewed by Phipps et al. [42]). However, contribution of these two components can be roughly estimated from energies calculated by different quantum chemistry methods. Electrostatic interactions are, in general, well addressed by uncorrelated and density functional theory (DFT) methods. In contrast, uncorrelated and DFT methods usually fail in modeling of interactions of dispersion nature and correlated *ab initio* methods must be used. In the other words, molecular complexes dominated by electrostatic interactions are modeled as attractive by both correlated and uncorrelated methods, whereas interactions of dispersion nature are modeled as attractive only by correlated methods. Uncorrelated and most DFT methods underestimate the strength of dispersion interactions. This fact can be used to elucidate the nature of the bonds.

Quantum chemical energies of interactions [31–40] in different systems shaped our view of carbohydrate–aromatic interactions. First, these interaction are surprisingly strong, i.e., comparable in terms of energy with classical hydrogen bonds. Second, these interactions are of dispersion nature with minor role of electrostatics. It supports the theory of dispersion CH/π interactions. Finally, these results are rather neutral in relation to the hydrophobic theory. It shows that hydrophobic interactions (mediated by solvation and desolvation processes) are not necessary for binding, i.e., these interactions can be stable enough even in vacuum. However, hydrophobic interaction can contribute and further strengthen interactions in real carbohydrate–protein complexes.

These results of quantum chemical calculations were later supported by experimental spectroscopic studies [43]. Screen et al. used an IR ion-dip (IRID) technique setup involving formation of a molecular beam containing a carbohydrate (methyl- or phenyl- α-D-gluco-, -galacto- and α-L-fucopyranosides and α-L-fucose) and an aromatic molecule (toluene), irradiation by IR and

UV laser and mass spectrometry. This setup makes it possible to obtain a conformation-resolved precise (gas phase) IR spectra of carbohydrate–aromatic complexes. They have shown that, for example, α-L-fucopyranose forms a stable CH/π complex with toluene even in a gas phase, i.e., without solvation and desolvation processes. This is an experimental evidence showing strength of carbohydrate–aromatic CH/π interactions. It clearly demonstrates that these interactions are not (solely) hydrophobic.

The majority of quantum chemistry studies in the field used geometries of carbohydrate–aromatic complexes derived from experimentally determined structures. In these studies, the monosaccharide moiety was usually taken directly from the experimental structure containing a monosaccharide ligand (or "dissected" from oligo-saccharide or glycoside). The aromatic amino acid residue was also taken from the experimental structure and typically represented as a side chain (i.e., as 3-methylindole or cresol for tryptophan or tyrosine, respectively). The geometry of such a complex was optimized by a suitable quantum chemistry method (while keeping overall orientation of both moieties) and then the stabilizing energy was calculated at the best available level of theory. However, this approach lacks a systematic evaluation of many possible carbohydrate–aromatic orientations. For this reason we carried out a systematic scanning of energy surface characterizing these interactions [33]. Energy of each point on the surface was calculated by placement of a benzene molecule onto one of the faces of β-D-glucopyranose, β-D-mannopyranose or α-L-fucopyranose. The geometry of such a complex (3757 for each face and saccharide) was optimized using density functional tight binding with a dispersion correction (DFTB-D) while restraining the center of benzene in the starting position. Energy of the final pose was re-calculated by DFT with a dispersion correction (DFT-D). Selected minima were then evaluated using coupled cluster with single, double and perturbative triple excitations (CCSD(T)). These results support the previously observed strength of these interactions (stabilized by 3.54–5.40 kcal/mol) [33]. They also showed that these interactions are weakly directional, i.e., comparable energies were obtained for different positions of benzene relative to the carbohydrate.

Carbohydrate–aromatic interactions are often modeled by biomolecular simulations using molecular mechanics force fields. The question arises how these force fields perform in this task. In order to test this we compared energies calculated by force fields and quantum chemistry [32]. Surprisingly, molecular mechanics force fields (OPLS-AA, GROMOS, CSFF/CHARMM and Glycam/AMBER) performed well with good correlation between quantum mechanics and molecular mechanics levels of theory. The only exceptions were special force fields not intended for routine use in biomolecular simulations.

Experimental IRID studies have shown that carbohydrate–aromatic interactions are stable enough to exist in vacuum. However, this finding does not rule out the possibility that hydrophobic effect contributes to stability of solvated carbohydrate–protein complexes. Pioneer computational studies on thermodynamics of carbohydrate–aromatic interactions on isolated moieties were done by Wohlert et al. [44]. In order to test this on a carbohydrate–protein complex, we carried out a series of molecular dynamics simulations of hevein domain HEV32 with β-D-GlcNAc trisaccharide [45]. This system is very interesting for its small size, which makes it possible to simulate relatively long time scales [46]. It is also an excellent object for NMR studies [47]. Beside the simulation of this complex under real conditions, we also carried out a series of simulations under hypothetical conditions of absence or weakening of selected interactions. The method of molecular dynamics simulation makes it possible to simulate systems under hypothetical conditions not possible in a real experiment. Černý et al. developed this approach to demonstrate the role of dispersion interactions in stability of DNA and proteins [48,49]. We simulated the complex of HEV32 with trisaccharide and in these simulations we weakened selected classical (OH/O) hydrogen bonds and/or carbohydrate–aromatic (CH/π) interactions. It was necessary to weaken both types of interactions simultaneously to completely destabilize the studied complex. Separate weakening of a single type of interactions was not sufficient and the complex was relatively stable through the whole simulation.

There are two possible scenarios in dynamics of an artificially destabilized complex. Either, the complex is stabilized solely by physical (electrostatic or dispersion) interaction. Then, its

simulation with weakened interaction is likely to lead to destabilisation and dissociation. In contrast, a complex stabilized mostly by hydrophobic (i.e., solvation/desolvation processes) can stay stable even when physical interactions are weakened. In our simulations we observed something between these two scenarios, which indicates that carbohydrate–aromatic interactions combine CH/π and hydrophobic interactions.

Experimental determination of these contributions became possible thanks to development of a technique based on Enhanced Aromatic Sequons (described later in this review). Chen et al. [50] analyzed selected carbohydrate–aromatic interactions and observed comparable stabilizing contribution of CH/π and hydrophobic interactions. Analysed carbohydrate–aromatic complexes were stabilized by free energy of 0.8–1.0 kcal/mol. CH/π interactions contributed by 42–72% (also stabilization by an intrinsic effect of glycosylation was observed in some complexes). These results completed the picture of carbohydrate–aromatic interactions also experimentally.

Another elegant approach used in the research of carbohydrate–aromatic interactions is dynamic combinatorial chemistry. The concept of dynamic chemistry is based on dynamic bonds, such as disulphides or Schiff bases. Asensio et al. studied reactions of molecules D-L-NH$_2$ with O=CH-L'-A, where D are donors and A are acceptors of carbohydrate–aromatic interactions and L and L' are suitable linkers [51–53]. The process of Schiff bases formation in a pool of donor or acceptor molecules is driven purely thermodynamically due to the dynamic nature of the bond. Populations of different products are driven by stabilities of interactions between donors D and acceptors A. Then, the equilibrium can be frozen by reduction of Schiff bases to corresponding secondary amines, which makes it possible to analyze the products. Asensio et al. used NMR, which makes it possible to determine not only composition but also to study the interactions between donors and acceptors.

These studies highlighted that the strength of carbohydrate–aromatic interactions depends strongly on the number of CH/π contacts [51,53]. They also showed that axial substituents may decrease stability by blocking the interacting phase (in contrast, axial -OH or -OCH$_3$ groups on the opposite face stabilize carbohydrate–aromatic interactions) [53]. Equatorial -OCH$_3$ moieties in methylglycosides contribute to strength of carbohydrate–aromatic interactions as they interact by CH/π interactions [53]. O/π interactions formed by ring oxygen are significantly weaker than CH/π interactions [53].

The most interesting results of dynamic combinatorial chemistry studies are related to the contribution of electrostatics. First, it was demonstrated that polarization of a CH group by attachment of an oxygen atom (ring oxygen or hydroxyl) stabilizes the complex by approximately 0.3 kcal/mol per activated C–H bond [52]. The strong role of electrostatics was also observed when aromatic acceptors were varied [53]. In order to rationalize these results, Asensio et al. calculated energies stabilizing corresponding complexes by quantum chemical methods and decomposed them into electrostatic, polarization and dispersive contributions (desolvation effect was also evaluated). They found that electrostatic (approx. 3–5 kcal/mol) and polarization (approx. 1–1.5 kcal/mol) quantum chemical energies strongly correlated with experimental stabilities (relative free energies of carbohydrate–aromatic complexes). The dispersion energy stabilizing these interactions was generally stronger (approx. 7–8 kcal/mol), but there was very weak correlation with experimental stabilities. This shows that dispersion interactions provide the main "glue" of carbohydrate–aromatic binding with very little variations caused by directionality or the electrostatic nature of donors and acceptors. On the other hand, electrostatic interactions are slightly weaker but they can further "tune" strength and the geometry of carbohydrate–aromatic complexes, because they are highly influenced by directionality and charge distribution on donor and acceptor molecules. This is also in agreement with the findings of recent bioinformatics study that highlighted the role of electrostatics in orientation and strengths of carbohydrate–aromatic complexes [11].

3. Applications

Understanding the physics of carbohydrate–aromatic interactions can be used in many applications. The above mentioned artificial carbohydrate receptor [23,24] is a prominent example of such applications. Another interesting application of carbohydrate–aromatic interactions is in Enhanced Aromatic Sequons [54]. Proteins can be *N*-glycosylated at asparagine residues of **N**XT or **N**XS motif. By systematic studies of glycoproteins the group of Jeffrey W. Kelly found that glycosylation of the motif FX**N**XT provides faster folding and higher stability of the corresponding glycoproteins. This can be explained by existence of carbohydrate–aromatic interactions between the phenylalanine residue and the *N*-acetylglucosamine residue attached to asparagine [55]. This motif was installed into various proteins by site-directed mutagenesis and successfully improved their stability [54].

Besides this, we must mention bioinformatic algorithms aimed at recognition of carbohydrate-binding sites in proteins on the basis of an amino acid sequence or 3D structure. For example, Doxey et al. [56] used a statistical approach to predict carbohydrate binding sites on a series of proteins. Among them, the tobacco pathogenesis-related protein PR-5d was selected and interaction with cellulose was demonstrated experimentally [56]. Aromatic amino acid residues played a major role in this prediction.

A direct application of knowledge of carbohydrate–aromatic CH/π interactions can be demonstrated on carbohydrate–protein docking. Carbohydrate docking is very difficult for a number of reasons, including generally weak binding, number of water-mediated interactions and pseudosymmetry of sugars. Involvement of CH/π interactions was proposed as another pitfall. This lead to development of the SLICK (Sugar–Lectin Interactions and doCKing) scoring function (i.e., structure-free energy relationship) tailored for carbohydrate docking [57]. This scoring function contains an explicit term for CH/π interactions in its equation. This method has been shown in combination with BALLDock to outperform general protein–ligand docking packages in carbohydrate–protein docking tasks.

4. Conclusions

Theoretical and experimental studies have formed our current view of carbohydrate–aromatic interactions. These interactions are widespread in carbohydrate–protein complexes. They are found in carbohydrate-binding proteins, carbohydrate-processing enzymes and carbohydrate transporters. They can be also found in complexes of carbohydrates with aromatic non-protein binding partners, natural as well as artificial. These interactions are examples of CH/π non-canonical hydrogen bonds. Carbohydrate–aromatic interactions can exist in vacuum without a hydrophobic effect, but this effect contributes to binding in natural carbohydrate–protein complexes. Understanding the nature and geometries of these interactions may be employed in searching for carbohydrate binding sites, docking of carbohydrates, design of carbohydrate receptors and other applications.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Karasová-Lipovová, P.; Strnad, H.; Spiwok, V.; Malá, Š.; Kralová, B.; Russell, N.J. The Cloning, Purification and Characterisation of a Cold-Active β-Galactosidase from the Psychrotolerant Antarctic Bacterium *Arthrobacter* sp. C2-2. *Enzym. Microb. Technol.* **2003**, *33*, 836–844. [CrossRef]
2. Petroková, H.; Vondráčková, E.; Skálová, T.; Dohnálek, J.; Lipovová, P.; Spiwok, V.; Strnad, H.; Králová, B.; Hašek, J. Crystallization and Preliminary X-Ray Diffraction Analysis of Cold-Active β-Galactosidase from *Arthrobacter* sp. C2-2. *Collect. Czech. Chem. Commun.* **2005**, *70*, 124–132. [CrossRef]
3. Skálová, T.; Dohnálek, J.; Spiwok, V.; Lipovová, P.; Vondráčková, E.; Petroková, H.; Dušková, J.; Strnad, H.; Králová, B.; Hašek, J. Cold-active β-Galactosidase from Arthrobacter sp. C2-2 Forms Compact 660 kDa Hexamers: Crystal Structure at 1.9 Å Resolution. *J. Mol. Biol.* **2005**, *353*, 282–294. [CrossRef] [PubMed]

4. Juers, D.H.; Heightman, T.D.; Vasella, A.; McCarter, J.D.; Mackenzie, L.; Withers, S.G.; Matthews, B.W. A structural view of the action of Escherichia coli (LacZ) β-galactosidase. *Biochemistry* **2001**, *40*, 14781–14794. [CrossRef] [PubMed]

5. Huber, R.E.; Hakda, S.; Cheng, C.; Cupples, C.G.; Edwards, R.A. Trp-999 of β-galactosidase (Escherichia coli) is a key residue for binding, catalysis, and synthesis of allolactose, the natural lac operon inducer. *Biochemistry* **2003**, *42*, 1796–1803. [CrossRef] [PubMed]

6. Muraki, M. The importance of CH/π interactions to the function of carbohydrate binding proteins. *Protein Peptide Lett.* **2002**, *9*, 195–209. [CrossRef]

7. Balaji, P.V. Contribution of C-H...π Interactions to the Affinity and Specificity of Carbohydrate Binding Sites. *Mini-Rev. Org. Chem.* **2011**, *8*, 222–228. [CrossRef]

8. Asensio, J.L.; Ardá, A.; Cañada, F.J.; Jiménez-Barbero, J. Carbohydrate-aromatic interactions. *Acc. Chem. Res.* **2013**, *46*, 946–954. [CrossRef] [PubMed]

9. Rojas-Macias, M.A.; Lütteke, T. Statistical analysis of amino acids in the vicinity of carbohydrate residues performed by GlyVicinity. *Methods Mol. Biol.* **2015**, *1273*, 215–226. [CrossRef] [PubMed]

10. Codon Usage Database. Available online: http://www.kazusa.or.jp/codon/ (accessed on 9 April 2017).

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

12. Rose, I.A.; Hanson, K.R.; Wilkinson, K.D.; Wimmer, M.J. A suggestion for naming faces of ring compounds. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 2439–2441. [CrossRef] [PubMed]

13. Pérez, S.; Tvaroška, I. Carbohydrate-protein interactions: Molecular modeling insights. *Adv. Carbohyd. Chem. Biochem.* **2014**, *71*, 9–136. [CrossRef]

14. Eschenmoser, A. Etiology of Potentially Primordial Biomolecular Structures: From Vitamin B12 to the Nucleic Acids and an Inquiry into the Chemistry of Life's Origin: A Retrospective. *Angew. Chem. Int. Ed.* **2011**, *50*, 12412–12472. [CrossRef] [PubMed]

15. Boraston, A.B.; Nurizzo, D.; Notenboom, V.; Ducros, V.; Rose, D.R.; Kilburn, D.G.; Davies, G.J. Differential Oligosaccharide Recognition by Evolutionarily-Related β-1,4 and β-1,3 Glucan-Binding Modules. *J. Mol. Biol.* **2002**, *319*, 1143–1156. [CrossRef]

16. Guerin, D.M.; Lascombe, M.B.; Costabel, M.; Souchon, H.; Lamzin, V.; Beguin, P.; Alzari, P.M. Atomic (0.94 Å) resolution structure of an inverting glycosidase in complex with substrate. *J. Mol. Biol.* **2002**, *316*, 1061–1069. [CrossRef] [PubMed]

17. Merritt, E.A.; Kuhn, P.; Sarfaty, S.; Erbe, J.L.; Holmes, R.K.; Hol, W.G. The 1.25 Å resolution refinement of the cholera toxin B-pentamer: Evidence of peptide backbone strain at the receptor-binding site. *J. Mol. Biol.* **1998**, *282*, 1043–1059. [CrossRef] [PubMed]

18. Dutzler, R.; Wang, Y.F.; Rizkallah, P.J.; Rosenbusch, J.P.; Schirmer, T. Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway. *Structure* **1996**, *4*, 127–134. [CrossRef]

19. Denker, K.; Orlik, F.; Schiffler, B.; Benz, R. Site-directed mutagenesis of the greasy slide aromatic residues within the LamB (maltoporin) channel of Escherichia coli: Effect on ion and maltopentaose transport. *J. Mol. Biol.* **2005**, *352*, 534–550. [CrossRef] [PubMed]

20. Robert, X.; Haser, R.; Gottschalk, T.E.; Ratajczak, F.; Driguez, H.; Svensson, B.; Aghajari, N. The structure of barley alpha-amylase isozyme 1 reveals a novel role of domain C in substrate recognition and binding: A pair of sugar tongs. *Structure* **2003**, *11*, 973–984. [CrossRef]

21. Bozonnet, S.; Jensen, M.T.; Nielsen, M.M.; Aghajari, N.; Jensen, M.H.; Kramhøft, B.; Willemoës, M.; Tranier, S.; Haser, R.; Svensson, B. The 'pair of sugar tongs' site on the non-catalytic domain C of barley alpha-amylase participates in substrate binding and activity. *FEBS J.* **2007**, *274*, 5055–5067. [CrossRef] [PubMed]

22. Besombes, S.; Mazeau, K. Molecular dynamics simulations of a guaiacyl β-*O*-4 lignin model compound: Examination of intramolecular hydrogen bonding and conformational flexibility. *Biopolymers* **2004**, *73*, 301–315. [CrossRef] [PubMed]

23. Davis, A.P.; Wareham, R.S. Carbohydrate Recognition through Noncovalent Interactions: A Challenge for Biomimetic and Supramolecular Chemistry. *Angew. Chem. Int. Edit.* **1999**, *38*, 2978–2996. [CrossRef]

24. Ferrand, Y.; Crump, M.P.; Davis, A.P. A Synthetic Lectin Analogue for Biomimetic Disaccharide Recognition. *Science* **2007**, *318*, 619–622. [CrossRef] [PubMed]

25. Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchison, G.R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **2012**, *4*, 17. [CrossRef] [PubMed]

26. Gabbanini, S.; Lucchi, E.; Guidugli, F.; Matera, R.; Valgimigli, L. Anomeric discrimination and rapid analysis of underivatized lactose, maltose, and sucrose in vegetable matrices by U-HPLC-ESI-MS/MS using porous graphitic carbon. *J. Mass. Spectrom.* **2010**, *45*, 1012–1018. [CrossRef] [PubMed]

27. Kim, O.K.; Je, J.; Baldwin, J.W.; Kooi, S.; Pehrsson, P.E.; Buckley, L.J. Solubilization of single-wall carbon nanotubes by supramolecular encapsulation of helical amylose. *J. Am. Chem. Soc.* **2003**, *125*, 4426–4427. [CrossRef] [PubMed]

28. Bhoi, V.I.; Kumar, S.; Murthy, C.N. The self-assembly and aqueous solubilization of [60]fullerene with disaccharides. *Carbohyd. Res.* **2012**, *359*, 120–127. [CrossRef] [PubMed]

29. Birchall, L.S.; Roy, S.; Jayawarna, V.; Hughes, M.; Irvine, E.; Okorogheye, G.T.; Saudi, N.; De Santis, E.; Tuttle, T.; Edwards, A.A.; et al. Exploiting CH-π interactions in supramolecular hydrogels of aromatic carbohydrate amphiphiles. *Chem. Sci.* **2011**, *2*, 1349–1355. [CrossRef]

30. Nishio, M.; Hirota, M.; Umezawa, Y. *The CH-Pi Interaction: Evidence, Nature, and Consequences*, 1st ed.; Wiley-VCH: New York, NY, USA, 1998.

31. Spiwok, V.; Lipovová, P.; Skálová, T.; Buchtelová, E.; Hašek, J.; Králová, B. Role of CH/π Interactions in Substrate Binding by Escherichia coli β-Galactosidase. *Carbohyd. Res.* **2004**, *339*, 2275–2280. [CrossRef] [PubMed]

32. Spiwok, V.; Lipovová, P.; Skálová, T.; Vondráčková, E.; Dohnálek, J.; Hašek, J.; Králová, B. Modelling of Carbohydrate-Aromatic Interactions: Ab Initio Energetics and Force Field Performance. *J. Comput. Aided Mol. Des.* **2005**, *19*, 887–901. [CrossRef] [PubMed]

33. Kozmon, S.; Matuška, R.; Spiwok, V.; Koča, J. Three-Dimensional Potential Energy Surface of Selected Carbohydrates' CH/π Dispersion Interactions Calculated by High-Level Quantum Mechanical Methods. *Chem. Eur. J.* **2011**, *17*, 5680–5690. [CrossRef] [PubMed]

34. Kozmon, S.; Matuška, R.; Spiwok, V.; Koča, J. Dispersion Interactions of Carbohydrates with Condensate Aromatic Moieties: Theoretical Study on the CH-π Interaction Additive Properties. *Phys. Chem. Chem. Phys.* **2011**, *13*, 14215–14222. [CrossRef] [PubMed]

35. Fernandez-Alonso, M.C.; Canada, F.J.; Jimenez-Barbero, J.; Cuevas, G. Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions. *J. Am. Chem. Soc.* **2005**, *127*, 7379–7386. [CrossRef] [PubMed]

36. Sujatha, M.S.; Sasidhar, Y.U.; Balaji, P.V. Insights into the role of the aromatic residue in galactose-binding sites: MP2/6-311G++** study on galactose- and glucose-aromatic residue analogue complexes. *Biochemistry* **2005**, *44*, 8554–8562. [CrossRef] [PubMed]

37. Sujatha, M.S.; Sasidhar, Y.U.; Balaji, P.V. MP2/6-311++G(d,p) study on galactose–aromatic residue analogue complexes in different position-orientations of the saccharide relative to aromatic residue. *J. Mol. Struct. THEOCHEM* **2007**, *814*, 11–24. [CrossRef]

38. Sharma, R.; McNamara, J.P.; Raju, R.K.; Vincent, M.A.; Hillier, I.A.; Morgado, C.A. The interaction of carbohydrates and amino acids with aromatic systems studied by density functional and semi-empirical molecular orbital calculations with dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2767–2774. [CrossRef] [PubMed]

39. Raju, R.K.; Ramraj, A.; Hillier, I.H.; Vincent, M.A.; Burton, N.A. Carbohydrate-aromatic π interactions: A test of density functionals and the DFT-D method. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3411–3416. [CrossRef] [PubMed]

40. Tsuzuki, S.; Uchimaru, T.; Mikami, M. Magnitude and nature of carbohydrate-aromatic interactions: Ab initio calculations of fucose-benzene complex. *J. Phys. Chem. B* **2009**, *113*, 5617–5621. [CrossRef] [PubMed]

41. Hobza, P.; Muller-Dethlefs, K. *Non-Covalent Interactions: Theory and Experiment*, 1st ed.; Royal Society of Chemistry: Cambridge, UK, 2009.

42. Phipps, M.J.; Fox, T.; Tautermann, C.S.; Skylaris, C.K. Energy decomposition analysis approaches and their evaluation on prototypical protein-drug interaction patterns. *Chem. Soc. Rev.* **2015**, *44*, 3177–3211. [CrossRef] [PubMed]

43. Screen, J.; Stanca-Kaposta, E.C.; Gamblin, D.P.; Liu, B.; Macleod, N.A.; Snoek, L.C.; Davis, B.G.; Simons, J.P. IR-spectral signatures of aromatic-sugar complexes: Probing carbohydrate-protein interactions. *Angew. Chem. Int. Ed.* **2007**, *46*, 3644–3648. [CrossRef] [PubMed]

44. Wohlert, J.; Schnupf, U.; Brady, J.W. Free energy surfaces for the interaction of D-glucose with planar aromatic groups in aqueous solution. *J. Chem. Phys.* **2010**, *133*, 155103. [CrossRef] [PubMed]

45. Mareška, V.; Tvaroška, I.; Králová, B.; Spiwok, V. Molecular Simulations of Hevein/(GlcNAc)3 Complex with Weakened OH/O and CH/π Hydrogen Bonds: Implications for their Role in Complex Stabilization. *Carbohyd. Res.* **2015**, *408*, 1–7. [CrossRef] [PubMed]

46. Colombo, G.; Meli, M.; Cañada, J.; Asensio, J.L.; Jimenez-Barbero, J. A dynamic perspective on the molecular recognition of chitooligosaccharide ligands by hevein domains. *Carbohyd. Res.* **2005**, *340*, 1039–1049. [CrossRef] [PubMed]

47. Asensio, J.L.; Siebert, H.C.; von Der Lieth, C.W.; Laynez, J.; Bruix, M.; Soedjanaamadja, U.M.; Beintema, J.J.; Cañada, F.J.; Gabius, H.J.; Jiménez-Barbero, J. NMR investigations of protein-carbohydate interactions: Studies on the relevance of Trp/Tyr variations in lectin binding sites as deduced from titration microcalorimetry and NMR studies on hevein domains. Determination of the NMR structure of the complex between pseudohevein and *N,N′,N′′*-triacetylchitotriose. *Proteins* **2000**, *40*, 218–236. [CrossRef] [PubMed]

48. Černý, J.; Kabeláč, M.; Hobza, P. Double-helical→ladder structural transition in the B-DNA is induced by a loss of dispersion energy. *J. Am. Chem. Soc.* **2008**, *130*, 16055–16059. [CrossRef] [PubMed]

49. Černý, J.; Vondrášek, J.; Hobza, P. Loss of dispersion energy changes the stability and folding/unfolding equilibrium of the Trp-cage protein. *J. Phys. Chem. B* **2009**, *113*, 5657–5660. [CrossRef] [PubMed]

50. Chen, W.; Enck, S.; Price, J.L.; Powers, D.L.; Powers, E.T.; Wong, C.H.; Dyson, H.J.; Kelly, J.W. Structural and energetic basis of carbohydrate-aromatic packing interactions in proteins. *J. Am. Chem. Soc.* **2013**, *135*, 9877–9884. [CrossRef] [PubMed]

51. Santana, A.G.; Jiménez-Moreno, E.; Gómez, A.M.; Corzana, F.; González, C.; Jiménez-Oses, G.; Jiménez-Barbero, J.; Asensio, J.L. A dynamic combinatorial approach for the analysis of weak carbohydrate/aromatic complexes: Dissecting facial selectivity in CH/π stacking interactions. *J. Am. Chem. Soc.* **2013**, *135*, 3347–3350. [CrossRef] [PubMed]

52. Jiménez-Moreno, E.; Gómez, A.M.; Bastida, A.; Corzana, F.; Jiménez-Oses, G.; Jiménez-Barbero, J.; Asensio, J.L. Modulating weak interactions for molecular recognition: A dynamic combinatorial analysis for assessing the contribution of electrostatics to the stability of CH-π bonds in water. *Angew. Chem. Int. Ed.* **2015**, *54*, 4344–4348. [CrossRef] [PubMed]

53. Jiménez-Moreno, E.; Jiménez-Osés, G.; Gómez, A.M.; Santana, A.G.; Corzana, F.; Bastida, A.; Jiménez-Barbero, J.; Asensio, J.L. A thorough experimental study of CH/π interactions in water: Quantitative structure-stability relationships for carbohydrate/aromatic complexes. *Chem. Sci.* **2015**, *6*, 6076–6085. [CrossRef]

54. Murray, A.N.; Chen, W.; Antonopoulos, A.; Hanson, S.R.; Wiseman, R.L.; Dell, A.; Haslam, S.M.; Powers, D.L.; Powers, E.T.; Kelly, J.W. Enhanced aromatic sequons increase oligosaccharyltransferase glycosylation efficiency and glycan homogeneity. *Chem. Biol.* **2015**, *22*, 1052–1062. [CrossRef] [PubMed]

55. Wyss, D.F.; Choi, J.S.; Li, J.; Knoppers, M.H.; Willis, K.J.; Arulanandam, A.R.; Smolyar, A.; Reinherz, E.L.; Wagner, G. Conformation and function of the N-linked glycan in the adhesion domain of human CD2. *Science* **1995**, *269*, 1273–1278. [CrossRef] [PubMed]

56. Doxey, A.C.; Cheng, Z.; Moffatt, B.A.; McConkey, B.J. Structural motif screening reveals a novel, conserved carbohydrate-binding surface in the pathogenesis-related protein PR-5d. *BMC Struct. Biol.* **2010**, *10*, 23. [CrossRef] [PubMed]

57. Kerzmann, A.; Fuhrmann, J.; Kohlbacher, O.; Neumann, D. BALLDock/SLICK: A new method for protein-carbohydrate docking. *J. Chem. Inf. Model.* **2008**, *48*, 1616–1625. [CrossRef] [PubMed]