

Carbohydrate – Protein aromatic ring interactions beyond CH/ $\pi$  interactions: A Protein Data Bank survey and quantum chemical calculations

Ivana M. Stanković, Jelena P. Blagojević Filipović, Snežana D. Zarić



PII: S0141-8130(20)32838-5

DOI: <https://doi.org/10.1016/j.ijbiomac.2020.03.251>

Reference: BIOMAC 15202

To appear in: *International Journal of Biological Macromolecules*

Received date: 18 January 2020

Revised date: 12 March 2020

Accepted date: 30 March 2020

Please cite this article as: I.M. Stanković, J.P. Blagojević Filipović and S.D. Zarić, Carbohydrate – Protein aromatic ring interactions beyond CH/ $\pi$  interactions: A Protein Data Bank survey and quantum chemical calculations, *International Journal of Biological Macromolecules* (2020), <https://doi.org/10.1016/j.ijbiomac.2020.03.251>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# CARBOHYDRATE – PROTEIN AROMATIC RING INTERACTIONS BEYOND CH/ $\pi$ INTERACTIONS: A PROTEIN DATA BANK SURVEY AND QUANTUM CHEMICAL CALCULATIONS

*Ivana M. Stanković<sup>a</sup>, Jelena P. Blagojević Filipović<sup>b</sup>, and Snežana D. Zarić<sup>c\*</sup>*

<sup>a</sup> ICTM, University of Belgrade, Njegoševa 12, Belgrade, Serbia;

<sup>b</sup> Innovation center of the Faculty of Chemistry, Studentski trg 12-16, 11000 Belgrade, Serbia;

<sup>c</sup> Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, Belgrade, Serbia

\*corresponding author: szaric@chem.bg.ac.rs

keywords: Carbohydrates, Aromatic Amino Acids, Stacking interactions, CH/ $\pi$  interactions

## Abstract

The geometries of the contacts between monosaccharides and aromatic rings of amino acids found in X-ray crystallography structures, in the Protein Data Bank (PDB), were analysed, while the energies of the interactions were calculated using quantum chemical method. We found 1913 sugar/aromatic ring contacts, 1054 of them (55%) with CH/ $\pi$  interactions and 859 of them (45%) without CH/ $\pi$  interactions. We showed that only the carbohydrate/aromatic contacts with CH/ $\pi$  interactions are preferentially parallel and enable sliding in the plane parallel to aromatic ring. The calculated interaction energies in systems with CH/ $\pi$  interactions are in the range from -1.7 kcal/mol to -6.8 kcal/mol, while in the systems without CH/ $\pi$  interactions are in the range -0.2 to -3.2 kcal/mol. Hence, the binding that does not include CH/ $\pi$  interactions, can also be important for aromatic amino acid and carbohydrate binding processes, since some of these interactions can be as strong as the CH/ $\pi$  interactions. At the same time, these interactions can be weak enough to enable releasing of small carbohydrate fragments after the enzymatic reaction. The analysis of the protein-substrate patterns showed that every second or third carbohydrate unit in long substrates stacks with protein aromatic amino acids.

## Abbreviations

PDB, Protein Data Bank; NMR, Nuclear Magnetic Resonance; RSL, Ralstonia solanacearum lectin; MD, Molecular Dynamics; VMD, Visual Molecular Dynamics; MP2, Møller-Plesset Perturbation Theory of second order; BSSE, Basis Set Superposition Error; CCSD(T), Coupled-Cluster with Single and Double and Perturbative Triple excitations; CBS, Complete Basis Set

## 1. Introduction

The recognition of saccharides by proteins has far reaching implications in biology, technology, and drug design. Carbohydrate - receptor interactions play an important role in many cellular processes, such as cell-cell adhesion, cell differentiation and in-cell signaling [1-11]. Unexpectedly, researchers observed that despite the highly hydrophilic character of most sugars, aromatic rings of the receptor interact with carbohydrate rings the most frequently [12]. Aromatic protein rings, phenylalanine, tyrosine and tryptophan, stack with the hydrophobic face (or a face) of a carbohydrate six-membered rings (Figure S1) giving rise to hydrophobic contacts that are loose enough to release substrate after the enzymatic reaction in order to proceed with hydrolysis on the next spot in sugar chain, as observed in molecular dynamics simulations [13-15]. This characteristic, called processivity, is as follows: after the hydrolysis event on a certain glycosidic unit, a carbohydrate chain shifts forward inside the enzyme cleft and exposes its other unit to the enzyme active site. The processivity, for example, was observed in the study of the cellulose.

Molecular dynamics simulations and normal mode analysis was used to examine the functional role of aromatic residues inside the active site tunnel of cellobiohydrolase 2 from *Trichoderma reesei* [16]. This enzyme contains three tryptophans which interact with a carbohydrate substrate and each one of them was mutated computationally to alanine residue. The calculations performed with the wild-type

enzyme and the three mutants showed that the free energy change was positive in case of the three mutants in relation to the wild-type enzyme. Therefore, the aromatic residues, associated with the substrate acquisition and product stabilization, exhibit large influence on the binding free energy and, accordingly, are prone to have great influence on the enzymatic activity. Consistent results were obtained with NMR experiments, which showed higher spatial proximity between substrate and aromatic residues than between substrate and other types of residues in mutated enzymes [17]. The aromatic platforms play a major role in determining the specificity of the molecular recognition process, both by aromatic-aromatic interactions or interactions with other moieties including carbohydrates [14,16,18-23].

The existence of stabilizing CH/ $\pi$  carbohydrate-aromatic interactions is demonstrated from both the theoretical and experimental viewpoints. The interaction energy between different aromatic rings and simple monosaccharides, based on quantum mechanical calculations in the gas phase ranges from -3.0 to -6.0 kcal/mol [19,24]. It was found that the interactions between carbohydrates and aromatic amino acid side chains are aided by nonconventional hydrogen bonds: the CH/ $\pi$  interactions [19,25-27]. These interactions are not highly directional, therefore enable sliding. The study of Wimmerová and collaborators [8] quantified how the CH/ $\pi$  interaction contributes to the carbohydrate - protein interaction in the complex of *Ralstonia solanacearum* lectin (RSL) with  $\alpha$ -L-Me-fucoside. The contribution of the CH/ $\pi$  interaction between Trp and  $\alpha$ -L-Me-fucoside was estimated by creating Trp to Ala mutants of RSL. Observed results suggest that in this and similar cases the carbohydrate-protein interaction can be driven mainly by a dispersion interaction and that the CH/ $\pi$  interaction is prevalent.

Here, we systematically study the geometry of monosaccharide-aromatic amino acid contacts, found in the PDB structures, resolved at atomic level by X-ray crystallography. We also perform quantum chemical calculations of interaction energies. Previously, a statistical analysis of carbohydrate-

aromatic CH/ $\pi$  interactions in PDB structures was performed [12], however, in our work we study both contacts with and without CH/ $\pi$  interactions. We also describe the pattern of number of aromatic-sugar contacts for each polysaccharide substrate. While previous quantum chemical studies calculated CH/ $\pi$  interaction energies on carbohydrate-aromatic model system [24] or systematically scanned potential energy surfaces of CH/ $\pi$  interactions [28] we performed calculations of interaction energies between carbohydrate and aromatic amino acid molecules, using geometries that are extracted from PDB, with and without CH/ $\pi$  interactions. The frequency of occurrence in the PDB structures and calculated energies indicate similar importance of interactions with and without CH/ $\pi$  interactions.

## 2. Methodology

### 2.1. PDB search

The carbohydrate-protein database has been constructed within Protein Data Bank (PDB) by searching for proteins resolved in complex with carbohydrates in the form of pyranose, on the 27th of April of 2018. Only crystallographic structures with resolution of up to 3.0 Å and with the  $R_{\text{work}}-R_{\text{free}}$  difference of up to 0.05 were taken. The missing hydrogen atoms were computationally added by using Reduce [29] software under Phenix [30].

Carbohydrates were defined as 6-membered sugar monomers not containing voluminous substituents. Glycoproteins were not taken into account, i. e. if the sugars are connected to proteins directly, by covalent bond(s), or through metal coordination. The linked sugar residues were detected as indicated in the PDB files and reconstructed employing the graph theory. We searched for aromatic rings within 4.0 Å around any carbohydrate unit: i.e. any aromatic ring heavy atom up to 4.0 Å of any heavy atom of a carbohydrate unit. We have excluded all sugar/aromatic contacts with any foreign

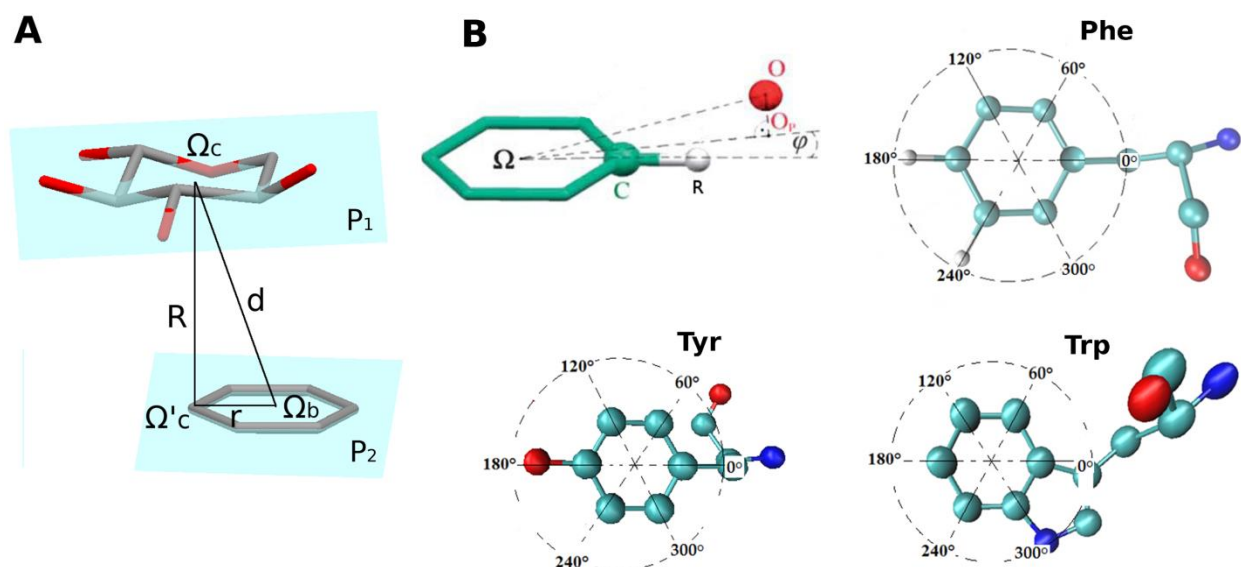
atoms (atoms that do not pertain to either the sugar or the aromatic ring studied) within the sphere of center-center distance diameter in order to analyze only the direct interactions between the two rings.

As carbohydrate rings do not possess one defined plane of the 6-membered ring, we defined an average sugar ring plane by choosing three equidistant points in the middle of the ring bonds, because the planes defined by any three ring equidistant atoms are very different. We determined the geometrical parameters such as the normal distance between the average planes of the interacting rings ( $R$ ), the offset ( $r$ ) the angle between planes ( $P_1/P_2$ ) (Figure 1A) and the  $\varphi$  angle (Figure 1B). For the offset value for tryptophan, we took the smaller of the two offset values, corresponding to the six- and five-membered ring. In the polar graph  $r(\varphi)$  we took only the six-membered ring offset for tryptophan.

The criteria for the CH/ $\pi$  interaction are the same as in previous work [12]: the distance between the C atom and the center of phenyl ring is  $\leq 4.5 \text{ \AA}$ , the angle between the CH vector and the phenyl ring plane normal vector is  $\leq 40^\circ$ , and the distance between C-projection to the ring and the ring center is  $\leq 2.0 \text{ \AA}$  ( $1.6 \text{ \AA}$  for the 5-membered tryptophane ring).

Some additional interactions, other than CH/ $\pi$  (classical hydrogen bonds, CH/N, OH/ $\pi$  and CH/O) can also occur in carbohydrate/aromatic ring contacts and their influence is also considered in this work. For the OH/ $\pi$  interactions, we considered the same criteria as for CH/ $\pi$  interactions. The criteria for the CH/N and CH/O interactions are: the distance between the H atom and O or N atom  $\leq 2.9 \text{ \AA}$ , and the angle C-H-O (or C-H-N) is  $>110^\circ$ , as in our previous paper about CH/O interactions in proteins [31].

The criteria for the classical hydrogen bonds are: the angle donor-H-acceptor  $> 110^\circ$  and the distance donor-acceptor  $< 4.0 \text{ \AA}$ . We do not consider the CH<sub>2</sub>OH carbohydrate out-of-ring group for the interaction analysis, as its position is not predictable due to bond rotation [12]. Also, we only consider amino acid aromatic ring, without backbone and without C $_{\beta}$  atom.



**Figure 1.** (A) Geometrical parameters used to define the interactions of aromatic and carbohydrate fragments: Parameter  $d$  is the intermolecular distance between the rings centers. The normal distance between the average planes of the interacting rings is  $R$ . The distance between the center of the aromatic ring ( $\Omega_b$ ) and the projection of the center of the carbohydrate ring onto the plane of the aromatic ring ( $\Omega'_c$ ) is the horizontal displacement (offset)  $r$ . P<sub>1</sub> and P<sub>2</sub> are monosaccharide and aromatic ring planes, respectively. (B)  $\phi$  angle is the angle between the sugar center projection onto aromatic plane ( $O_p$ ), aromatic ring center ( $\Omega$ ) and the reference C atom of the aromatic ring, labeled as  $0^\circ$ . The reference C atom is the one bound to the protein backbone.

We observed the three amino acids (phenylalanine, tyrosine and tryptophane) separately because of the difference in their aromatic rings. Histidine was not taken into account, as it is often charged at physiological conditions, and this charge can shield delicate aromatic interactions.

We defined the  $\alpha$  face of the sugar monomer in this way: when the  $\alpha$  face is turned towards aromatic amino acid, the sum of distances between sugar ring C atoms and the aromatic center is smaller than the sum of distances between the sugar ring O atoms and the aromatic center. Homemade scripts for the search were written in Python (<http://www.python.org>) and MDAnalysis [32] python library was used for PDB file parsing. The structural alignment of different PDB structures was performed by a homemade script written in TCL language (<https://www.tcl.tk/>) embedded in VMD software [33]. We defined the  $\alpha$  face of the sugar monomer in this way: when the  $\alpha$  face is turned towards aromatic amino acid, the sum of distances between sugar ring C atoms and the aromatic center is smaller than the sum of distances between the sugar ring O atoms and the aromatic center, Figure S1.

## 2.2. Calculations

In order to evaluate the energies of carbohydrate/aromatic contacts with and without CH/ $\pi$  interactions, quantum-chemical calculations were performed using the MP2 method and the 6-31G(d,p) basis set in the Gaussian (version 09) program [34]. The BSSE correction was done [35]. The method and basis set have been chosen according to the methodology used in previous calculations in carbohydrate/aromatic contacts [24]. Model systems have been made from crystal structures by neutralizing monosaccharide units with H and OH group, and amino acids were patched with COOH and NH<sub>2</sub> groups. No geometry optimizations were made, hence the interaction energies were calculated on the geometries identical to the geometries in the crystal structures.



### 3. Results and discussion

#### 3.1. Geometry and type of carbohydrate/aromatic interactions

Using the criteria described in the Methodology section, number of aromatic/sugar contacts in the crystal structures from the PDB was 1913 (Table 1). The separate analyses of the three amino acids (phenylalanine, tyrosine and tryptophane) indicate that the tryptophane interactions are the most frequent (Table 1) because it contains two aromatic rings and also its six-membered ring has the most negative surface as shown in the calculated electrostatic potential [12,31]. The Table 2 shows the most frequent monosaccharides found in the structures, and the number of different kind of interactions for each one indicating that contacts of  $\beta$ -D-glucose (BGC) (39.7%) and  $\alpha$ -D-glucose (GLC) (38.1%) are the most frequent.

In large number of the contacts there are CH/ $\pi$  interactions; 1054 contacts (55%) have CH/ $\pi$  interactions, while 859 contacts are without CH/ $\pi$  interactions (45%). These data on large number of CH/ $\pi$  interactions is in accordance with an atomic force microscopic study suggesting that the CH/ $\pi$  interaction between the aromatic residues in chitinase from *Thermococcus kodakarensis* and pyranose ring of chitin/cellulose is the major interaction and that the side chain groups of the polysaccharide do not affect the binding event [26].

Table 1 shows also data on other types of the interactions that are possible between aromatic rings and sugar, classical hydrogen bonds, CH/N, OH/ $\pi$ , and CH/O interactions. In addition, there are some close contacts that do not satisfy criteria for any type of the interactions (CH/N, OH/ $\pi$ , CH/O or CH/ $\pi$  interactions, or hydrogen bonds). The mapping of contacts with these particular types of interactions is given in Figures S2-S5. In some of the contacts there is more than one type of the interactions, while

the number of contacts with combined interactions, shown in Table 1, indicates that simultaneous interactions are quite common with CH/ $\pi$  interactions.

**Table 1.** Number of interacting carbohydrate-aromatic pairs and the number of contacts with simultaneous interactions of different types. The percentages in the brackets are in relation to the total number of interactions (1913).

Number of interacting carbohydrate-aromatic pairs					Number of contacts with simultaneous interactions of given type			
	total	PHE	TYR	TRP	Classical hydrogen bonds	CH/N	OH/ $\pi$	CH/O
total number of contacts	1913	321 (16.8%)	612 (32.0%)	980 (51.2%)				
with CH/ $\pi$	1054 (55.1%)	117 (6.1%)	323 (16.9%)	614 (32.1%)	66 (3.5%)	286 (15.0%)	33 (1.7%)	40 (2.1%)
with classical hydrogen bonds	196 (10.2%)	0	87 (4.5%)	109 (5.7%)		77 (4.0%)	16 (0.8%)	63 (3.3%)
with CH/N	366 (19.1%)	0	0	366 (19.1%)			17 (0.9%)	21 (1.1%)
with OH/ $\pi$	53 (2.8%)	1 (0.1%)	11 (0.6%)	41 (2.1%)				1 (0.1%)
with CH/O	238 (12.4%)	37 (1.9%)	142 (7.4%)	59 (3.1%)				
other contacts	553 (28.9%)	169 (8.8%)	155 (8.1%)	229 (12.0%)				

**Table 2.** Number of interacting carbohydrate-aromatic pairs with different sugar units. The percentages in the brackets are in relation to the total number of interactions (1913).

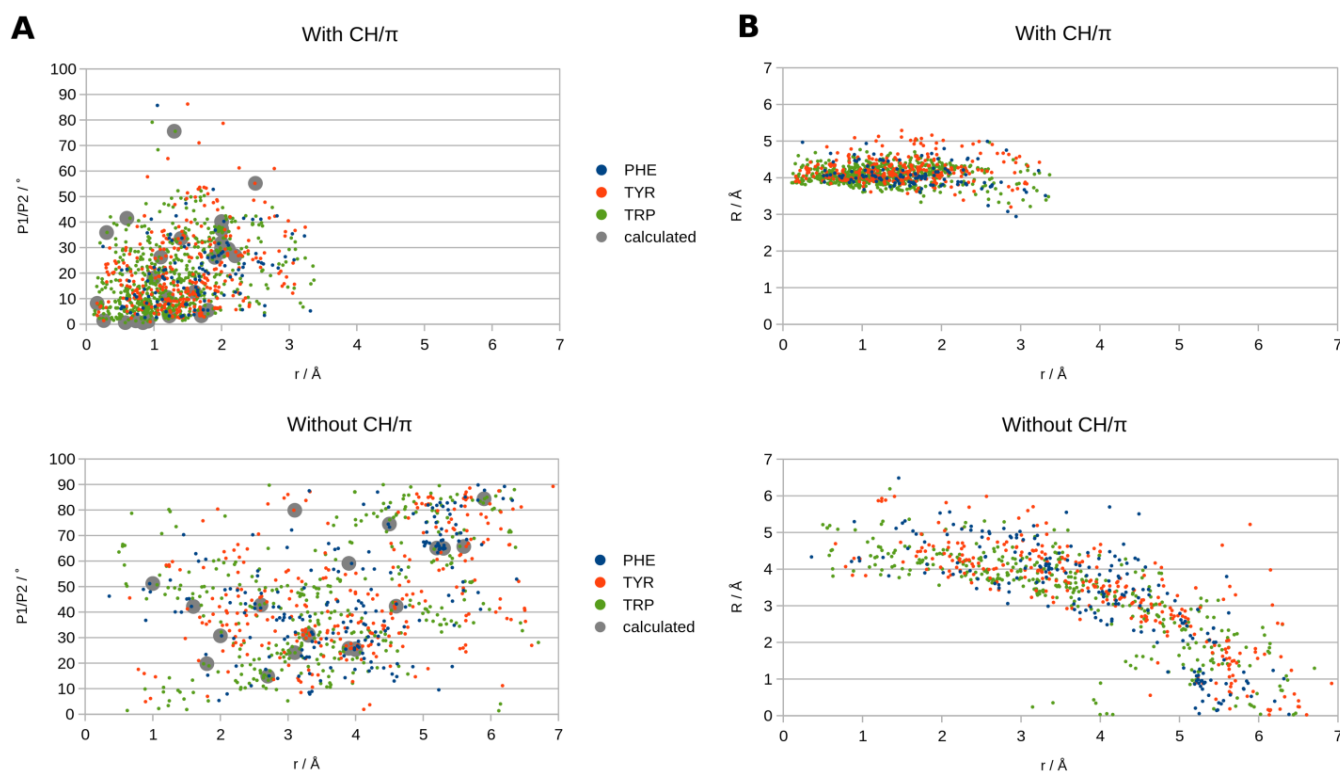
sugar unit	BGC	GLC	GAL	XYS	FUC	GLA	FUL	other
total number of contacts	759 (39.7%)	729 (38.1%)	97 (5.1%)	52 (2.7%)	51 (2.7%)	45 (2.4%)	30 (1.6%)	149 (7.8%)
with CH/ $\pi$	409 (21.4%)	386 (20.2%)	91 (4.8%)	21 (1.1%)	35 (1.8%)	34 (1.8%)	18 (0.9%)	59 (3.1%)
with classical hydrogen bonds	93 (4.9%)	67 (3.5%)	2 (0.1%)	4 (0.2%)	4 (0.2%)	3 (0.2%)	7 (0.4%)	16 (0.8%)
with CH/N	155 (8.1%)	89 (4.7%)	44 (2.3%)	5 (0.3%)	11 (0.6%)	18 (0.9%)	12 (0.6%)	30 (1.6%)
with OH/ $\pi$	21 (1.1%)	21 (1.1%)	2 (0.1%)	1 (0.1%)	0	0	4 (0.2%)	3 (0.2%)
with CH/O	81 (4.2%)	91 (4.8%)	5 (0.3%)	8 (0.4%)	12 (0.6%)	4 (0.2%)	3 (0.2%)	34 (1.8%)
other contacts	229 (12.0%)	233 (12.2%)	4 (0.2%)	22 (1.2%)	8 (0.4%)	5 (0.3%)	5 (0.3%)	47 (2.5%)

The number of the classical hydrogen bonds is relatively small, only 196 contacts are with classical hydrogen bonds between aromatic ring nitrogen or oxygen and carbohydrate oxygen atoms (Table 1). One can anticipate that small number of hydrogen bonds is caused by large tendency of OH groups to interact with polar amino acids or water. Large number of the classical hydrogen bonds are simultaneous with CH/ $\pi$  interactions, 66 of these interactions are simultaneous (Table 1).

The CH/N interactions are in 366 aromatic/sugar contacts (Table 1), while these interactions are mostly present in contacts simultaneously with CH/ $\pi$  interactions (286, Table 1). Analysis performed an OH/ $\pi$  interactions shows only 53 sugar-aromatic pairs that have OH/ $\pi$  interactions (Table 1), while 33 of them are simultaneous with CH/ $\pi$  interactions (Table 1). One can anticipate that small number of

OH/ $\pi$  interactions, similar to small number of hydrogen bond, is caused by OH group large tendency to interact with polar amino acids or water.

There are 238 aromatic/sugar contacts with CH/O interactions (Table 1). Relatively small number of contacts with simultaneous CH/ $\pi$  and CH/O interactions are present in the PDB (40 contacts), while there are no simultaneous CH/O interactions with any other type of the interactions (Table 1). Hence, CH/O interactions are mostly present in contacts without any other type of interactions. Contacts of  $\beta$ -D-glucose (BGC) (39.7%) and  $\alpha$ -D-glucose (GLC) (38.1%) are the most frequent in the total number of contacts and also in contacts with CH/ $\pi$  interactions (21.4% and 20.2%, respectively, Table 2). Since, from all studied interactions, CH/ $\pi$  interactions are the most frequent, we have split the contacts into two groups, with and without CH/ $\pi$  interactions, and we have analyzed geometric parameters separately for these two groups. Also we analyzed separately interactions for aromatic side chains of phenylalanine, tyrosine and tryptophane. One can notice that there is no clear difference between these three amino acid residues phenylalanine, tyrosine and tryptophane for both contacts with and without CH/ $\pi$  interactions (Figure 2). However, the two groups with and without CH/ $\pi$  interactions are quite different.



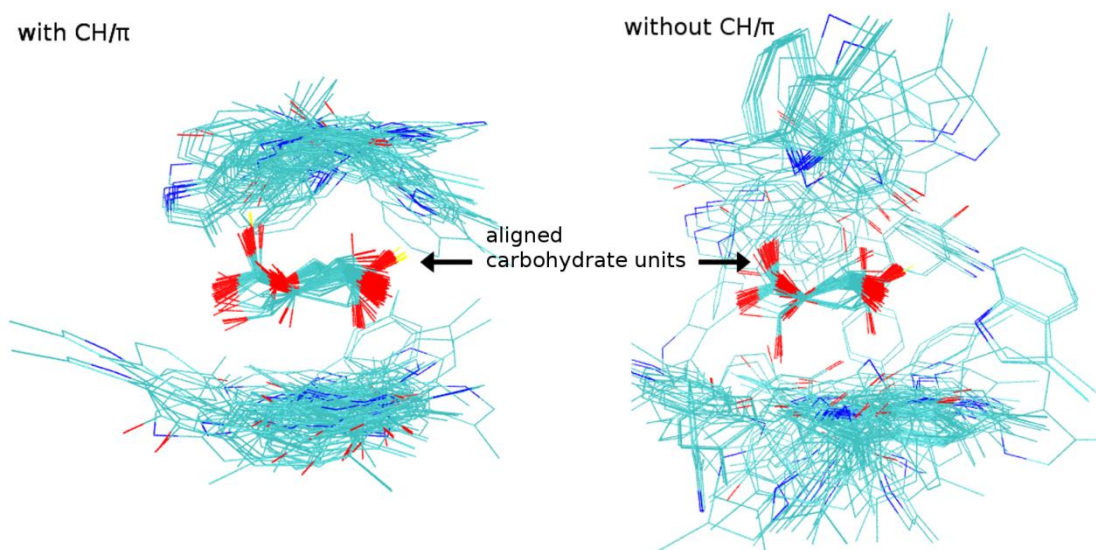
**Figure 2.** Monosaccharide/aromatic contacts with CH/ $\pi$  interactions and without CH/ $\pi$  interactions for phenylalanine, tyrosine and tryptophan found in PDB. **(A)** The angle between rings ( $P_1/P_2$ ) dependence on the offset ( $r$ ). **(B)** The normal distance ( $R$ ) dependence on the offset ( $r$ ). The contacts whose interaction energies were calculated are marked with grey circles.

The contacts with CH/ $\pi$  interactions shows strong tendency towards parallel arrangements with small offsets,  $P_1/P_2$  in the range from 0 to 20°, while offset  $r$  is in the range from 0.0 to 2.0 Å (Figure 2A). An example of the higher angle contact that is between deoxy sugar and the aromatic ring, is shown in Figure S6. This offset distribution agrees with the narrow offset distribution and energy curve for the benzene/cyclohexane interactions [36] that show maximum of frequency and energy for the offset below 2.0 Å. The graph  $R(r)$ , Figure 2B, confirms the sliding character of CH/ $\pi$  interactions described in previous works [19,25,27];  $r$  changes in the range of up to approximately 2.0 Å, while the normal distance tends to be around 4.0 Å. The offsets  $r$  has high tendency to be in the range between

0.0 and 2.0 Å corresponding with the sugar CH groups projection in the region of aromatic ring. The distribution of the number of CH/ $\pi$  interactions per sugar-aromatic pair demonstrates tendency towards only one or two CH/ $\pi$  bonds per sugar-aromatic pair (Figure S7).

The contacts without CH/ $\pi$  interactions show preferences towards high  $P_1/P_2$  angles and offsets (Figure 2A). Although the angles span from 0 to 90° and the offsets span from 0.0 to 7.0 Å, small number of contacts have small angles and small offsets. They are not horizontally sliding (the normal distance R drops with the increase of offset r), differently than for the contacts with CH/ $\pi$  interactions (Figure 2B).

The structural alignment of 200 random contacts with CH/ $\pi$  interactions and 200 random contacts without CH/ $\pi$  interactions found in PDB is shown in Figure 3. In accordance with the data in Figure 2, the data in Figure 3 also show that the interactions with CH/ $\pi$  are more localized, the aromatic rings are parallel to sugar, and sliding, while for the contacts that do not have CH/ $\pi$  interactions, the interactions are quite delocalized, with numerous interactions where two rings are not parallel.

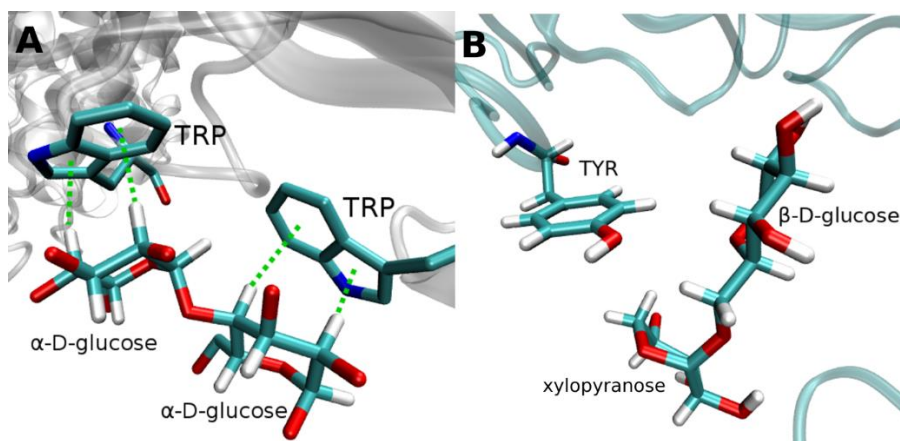


**Figure 3.** Structurally aligned carbohydrate units interacting with aromatic rings, 200 random structures found in PDB represented for the case with (left) and without CH/ $\pi$  interactions (right).

The sugars usually interact with aromatic rings only with their hydrophobic,  $\alpha$ , face exposing their polar groups to polar amino acids or solvent: we found 1772 such contacts out of total 1913. The preference for  $\alpha$  face is independent of amino acid type and number of aromatics per sugar. These results are in agreement with previous analysis [24,27,37]. Sandwich arrangements, where one sugar is in between two aromatic rings, are rare, one example is shown in Figure S8. The projection of the monosaccharide centers onto aromatic ring planes ( $r(\varphi)$  graphs in Figure S9) confirm the sliding character of these interactions: there is no clear specificity in offset nor  $\varphi$  angle.

Typical pyranose-aromatic interactions with CH/ $\pi$  and without CH/ $\pi$  are represented in Figure 4. In Figure 4A interactions of a disaccharide with two tryptophan rings are shown, where each tryptophan stacks with a glucose unit through two CH/ $\pi$  interactions. The rings are almost parallel ( $P_1/P_2 = 3.52^\circ$  and  $13.60^\circ$ ) with small offset ( $r = 1.20$  and  $0.23 \text{ \AA}$ ). Figure 4B shows a typical example of the carbohydrate - aromatic amino acid contact without CH/ $\pi$  interactions where Tyr ring interacts with a disaccharide. Not only that this contact is without CH/ $\pi$  interactions, but one cannot recognize neither

hydrogen bond, CH/N, OH/ $\pi$ , nor CH/O interactions. The rings are not parallel ( $P_1/P_2 = 65.74^\circ$  and  $72.60^\circ$ ), and offsets are relatively large ( $r = 5.64$  and  $6.15 \text{ \AA}$ ).



**Figure 4.** (A) A typical carbohydrate-aromatic interaction with CH/ $\pi$ , PDB ID: 1VEM. (B) A typical carbohydrate-aromatic interaction without CH/ $\pi$ , PDB ID: 5FKS. Green dotted lines represent CH/ $\pi$  interactions.

### 3.2. Protein/substrate pattern

We have performed an analysis of the number of aromatic amino acids per each full-length carbohydrate substrate, which can be monomer or oligomer. As one can anticipate, there is higher number of aromatics for longer substrates. The analysis shows that proteins arrange more or less evenly their aromatic rings onto every second carbohydrate monomer, since number of aromatics per monomer is in the range 0.39-0.73 (Table S1). For example, a six units long carbohydrate substrate interacts with three aromatic amino acids from the protein cleft (Figure S10). Some of these contacts are with CH/ $\pi$  interactions, while some of them are without CH/ $\pi$  interactions.

The data indicate that somewhat larger number of aromatic rings interact per monomer in contacts with CH/ $\pi$  interactions that is in accordance with large number of contacts with CH/ $\pi$  interactions (Table 1). If we take only aromatic rings which form CH/ $\pi$  interactions with sugar substrates, the



average pattern is approximately one aromatic ring on every third or fourth carbohydrate monomer unit, since number of aromatics per monomer is in the range 0.23-0.33, with exceptions for dimers and for octamers (Table S1). The average pattern is slightly different for contacts without CH/ $\pi$  interactions; the number of aromatics per monomer is in the range 0.15-0.22, with exception for monomers where the average is 0.36 (Table S1). It is worth to notice that only for monomers there is a larger number of contacts without CH/ $\pi$  interactions than with CH/ $\pi$  interactions. The branched or linear character of the substrate was not shown to play any role on the interaction pattern.

Furthermore, recent studies have suggested that aromatic amino acids might have a role in the stabilization of glycosyl cation transition states, through cation/ $\pi$  interactions [38,39]. Thus, we have performed a survey of the aromatic contacts at the catalytic subsite -1 in glycosidases, or in vicinity of the reducing end, i.e. the C1 carbohydrate atom [40]. As the cation transition state lifetime is extremely short ( $10^{-10}$  s to  $10^{-20}$  s [41]), it is hardly found in PDB, therefore we have only observed structures with resolved products after the hydrolysis event. We took the C1 atom which is up to 6 Å from another carbohydrate, and we counted aromatic rings at up to 6 Å near the C1 atom. The survey results show that a great deal of the structures (79.4%) possess at least one aromatic residue at the catalytic subsite -1, i.e. in position to interact with the cation transition state of carbohydrate via cation/ $\pi$  interactions.

### 3.3. Quantum chemical calculations of interaction energies

Previous calculations [19,24,25,27] and structural analysis [12] focus only on the carbohydrate/aromatic contacts through CH/ $\pi$  interactions. Previously calculated energies for the systems involving CH/ $\pi$  interactions are always stronger than -3.00 kcal/mol; the values are -3.0 kcal/mol for fucose/benzene calculated at MP2/6-31G(d,p) level [24], -3.54 to -5.40 kcal/mol for

various monosaccharides ( $\beta$ -d-glucopyranose,  $\beta$ -d-mannopyranose and  $\alpha$ -l-fucopyranose) with benzene [28] and -4.05 to -8.20 kcal/mol with naphthalene at CCSD(T)/CBS level [25]. Interactions of tryptophane with different monosaccharides were estimated to -3.2 to -6.4 kcal/mol at the MP2/6-311+G(d) level [42]; -2.4 to -5.2 kcal/mol at the MP2/6-31+G(d) level [43]; -3.2 to -8.2 kcal/mol at the MP2/6-311G++\*\* level [44]; and -0.8 to -7.5 kcal/mol at the MP2/6-311++G(d,p) level [45].

In order to assess strength of interactions observed in PDB, in contacts with CH/ $\pi$  interactions, as well as in contacts without the CH/ $\pi$  interactions, we have calculated interaction energies for the monosaccharide/aromatic contacts at the MP2 6-31G(d,p) level. We have chosen several typical geometries from the angle-offset graph (Figure 2A) as model systems for the three groups of contacts; two types of contacts with CH/ $\pi$  interactions, and contacts without CH/ $\pi$  interactions. Two types of contacts with CH/ $\pi$  interactions are favorable and interactions in which planes of the fragments (P1 and P2, Figure 1) are not parallel. Favorable CH/ $\pi$  interactions are defined by interring angles  $P_1/P_2$  smaller than  $10^\circ$  and offset values smaller than 2.0 Å, while non-parallel CH/ $\pi$  interactions are defined to be out of this range. The calculated energies of carbohydrate/aromatic contacts with CH/ $\pi$  interactions and without CH/ $\pi$  interactions are represented in Tables 3 and 4, respectively. Model systems are made from crystal structures, without geometry optimizations (Figure 2A).

**Table 3** Calculated energies of carbohydrate/aromatic contacts ( $\Delta E$ , in kcal/mol) with typical CH/ $\pi$  interactions (angles  $P_1/P_2$  (in  $^\circ$ ) smaller than  $10^\circ$  and offset values ( $r$ , in  $\text{\AA}$ ) smaller than  $2.0 \text{\AA}$ ). Monosaccharide and aromatic residues are given as residue name, residue index and chain as in PDB, and full name in brackets.

PDBid	monosaccharide	aromatic	$P_1/P_2$	$r$	$\Delta E$
4pey <sup>a,d</sup>	BGC703A ( $\beta$ -D-glucose)	TRP196A	1.2	0.9	-6.81
3wq1 <sup>c</sup>	BGC413A ( $\beta$ -D-glucose)	TRP84A	0.6	0.6	-6.38
1oh3 <sup>a,c</sup>	BGC1148A ( $\beta$ -D-glucose)	TRP26B	0.6	0.8	-4.35
2azd <sup>c</sup>	GLC996A ( $\alpha$ -D-glucose)	TYR256A	1.4	0.2	-4.15
4gx0 <sup>b</sup>	GLC604D ( $\alpha$ -D-glucose)	TRP46D	1.3	0.7	-3.68
2pc8 <sup>c</sup>	BGC402A ( $\beta$ -D-glucose)	PHE144A	5.5	1.8	-3.53
1qsg <sup>b</sup>	GLC1318F ( $\alpha$ -D-glucose)	PHE94F	3.2	1.2	-3.25
3wwg <sup>c</sup>	GLC1009B ( $\alpha$ -D-glucose)	TYR438B	8.2	0.2	-2.80

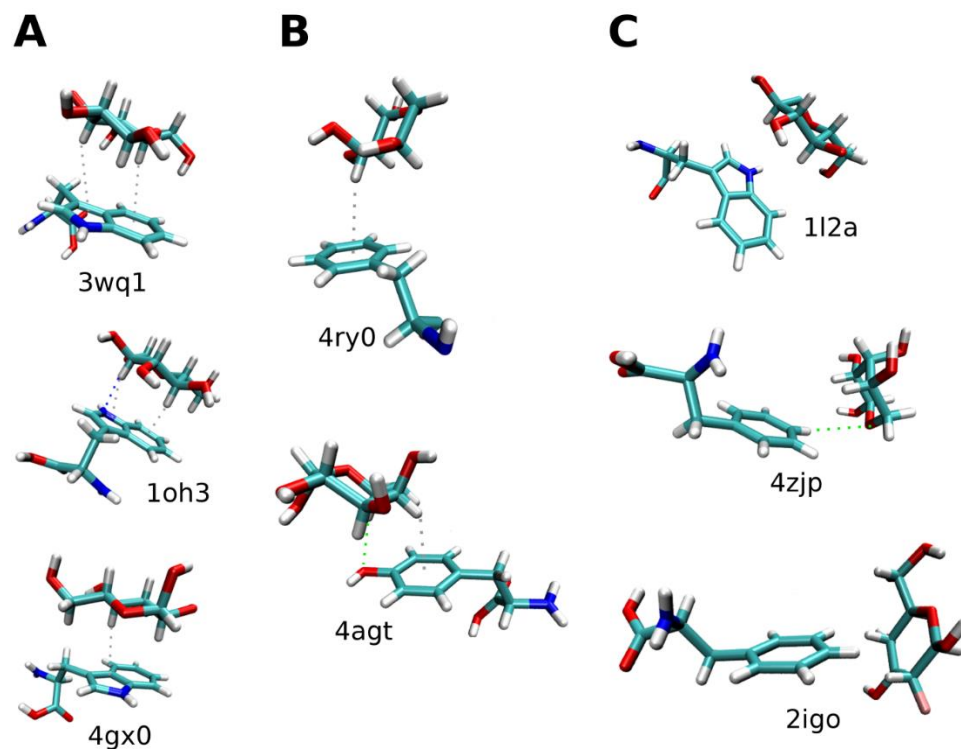
<sup>a</sup>Contacts with one CH/N interaction, <sup>b</sup>Contacts with one CH/O interaction, <sup>c</sup>Contacts with one hydrogen bond, <sup>d</sup>Contacts with one CH/ $\pi$  interaction

Interactions of systems with typical CH/ $\pi$  interactions are relatively strong (from -3.68 kcal/mol to -6.81 kcal/mol, Table 3), similarly to previously calculated carbohydrate/aromatic CH/ $\pi$  interactions, which are always stronger than -3.0 kcal/mol [19,24,25,28,42-45]. In some of the model systems used for the calculations interaction can be very strong, even if there are no additional interactions, as in the case of 3wq1 model system, with two CH/ $\pi$  interactions and interaction energy of -6.31 kcal/mol (Figure 5A, Table 3). Three examples of carbohydrate/aromatic contacts with typical CH/ $\pi$  interactions are shown in Figure 5A.

**Table 4.** Calculated energies of carbohydrate/aromatic contacts ( $\Delta E$ , in kcal/mol) with CH/ $\pi$  interactions and non-parallel planes (any geometries out of the range of angles P1/P2 (in  $^\circ$ ) smaller than  $10^\circ$  and offset values ( $r$ , in  $\text{\AA}$ ) smaller than 2.0  $\text{\AA}$ ). Monosaccharide and aromatic residues are given as residue name, residue index and chain as in PDB, and full name in brackets.

PDBid	monosaccharide	aromatic	P <sub>1</sub> /P <sub>2</sub>	r	$\Delta E$
5d61 <sup>a,f</sup>	GAL302A ( $\beta$ -D-galactose)	TRP35A	37.4	2.0	-4.34
1cqf <sup>f</sup>	GAL581E ( $\beta$ -D-galactose)	TRP534E	41.5	0.6	-4.23
1gu3 <sup>d</sup>	BGC601A ( $\beta$ -D-glucose)	TYR85A	3.3	1.7	-3.70
1v6l <sup>b,e</sup>	GAL235D ( $\beta$ -D-galactose)	TYR125D	33.6	1.4	-3.63
2bs6 <sup>e</sup>	FUC803A ( $\alpha$ -L-fucose)	TRP76A	26.3	1.1	-3.62
4xzs <sup>d</sup>	GLC501B ( $\alpha$ -D-glucose)	TYR155B	12.6	1.6	-3.53
3vgf <sup>f</sup>	GLC604A ( $\alpha$ -D-glucose)	TYR152A	18.1	1.0	-3.37
1ugy <sup>d</sup>	BGC201G ( $\beta$ -D-glucose)	TYR122G	26.8	2.2	-3.16
1lax <sup>a,c,d</sup>	GLC400A ( $\alpha$ -D-glucose)	TRP340A	10.5	1.2	-2.98
4zzu <sup>a,d</sup>	SGC1441A (4-deoxy-4-thio- $\beta$ -D-glucopyranose)	TRP38A	35.9	0.3	-2.79
4ry0 <sup>d</sup>	RIP401A (ribose)	PHE41A	40.2	2.0	-2.64
5a05 <sup>d</sup>	GLC1341E ( $\alpha$ -D-glucose)	PHE163E	26.2	1.9	-2.49
3wmp <sup>d</sup>	GLA103F ( $\alpha$ -D-galactose)	TYR79D	55.2	2.5	-2.41
5a05 <sup>d</sup>	GLC1341C ( $\alpha$ -D-glucose)	PHE163C	28.4	2.0	-2.19
4z4s <sup>e</sup>	FUL606A ( $\beta$ -L-fucose)	TRP381A	29.8	2.0	-2.06
3w27 <sup>d</sup>	XYS402A (xylose)	TRP300A	75.6	1.3	-2.05
1l2a <sup>d</sup>	BGC680C ( $\beta$ -D-glucose)	TRP439C	28.5	2.0	-2.04
1qot <sup>d</sup>	GAL502D ( $\beta$ -D-galactose)	TYR135D	37.0	2.0	-1.97
1l1y <sup>d</sup>	BGC680E ( $\beta$ -D-glucose)	TRP439E	31.8	2.0	-1.69
4agt <sup>b,d</sup>	FUC910A ( $\alpha$ -L-fucose)	TYR88A	29.4	2.1	-1.68

<sup>a</sup>Contacts with one CH/N interaction, <sup>b</sup>Contacts with one CH/O interaction, <sup>c</sup>Contacts with one hydrogen bond, <sup>d</sup>Contacts with one CH/ $\pi$  interaction, <sup>e</sup>Contacts with two CH/ $\pi$  interactions, <sup>f</sup>Contacts with three CH/ $\pi$  interactions



**Figure 5.** Examples of carbohydrate/aromatic contacts with (A) typical CH/π interactions, representing small offset and angle values (Table 3); with (B) CH/π interactions representing high offset and angle values (Table 4); and (C) without CH/π interactions (Table 5). CH/π interactions represented in gray dots, CH/N interaction in blue dots, and CH/O interactions in green dots.

The energies of interactions in systems with CH/π interactions and non-parallel geometries between two fragments are in the range from -1.68 kcal/mol to -4.34 kcal/mol (Table 4). The strongest non-parallel CH/π geometries can be comparable with the average interactions in systems with favorable CH/π interactions shown in Table 3 and with previously calculated [19,24,25]. Significant number of these interactions, even without additional interactions are relatively strong, they are stronger than -3.0 kcal/mol. Some examples of the model systems used to calculate interaction energies of the non-parallel CH/π interactions, given in Table 4, are shown in Figure 5B.

**Table 5.** Calculated energies ( $\Delta E$ ) (in kcal/mol) of carbohydrate/aromatic contacts without CH/ $\pi$  interactions;  $P_1/P_2$  (in  $^\circ$ ) is the angle between aromatic amino acid ring planes and carbohydrate ring mean planes;  $r$  (in  $\text{\AA}$ ) is the offset value between centers of amino acid and carbohydrate rings. Model systems were made from crystal structures (Figure 2A).

Monosaccharide and aromatic residues are given as residue name, residue index and chain as in PDB, and full name in brackets.

PDBid	monosaccharide	aromatic	$P_1/P_2$	$r$	$\Delta E$
1l2a	BGC679E ( $\beta$ -D-glucose)	TRP439E	14.9	2.7	-3.22
1uh4	GLC804A ( $\alpha$ -D-glucose)	TRP51A	19.8	1.8	-3.11
2c3w	GLC1104A ( $\alpha$ -D-glucose)	TRP20A	24.1	3.1	-2.45
1geg	GLC303B ( $\alpha$ -D-glucose)	PHE67B	30.7	2.0	-2.45
4gx1	GLC609C ( $\alpha$ -D-glucose)	PHE377C	42.2	1.6	-2.43
1l2a	BGC681C ( $\beta$ -D-glucose)	PHE206C	25.3	4.0	-2.31
5fks <sup>a</sup>	BGC1770A ( $\beta$ -D-glucose)	TYR298A	65.7	5.6	-2.18
2fn8	RIP303A (ribose)	PHE172A	51.2	1.0	-1.88
4zjp <sup>b</sup>	RIP301A (ribose)	PHE39A	59.1	3.9	-1.85
4bfn	GLC1109A ( $\alpha$ -D-glucose)	TYR83A	30.9	3.3	-1.57
1iuc <sup>b</sup>	FUC403A ( $\alpha$ -L-fucose)	TYR241A	31.3	3.3	-1.33
5ayi <sup>a</sup>	BGC503A ( $\beta$ -D-glucose)	TYR295A	25.6	4.0	-1.15
2igo	SHG808G (2-deoxy-2-fluoro- $\beta$ -D-glucose)	PHE474G	65.1	5.2	-0.86
4agt	FUC940B ( $\alpha$ -L-fucose)	TYR199B	79.9	3.1	-0.80
2yih	BGC1541A ( $\beta$ -D-glucose)	TYR234A	42.3	4.6	-0.77
4gvx <sup>c</sup>	FUL303D ( $\beta$ -L-fucose)	TYR191D	25.8	3.9	-0.76
2igo	SHG805E (2-deoxy-2-fluoro- $\beta$ -D-glucose)	PHE474E	65.0	5.3	-0.70
5cps	GLC708A ( $\alpha$ -D-glucose)	TRP546A	84.4	5.9	-0.64
4gbz	BGC501A ( $\beta$ -D-glucose)	PHE24A	74.5	4.5	-0.42
3l2m <sup>b</sup>	GLC803A ( $\alpha$ -D-glucose)	TRP134A	42.8	2.6	-0.17

<sup>a</sup>Contacts with two CH/O interactions, <sup>b</sup>Contacts with one CH/O interaction, <sup>c</sup>Contact with two hydrogen bonds

The calculated interactions energies for carbohydrate/aromatic contacts without CH/ $\pi$  interactions (according to criteria [12], given in Methodology section) are in the range from -0.17 kcal/mol to -3.22 kcal/mol (Table 5). The weakest calculated interaction is -0.17 kcal/mol (Table 5), while the significant number of contacts have interactions stronger than -2.0 kcal/mol. For the contacts without CH/ $\pi$  interactions, calculated interaction energies (Table 5) are somewhat weaker than contacts with non-parallel CH/ $\pi$  interactions (from -1.68 kcal/mol to -4.34 kcal/mol, Table 4), although the two ranges overlap significantly. As one can anticipate, the contacts without CH/ $\pi$  interactions are significantly weaker than the contacts with favorable geometries of CH/ $\pi$  interactions calculated in this work (from -3.68 kcal/mol to -6.81 kcal/mol, Table 3) and previously calculated CH/ $\pi$  interactions [19,24,25]. It is interesting that large number of contacts in Table 5 do not satisfy geometrical criteria for any of the interactions (classical hydrogen bond, CH/N, OH/ $\pi$ , or CH/O), while the interaction energies are significant (Table 5).

That suggests that the geometric criteria for CH/ $\pi$  interactions, as well as other interactions, classical hydrogen bond, CH/N, OH/ $\pi$ , or CH/O, do not include the contacts that can be relevant for carbohydrate/protein binding. Examples of carbohydrate/aromatic contacts without CH/ $\pi$  interactions are shown in Figure 5C. In two of the examples, 112a and 2igo, one cannot recognize any type of interaction (CH/ $\pi$ , classical hydrogen bond, CH/N, OH/ $\pi$ , or CH/O), although the interactions energies are -3.22 and -0.86 kcal/mol (Table 5).

#### 4. Conclusions

We analyzed geometries of carbohydrate/aromatic contacts in PDB and calculated interactions energies for those contacts. We presented the results on all contacts between aromatic rings and carbohydrate units (the distance less than to 4.0 Å between any aromatic ring heavy atom and any

heavy atom of a carbohydrate unit), including contacts without CH/ $\pi$  interactions, while previous studies on carbohydrate/aromatic interactions [19,24,25,27] focus on the CH/ $\pi$  interactions. Hence, in our analyses we have both contacts with and without CH/ $\pi$  interactions; there are 1054 contacts with and 859 without CH/ $\pi$  interactions in the PDB.

The analysis of the geometries of carbohydrate/aromatic contacts with CH/ $\pi$  interactions shows sliding character in the plane parallel to the ring plane. Namely, the normal distance tends to be around 4.0 Å, while the parallel displacement is in the range from 0.0 to up to 2.0 Å. On the other hand the contacts without CH/ $\pi$  interactions do not tend to be parallel and sliding in the ring plane and have much larger ranges of normal distances and parallel displacements. Analysis of the aromatic amino acid nature shows that it does not influence sugar/aromatic interaction geometry.

In the case of carbohydrate oligomers, proteins arrange more or less evenly their aromatic rings onto every second carbohydrate unit. This could be important for the enzymatic binding of a carbohydrate substrate prior to the hydrolysis reaction.

Our quantum chemical calculations at MP2/6-31G(d,p) level indicate significant attraction in carbohydrate/aromatic contacts, even when CH/ $\pi$  interactions are not present. In contacts with CH/ $\pi$  interactions the interaction energy is in the range from -1.68 kcal/mol to -6.81 kcal/mol, while in contacts with no CH/ $\pi$  interactions the energy is from -0.17 kcal/mol to -3.22 kcal/mol. These data show that contacts without CH/ $\pi$  interactions should not be neglected since their interaction energies are significant, sometimes even stronger than in contacts with CH/ $\pi$  interactions.

Our structural analysis and quantum chemical calculations pointed to the importance of carbohydrate/aromatic amino acid contacts without CH/ $\pi$  interactions in protein-carbohydrate recognition and binding: they are strong enough (although can be weaker than in systems with CH/ $\pi$  interactions) to hold the substrate in the protein cleft prior to catalysis, so they should not be neglected



in carbohydrate-protein recognition considerations. More importantly, the release of carbohydrate products of enzymatic reaction is easier when no typical CH/ $\pi$  interactions are present, since the binding is weaker.

## ASSOCIATED CONTENT

**Supplemental Information.** SI.doc

### **Acknowledgments and funding information**

This work was supported by the Serbian Ministry of Education, Science and Technological Development (grant 172065). The HPC resources and services used in this work were partially provided by the IT Research Computing group in Texas A&M University at Qatar. IT Research Computing is funded by the Qatar Foundation for Education, Science and Community Development (<http://www.qf.org.qa>).

### **Declaration of Interest**

The authors declare no conflicts of interest.

### **Author Contributions**

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

### **References**

- [1] Karlsson, K. A. Animal Glycosphingolipids as Membrane Attachment Sites for Bacteria. *Annu. Rev. Biochem.*, 58 (1), (1989), 309–350. <https://doi.org/10.1146/annurev.bi.58.070189.001521>.
- [2] Neufeld, E. F. Lysosomal Storage Diseases. *Annu. Rev. Biochem.*, 60 (1), (1991), 257–280.

- <https://doi.org/10.1146/annurev.bi.60.070191.001353>.
- [3] Dwek, R. A. Glycobiology: Toward Understanding the Function of Sugars. *Chem. Rev.*, *96* (2), (1996), 683–720.
- [4] Campbell, B. J.; Yu, L. G.; Rhodes, J. M. Altered Glycosylation in Inflammatory Bowel Disease: A Possible Role in Cancer Development. *Glycoconj. J.*, *18* (11–12), (2001), 851–858.
- [5] Dreitlein, W. B.; Maratos, J.; Brocavich, J. Zanamivir and Oseltamivir: Two New Options for the Treatment and Prevention of Influenza. *Clin. Ther.*, *23* (3), (2001), 327–355.
- [6] Haltiwanger, R. S.; Lowe, J. B. Role of Glycosylation in Development. *Annu. Rev. Biochem.*, *73* (1), (2004), 491–537. <https://doi.org/10.1146/annurev.biochem.73.011303.074043>.
- [7] Dzhambazov, B.; Holmdahl, M.; Yamada, H.; Lu, S.; Vestberg, M.; Holm, B.; Johnell, O.; Kihlberg, J.; Holmdahl, R. The Major T Cell Epitope on Type II Collagen Is Glycosylated in Normal Cartilage but Modified by Arthritis in Both Rats and Humans. *Eur. J. Immunol.*, *35* (2), (2005), 357–366. <https://doi.org/10.1002/eji.200425637>.
- [8] Wimmerová, M.; Kozmon, S.; Nečasová, I.; Mishra, S. K.; Komárek, J.; Koča, J. Stacking Interactions between Carbohydrate and Protein Quantified by Combination of Theoretical and Experimental Methods. *PLoS One*, *7* (10), (2012), 1–9. <https://doi.org/10.1371/journal.pone.0046032>.
- [9] Noll, A. J.; Yu, Y.; Lasanajak, Y.; Duska-McEwen, G.; Buck, R. H.; Smith, D. F.; Cummings, R. D. Human DC-SIGN binds specific human milk glycans. *Biochem. J.*, *473* (10), (2016), 1343–1353, <https://doi.org/10.1042/BCJ20160046>.
- [10] Elola, M. T.; Blidner, A. G.; Ferragut, F.; Bracalente, C.; Rabinovich, G. A. Assembly,

- organization and regulation of cell-surface receptors by lectin–glycan complexes. *Biochem. J.*, 469 (1), (2015), 1-16. <https://doi.org/10.1042/BJ20150461>.
- [11] Shinya, S.; Nishimura, S.; Kitaoku, Y.; Numata, T.; Kimoto, H.; Kusaoke, H.; Ohnuma, T.; Fukamizo, T. Mechanism of chitosan recognition by CBM32 carbohydrate-binding modules from a *Paenibacillus* sp. IK-5 chitosanase/glucanase. *Biochem. J.*, 473 (8), (2016), 1085-1095. <https://doi.org/10.1042/BCJ20160045>.
- [12] 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.*, 137 (48), (2015), 15152–15160. <https://doi.org/10.1021/jacs.5b08424>.
- [13] Bu, L.; Nimlos, M. R.; Shirts, M. R.; Ståhlberg, J.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T. Product Binding Varies Dramatically between Processive and Nonprocessive Cellulase Enzymes. *J. Biol. Chem.*, 287 (29), (2012), 24807–24813. <https://doi.org/10.1074/jbc.M112.365510>.
- [14] Prates, É. T.; Stankovic, I.; Silveira, R. L.; Liberato, M. V.; Henrique-Silva, F.; Pereira, N.; Polikarpov, I.; Skaf, M. S. X-Ray Structure and Molecular Dynamics Simulations of Endoglucanase 3 from *Trichoderma Harzianum*: Structural Organization and Substrate Recognition by Endoglucanases That Lack Cellulose Binding Module. *PLoS One*, 8 (3), (2013), e59069, <https://doi.org/10.1371/journal.pone.0059069>.
- [15] Zhang, H.; Zhang, J. L.; Sun, L.; Niu, X. Di; Wang, S.; Shan, Y. M. Molecular Dynamics Simulation of the Processive Endocellulase Cel48F from *Clostridium Cellulolyticum*: A Novel “Water-Control Mechanism” in Enzymatic Hydrolysis of Cellulose. *J. Mol. Recognit.*, 27 (7), (2014), 438–447. <https://doi.org/10.1002/jmr.2364>.

- [16] Payne, C. M.; Bomble, Y. J.; Taylor, C. B.; McCabe, C.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T. Multiple Functions of Aromatic-Carbohydrate Interactions in a Processive Cellulase Examined with Molecular Simulation. *J. Biol. Chem.*, 286 (47), (2011), 41028–41035. <https://doi.org/10.1074/jbc.M111.297713>.
- [17] Laughrey, Z. R.; Kiehna, S. E.; Riemen, A. J.; Waters, M. L. Carbohydrate- $\pi$  Interactions: What Are They Worth? *J. Am. Chem. Soc.*, 130 (44), (2008), 14625–14633. <https://doi.org/10.1021/ja803960x>.
- [18] Muehldorf, A. V.; Van Engen, D.; Warner, J. C.; Hamilton, A. D. Aromatic-Aromatic Interactions in Molecular Recognition: A Family of Artificial Receptors for Thymine That Shows Both Face-to-Face and Edge-to-Face Orientations. *J. Am. Chem. Soc.*, 110 (19), (1988), 6561–6562. <https://doi.org/10.1021/ja00227a045>.
- [19] Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J. Carbohydrate-Aromatic Interactions. *Acc. Chem. Res.*, 46 (4), (2013), 946–954. <https://doi.org/10.1021/ar300024d>.
- [20] Hsu, C.-H.; Park, S.; Mortenson, D. E.; Foley, B. L.; Wang, X.; Woods, R. J.; Case, D. A.; Powers, E. T.; Wong, C.-H.; Dyson, H. J.; et al. The Dependence of Carbohydrate–Aromatic Interaction Strengths on the Structure of the Carbohydrate. *J. Am. Chem. Soc.*, 138 (24), (2016), 7636–7648. <https://doi.org/10.1021/jacs.6b02879>.
- [21] Xu, X.; Liu, W.; Zhong, J.; Luo, L.; Liu, C.; Luo, S.; Chen, L. Binding interaction between rice glutelin and amylose:Hydrophobic interaction and conformational changes. *Int. J. Biol. Macromol.*, 81, (2015), 942–950. <https://doi.org/10.1016/j.ijbiomac.2015.09.041>.
- [22] Furtado, G. P.; Lourenzoni, M. R.; Fuzo, C. A.; Fonseca-Maldonado, R.; Guazzaroni, M.-E.;

- Ribeiro, L. F.; Ward, R. J. Engineering the affinity of a family 11 carbohydrate binding module to improve binding of branched over unbranched polysaccharides. *Int. J. Biol. Macromol.*, *120 B*, (2018), 2509–2516. <https://doi.org/10.1016/j.ijbiomac.2018.09.022>.
- [23] Loch, J. I.; Bonarek, P.; Polit, A.; Jabłoński, M.; Czub, M.; Ye, X.; Lewiński, K.  $\beta$ -Lactoglobulin interactions with local anaesthetic drugs—Crystallographic and calorimetric studies. *Int. J. Biol. Macromol.*, *80*, (2015), 87–94. <https://doi.org/10.1016/j.ijbiomac.2015.06.013>.
- [24] Fernández-Alonso, M. D. C.; Cañada, F. J.; Jiménez-Barbero, J.; Cuevas, G. Molecular Recognition of Saccharides by Proteins. Insights on the Origin of the Carbohydrate-Aromatic Interactions. *J. Am. Chem. Soc.*, *127* (20), (2005), 7379–7386. <https://doi.org/10.1021/ja051020+>.
- [25] Kozmon, S.; Matuška, R.; Spiwok, V.; Koča, J. Dispersion Interactions of Carbohydrates with Condensate Aromatic Moieties: Theoretical Study on the CH– $\pi$  Interaction Additive Properties. *Phys. Chem. Chem. Phys.*, *13* (31), (2011), 14215. <https://doi.org/10.1039/c1cp21071h>.
- [26] Kikkawa, Y.; Fukuda, M.; Kimura, T.; Kashiwada, A.; Matsuda, K.; Kanesato, M.; Wada, M.; Imanaka, T.; Tanaka, T. Atomic Force Microscopic Study of Chitinase Binding onto Chitin and Cellulose Surfaces. *Biomacromolecules*, *15* (3), (2014), 1074–1077. <https://doi.org/10.1021/bm500046f>.
- [27] 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/ $\pi$  Interactions in Water: Quantitative Structure–Stability Relationships for Carbohydrate/Aromatic Complexes. *Chem. Sci.*, *6* (11), (2015), 6076–6085. <https://doi.org/10.1039/C5SC02108A>.

- [28] Kozmon, S.; Matuška, R.; Spiwok, V.; Koča, J. Three-Dimensional Potential Energy Surface of Selected Carbohydrates' CH/ $\pi$  Dispersion Interactions Calculated by High-Level Quantum Mechanical Methods. *Chem. - A Eur. J.*, *17* (20), (2011), 5680–5690.  
<https://doi.org/10.1002/chem.201002876>.
- [29] Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-Chain Amide Orientation. *J. Mol. Biol.*, *285* (4), (1999), 1735–1747. <https://doi.org/10.1006/jmbi.1998.2401>.
- [30] Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; et al. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, *66* (2), (2010), 213–221. <https://doi.org/10.1107/S09074444909052925>.
- [31] Dragelj, J. L.; Stanković, I. M.; Božinovski, D. M.; Meyer, T.; Veljković, D. Z.; Medaković, V. B.; Knapp, E. W.; Zarić, S. D. C-H/O Interactions of Aromatic CH Donors within Proteins: A Crystallographic Study. *Cryst. Growth Des.*, *16* (4), (2016), 1948–1957.  
<https://doi.org/10.1021/acs.cgd.5b01543>.
- [32] Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J Comput Chem*, *32* (10), (2011), 2319–2327.  
<https://doi.org/10.1002/jcc.21787>.MDAnalysis.
- [33] Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.*, *14* (1), (1996), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [34] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G.

- Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Had, and D. J. F. Gaussian 09, Revision A.1. Gaussian, Inc.: Wallingford CT 2009.
- [35] Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.*, *19* (4), (1970), 553–566. <https://doi.org/10.1080/00268977000101561>.
- [36] Ninković, D. B.; Vojislavljević-Vasilev, D. Z.; Medaković, V. B.; Hall, M. B.; Brothers, E. N.; Zarić, S. D.; Pearson, J.; Taylor, R. Aliphatic–Aromatic Stacking Interactions in Cyclohexane–Benzene Are Stronger than Aromatic–Aromatic Interaction in the Benzene Dimer. *Phys. Chem. Chem. Phys.*, *18* (37), (2016), 25791–25795. <https://doi.org/10.1039/C6CP03734H>.
- [37] Simons, J. P.; Davis, B. G.; Cocinero, E. J.; Gamblin, D. P.; Stanca-Kaposta, E. C. Conformational Change and Selectivity in Explicitly Hydrated Carbohydrates. *Tetrahedron Asymmetry*, *20* (6–8), (2009), 718–722. <https://doi.org/10.1016/j.tetasy.2009.02.032>.
- [38] Iglesias-Fernández, J.; Hancock, S. M.; Lee, S. S.; Khan, M.; Kirkpatrick, J.; Oldham, N. J.; McAauley, K.; Fordham-Skelton, A.; Rovira, C.; Davis, B. G. A front-face "S<sub>N</sub>i synthase" engineered from a retaining "double-S<sub>N</sub>2" hydrolase. *Nat. Chem. Biol.*, *13* (8), (2017), 874–886. <https://doi.org/10.1038/nchembio.2394>.
- [39] Montalvillo-Jiménez, L.; Santana, A. G.; Corzana, F.; Jiménez-Osés, G.; Jiménez-Barbero, J.; Gómez, A. M.; Asensio, J. L. Impact of Aromatic Stacking on Glycoside Reactivity: Balancing CH/π and Cation/π Interactions for the Stabilization of Glycosyl-Oxocarbenium Ions. *J. Am. Chem. Soc.*, *141* (34), (2019), 13372–13384. <https://doi.org/10.1021/jacs.9b03285>.

- [40] Naumoff, D. G. Hierarchical Classification of Glycoside Hydrolases. *Biochemistry (Moscow)*, 76 (6), (2011), 622-635, link
- [41] Sinnott M. L. Catalytic Mechanisms of Enzymic Glycosyl Transfer. *Chem. Rev.*, 90 (7), (1990), 1171-1202, <https://doi.org/10.1021/cr00105a006>.
- [42] 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.*, 19 (12), (2006), 887–901. <https://doi.org/10.1007/s10822-005-9033-z>.
- [43] Spiwok, V.; Lipovová, P.; Skálová, T.; Buchtelová, E.; Hašek, J.; Králová, B. Role of CH/ $\pi$  Interactions in Substrate Binding by Escherichia Coli  $\beta$ -Galactosidase. *Carbohydr. Res.*, 339 (13), (2004), 2275–2280. <https://doi.org/10.1016/j.carres.2004.06.016>.
- [44] 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*, 44 (23), (2005), 8554–8562. <https://doi.org/10.1021/bi050298b>.
- [45] Sujatha, M. S.; Sasidhar, Y. U.; Balaji, P. V. MP2/6-311++G(d,p) Study on Galactose–Aromatic Residue Analog Complexes in Different Position-Orientations of the Saccharide Relative to Aromatic Residue. *J. Mol. Struct. THEOCHEM*, 814 (1–3), (2007), 11–24. <https://doi.org/10.1016/j.theochem.2007.02.032>.



Authorship statements

**Ivana M. Stanković:** Conceptualization; Data curation; Formal analysis; Investigation; Software; Visualization; Writing - original draft

**Jelena P. Blagojević Filipović:** Data curation; Formal analysis; Investigation; Visualization; Writing - original draft

**Snežana D. Zarić:** Conceptualization; Formal analysis; Funding acquisition; Methodology; Project administration; Resources; Supervision; Funding acquisition; Writing - review & editing

Journal Pre-proof

### Highlights

- The frequency of occurrence of the carbohydrate/aromatic contacts with and without CH/ $\pi$  interactions in the PDB structures indicate similar importance of interactions with and without CH/ $\pi$  interactions.
- Carbohydrate/aromatic contacts with CH/ $\pi$  interactions are preferentially parallel and have a sliding character in the plane parallel to the ring plane, while carbohydrate/aromatic contacts without CH/ $\pi$  interactions do not tend to be parallel and sliding in the ring plane and have much larger ranges of normal distances and parallel displacements.
- In the case of carbohydrate oligomers, proteins arrange more or less evenly their aromatic rings onto every second carbohydrate unit.
- The calculated interaction energies reveal that the interactions between carbohydrates and aromatic amino groups with geometries extracted from the PDB are significant (up to 3.2 kcal/mol) and should not be neglected in carbohydrate-protein recognition considerations.

Journal Pre-proof