Article

Facial Selectivity between Carbohydrates and Aromatic Amino Acids Explained by a Combination of NCI, NBO and EDA Techniques with NMR Spectroscopy

Thiago C. Rozada, [a] *Ulisses Z. de Melo,* [b] *Rodrigo M. Pontes,*[b] *Roberto Rittner*[c] *and Ernani A. Basso* [*,b]

*aDepartamento de Química, Universidade Estadual de Ponta Grossa,
Av. General Carlos Cavalcanti, 4748, Uvaranas, 84030-900 Ponta Grossa-PR, Brazil*

*bDepartamento de Química, Universidade Estadual de Maringá,
Av. Colombo, 5790, Jd. Universitário, 87020-900 Maringá-PR, Brazil*

*cInstituto de Química, Universidade Estadual de Campinas,
Cidade Universitária, 13083-872 Campinas-SP, Brazil*

The influence of electrostatic and dispersion components of intermolecular interactions on the recognition of carbohydrates by aromatic protein residues is important for many biological processes. Interactions between glucose and galactose and aromatic moieties of tryptophan, phenylalanine and histidine were investigated through ^{1}H nuclear magnetic resonance (NMR) chemical shift perturbation and fully explained by molecular modelling at the density functional theory (DFT) level. According to NMR experiments, aromatic amino acids interact preferably with one face of the carbohydrate and the calculations showed how intermolecular interactions were determinant in explaining the selectivity. Non-covalent interaction surfaces revealed that a CH bond oriented toward the center of the aromatic ring maximized the attractive interaction while minimizing the steric repulsion. Energy decomposition analyses showed that the dispersion component was stronger than the electrostatic component and contributes more when hydrogen bonds are not present in the studied complexes. However, it was the electrostatic component that correlated with the facial preference, especially for the complexes with tryptophan.

Keywords: carbohydrate, amino acid, intermolecular interactions, NMR, theoretical calculations, DFT, NBO, NCI, EDA, SAPT0

Introduction

Interactions between carbohydrates and proteins are very important to life, since they mediate a range of biological processes including allergenic reactions, embryogenesis, tissue development, fertilization, metathesis, cell wall recognition and hydration, and the stability of proteins.[1,2] Since these interactions are ubiquitous in biology and chemistry, the understanding of their nature and the factors that modulate their strength can lead to new therapeutics methods, supramolecular devices and advanced materials.

However, from a molecular point of view, the many factors involved in the recognition process are not fully understood.[3] Late in 1980's, X-ray analysis[4,5] revealed that the selectivity and specificity of the carbohydrate-protein interaction is due to a series of factors such as hydrogen bonding, interactions between the CH sugar bonds with aromatic residues, metal coordination and van der Waals forces. These interactions are relatively weak when compared with other biological interactions, with association constants on the order of 10^3-10^4 M^{-1},[6] but in the case of proteins and carbohydrates, the strength and specificity increase when the multiple interactions occur simultaneously.[3]

To date, several papers involving this topic have been published. Among all kinds of intermolecular interactions, the following two are recognized as most important in carbohydrate-protein pairs: the hydrogen bonding and the CH-π interaction. The importance of hydrogen bonding is well accepted and is due to the fact that polar and charged residues can interact with hydroxyl groups of carbohydrates.

CH-π interactions arise from the following two remarkable characteristics of each of these molecules:

the carbohydrates possess polarized CH bonds with lower electronic density on hydrogen atoms, and the aromatic moieties offer large electronic density in their π bonds. These two kinds of molecules often stack against each other in a contact ruled mostly by dispersive forces.[7] There is also an electrostatic component, but it is rather small when compared to the dispersion component.[8,9] Solvophobic effects must also be present in aqueous solution, because the carbohydrate molecule must displace the surrounding water molecules to interact with highly hydrophobic aromatic chains. Thus, since the solute-solvent interaction is rather strong, there must be a compensating driving force to accomplish it. There is an enhancement in the carbohydrate-aromatic stacking with an increase in sugar hydrophobicity through methylation, as observed by Morales and co-workers.[10]

The interaction of carbohydrates with model compounds, such as benzene,[11-13] toluene,[14] naphthalene,[15] *p*-hydroxy-toluene,[16] and 3-methylindole,[17] has been studied by both theoretical and experimental approaches, and it was determined that the CH-π interaction occurs and is more relevant when three CH bonds of the carbohydrate are oriented toward the aromatic system.

A mimetic experiment performed under molecular beam conditions by Simons and co-workers[14] detected complexes between monosaccharides and toluene. Under such conditions, these molecules cannot establish a hydrogen bonding network, so the detected complexes are maintained by CH-π interactions. This points out that this interaction is not necessarily driven by hydrophobic effects.[15]

Although tryptophan is the least abundant amino acid in the human proteome,[15] a study by Hudson *et al.*[18] revealed that it is the amino acid with the largest propensity to be proximal to the carbohydrate unity in crystal structures of protein-carbohydrate complexes. The next in this order of propensity are tyrosine and histidine. The fact that these three amino acids are aromatic is unquestionable evidence for the importance of the CH-π interaction in molecular recognition of carbohydrates.

The substituent effects on both the carbohydrate and amino acid molecules are relevant parameters, not only because of their stereochemical influence, but also because of a series of electronic effects. Barwell and Davis[19] found evidence that aromatic electronic π density is one key to modulating this interaction: a π-electron rich system can lead to stronger binding. A dynamic combinatorial approach reported by Asensio and co-workers[20,21] not only corroborated the proposal of Barwell and Davis, but also showed that the interaction can be modulated by substitution at the carbon atom, since the attachment of electron-withdrawing atoms to the CH bond is a tool used to increase the strength of the CH-π interaction.

Moreover, Asensio and co-workers work[20] showed that facial selectivity is highly dependent on the number of axial CH bonds in the carbohydrate moiety, while repulsive electrostatic and solvation effects play a minor role. These findings are in accordance with those of Hudson *et al.*,[18] who highlighted the relevance of CH bonds in forming a "non-polar patch", which can be interpreted as a group or surface of CH bonds close to each other that are able to interact with the aromatic π electron cloud.

The existence of this "patch" is directly related to the carbohydrate stereochemistry and, consequently, to anomeric and facial selectivity in the sugar binding process. Once the pyranosidic ring is considered as a kind of plane, we could think on two different faces in the carbohydrate ring and, therefore, one could expect some geometric arrangements in relation to the aromatic plane, as well. In the case of D-glucose (Glc) and D-galactose (Gal), for example, most of the CH bonds are axial, which contributes greatly to stacking via CH-π interactions. The stereochemical differences between Glc and Gal are only at the C4 position. In glucopyranose, all CH are axial (except the anomeric CH in the β anomer), and in galactopyranose, the C4H is equatorial. Only this difference is enough to cause significantly different interactions in geometry and strength. On this basis, factors can cause a protein to recognize the different carbohydrates in a selective way. Therefore, to better understand the recognition process as a whole, it is pivotal to elucidate this geometric dependence, as well as how it can influence intermolecular forces.

There are several papers applying theoretical methods (*ab initio*,[13] density functional theory (DFT),[17,22,23] molecular mechanics)[22] to understand the physical aspects involved in carbohydrate recognition by aromatic systems. However, to the best of our knowledge, none of them explore theoretical tools such as natural bond orbital (NBO), non-covalent interactions (NCI) or energy decomposition analyses (EDA) of the intermolecular interactions that are of unequivocal importance to an effective recognition process. Therefore, we would like to gain a deeper understanding of how non-covalent intermolecular interactions influence the recognition of carbohydrates by aromatic amino acids.

This paper employs high-level DFT methods to systematically optimize the complex geometry of both α and β anomers of D-glucose and D-galactose with L-tryptophan (Trp), L-phenylalanine (Phe) and L-histidine (His). All possible combinations of carbohydrate conformation and position of the amino acids were systematically evaluated. All complexes were then submitted to NCI, NBO, and EDA analysis, and the results

of all methods were carefully compared with ^{1}H nuclear magnetic resonance (NMR) chemical shift perturbation of the carbohydrate molecules. It is worth mentioning that this work deals with the carbohydrate in its natural anomeric form, in both experiment and theoretical approaches, without any conformational constrains, such as *O*-methylation.[23] By application of one-dimensional total correlation spectroscopy (TOCSY) techniques, it was possible to measure the ^{1}H resonances of each anomer. This paper aimed to provide a solid comparison among the stereoelectronic effects that are pivotal in the six different complexes studied, and how carbohydrate stereochemistry affects the strength and selectivity of the recognition process.

Experimental

NMR analyses

Carbohydrates, amino acids, and D_2O were purchased from Sigma-Aldrich and employed without further purification.

NMR experiments were performed on a Bruker Avance III HD nuclear magnetic resonance spectrometer operating at 500.13 MHz for the proton. The concentration employed for the carbohydrate was 2 mmol L^{-1}, while concentrations for the amino acids were 0, 2, 8, 16, and 32 mmol L^{-1} in 0.5 mL D_2O solutions in a 5 mm NMR tube at 25 °C.

A coaxial insertion tube with the sodium salt of 3-(trimethylsilyl)propionic-2,2,3,3-*d* acid in D_2O was employed as an external reference so that only the carbohydrate and the amino acid were in contact in the D_2O solution. ^{1}H NMR analyses were performed with 128k acquisition points, a spectral width of 5000 Hz, a recycling delay (D1) of 2.0 s, and 128 scans, employing a water pre-saturation pulse sequence (zgcppr).

The double-pulse field-gradient spin-echo (DPFGSE) 1D TOCSY[24] experiments were performed with the selective excitation of the anomeric hydrogen atom of the α and β anomers of the saccharides with the same conditions as the ^{1}H NMR analysis. Mixing times of 250 and 450 ms were employed for D-glucose and D-galactose, respectively. The power and duration of each pulse were optimized for each analysis.

Exponential apodization functions of 0.1 and 0.2 Hz were employed for the processing of the ^{1}H and 1D TOCSY spectra, respectively.

Theoretical calculations

The Gaussian 09[25] software package was employed for electron structure calculations and the NBO 5.9[26] program was used for the analysis involving NBO. Data for the non-covalent interactions (NCI) analysis was obtained with the program NCIPlot 3.0[27] and symmetry-adapted perturbation theory (SAPT) calculations were performed with Psi4.[28] The PyMOL[29] program was employed to visualize structures and surfaces.

The complexes formed by the carbohydrate and the amino acid were optimized with the M06-2X[30] method and the 6-31++G(d,p)[31] basis set functions in water with the integral equation formulation of the polarizable continuum model (IEF-PCM)[32] solvation method using the Bondi's radii for the description of the molecular cavities. The hybrid functional M06-2X is adequate for obtainment of thermodynamic and kinetic data of main group elements and for cases in which non-covalent interactions are important.[33] In addition, this method shows excellent performance for the energy evaluation of carbohydrates[34] with a favorable computational cost. Frequency calculations at the same theory level were performed to characterize the optimized structures as stationary points, and also to obtain the zero point energy correction (ZPE)[35] and thermal corrections. To precisely determine the energy of the complexes, the basis set superposition error (BSSE)[36] was obtained at the same level of theory of the optimization. NBO analysis was performed in vacuum at the M06-2X/6-31++G(d,p) level with the structures optimized in water at the same level of theory. The wave function utilized for the NCI was obtained at the M06-2X/6-31++G(d,p) level in water (IEF-PCM/Bondi).

SAPT analyses[37] were performed with the 6-31++G(d,p) basis set and density fitting with SAPT0.[38,39]

Results and Discussion

Study of the interaction between the carbohydrate and amino acid was performed through analysis of the ^{1}H NMR chemical shift perturbations and theoretical calculations. The first subsection presents the results of NMR analysis involving the monosaccharides D-glucose (Glc) and D-galactose (Gal) given in Figure 1 and the amino acids L-tryptophan (Trp), L-phenylalanine (Phe), and L-histidine (His). The next subsection presents the theoretical calculations for these complexes.

NMR analyses

The interaction between the carbohydrate and amino acid was investigated through the chemical shift perturbation of the carbohydrate hydrogen atoms due to the presence of an amino acid. At equilibrium in water,
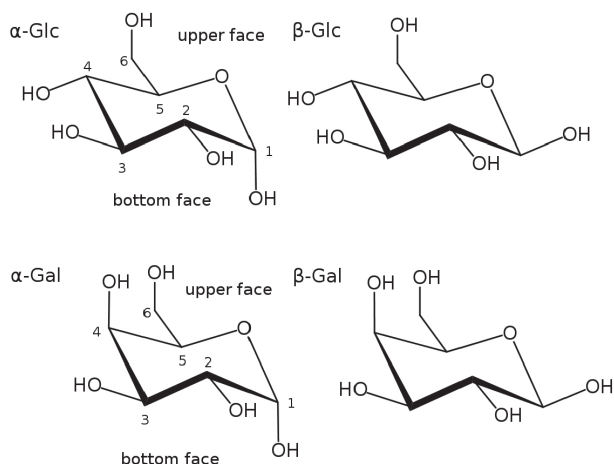
Figure 1. Structures of the monosaccharides studied in the $^4C_1$ pyranosidic conformation.

the monosaccharides D-glucose and D-galactose present two different pyranosidic anomers, α and β. To precisely determine the chemical shifts of each anomer and to avoid superposition with the amino acid signals, the 1D DPFGSE-TOCSY NMR[24] was used. This experiment can be used to analyze mixtures or identify impurities, substances with glycosidic or peptide units, overlapping signals, and $^{2,3}J_{CH}$ coupling constants.[40]

The ^1H chemical shift perturbation of each signal in the presence of amino acid was calculated as the difference in hertz (Hz) of the corresponding signal of the carbohydrate

without the presence of the amino acid (Tables S1-S8, Supplementary Information (SI) section). It was observed that the chemical shift perturbation for each carbohydrate's hydrogen atom had a linear dependency with the concentration of the amino acids studied (Figure 2). The angular coefficient of the linear regression of each ^1H chemical shift perturbation (multiplied by −100 for ease of visualization) indicates how the hydrogen atoms of the carbohydrate experienced the shielding effect due to the presence of the amino acid. Tables 1 and 2 show these coefficients for each anomer of Glc and Gal with the different amino acids. The higher the values, the higher the shielding effect experienced by the hydrogen atoms of the carbohydrate.

The data analysis showed that the ^1H chemical shift perturbation for each hydrogen atom of the carbohydrate was different. In general, the presence of the amino acid led to shielding of the hydrogen atoms of the carbohydrate. For α-Glc with Trp, this shielding effect was more pronounced for the hydrogen atoms on the upper face of the carbohydrate (H1, H2, and H4) than for the hydrogen atoms on the bottom face (H3 and H5). For β-Glc, the hydrogen atoms of the bottom face (H1, H3, and H5) showed a more pronounced shielding effect than the hydrogen atoms of the upper face (H2 and H4). This effect is consistent with the model in which the shielding cone generated by the aromatic group of the amino acid is positioned at each
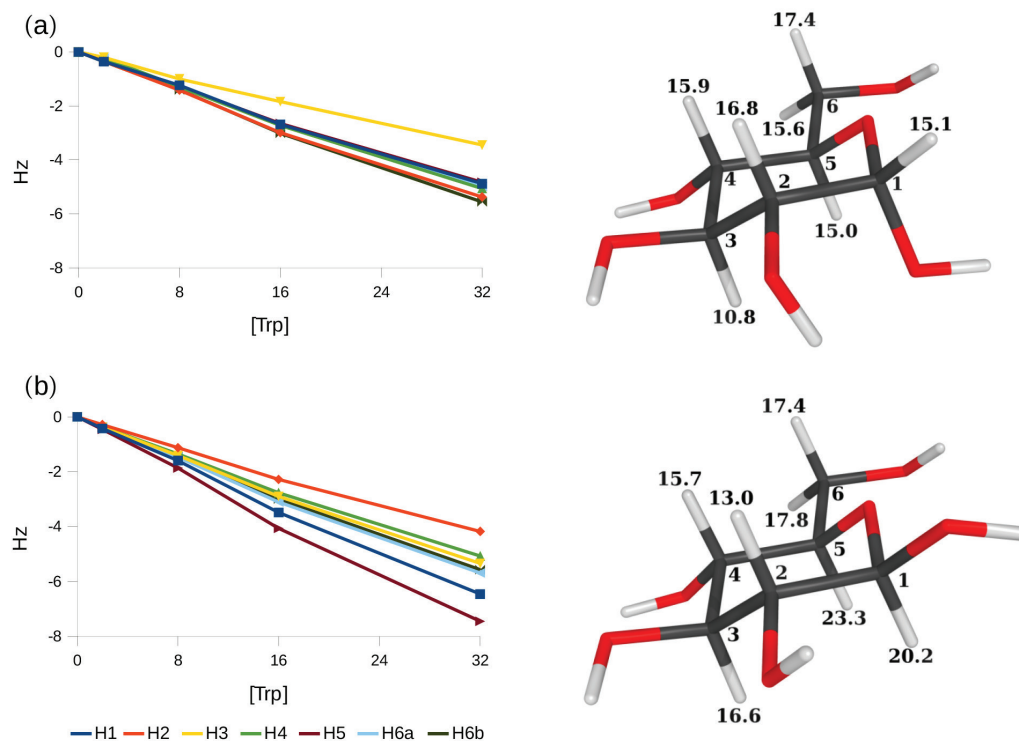

Figure 2. Chemical shift perturbation (Hz) and angular coefficient of the linear regression of each hydrogen atom (multiplied by −100) for (a) α-D-glucose and (b) β-D-glucose with L-tryptophan (mmol L⁻¹).

Table 1. Linear regression of the ^{1}H chemical shift perturbation of D-glucose with different amino acids and the correspondent coefficient of determination (R^2) value (between parentheses)

Complex	Hydrogen						
	H1	H2	H3	H4	H5	H6a	H6b
α-Glc-Trp	15.1 (0.9954)	16.8 (0.9961)	10.8 (0.9974)	15.9 (0.9972)	15.0 (0.9966)	15.6 (0.9966)	17.4 (0.9976)
β-Glc-Trp	20.2 (0.9972)	13.0 (0.9969)	16.6 (0.9973)	15.7 (0.9969)	23.3 (0.9970)	17.8 (0.9969)	17.4 (0.9981)
α-Glc-Phe	6.5 (0.9464)	6.7 (0.9379)	5.5 (0.9453)	6.6 (0.9630)	7.4 (0.9546)	7.2 (0.9360)	7.2 (0.9507)
β-Glc-Phe	7.3 (0.9305)	5.2 (0.9306)	6.0 (0.9028)	6.2 (0.9081)	7.9 (0.9210)	7.2 (0.9445)	7.0 (0.9298)
α-Glc-His	4.3 (0.9884)	3.5 (0.9607)	3.9 (0.9768)	4.2 (0.9907)	5.8 (0.8783)	5.0 (0.9864)	4.7 (0.9918)
β-Glc-His	4.3 (0.9757)	4.4 (0.9526)	3.7 (0.9375)	4.2 (0.9271)	4.2 (0.9697)	5.0 (0.9953)	4.7 (0.9630)

Glc: D-glucose; Trp: L-tryptophan; Phe: L-phenylalanine; His: L-histidine.

Table 2. Linear regression of the ^{1}H chemical shift perturbation of D-galactose with different amino acids and the correspondent coefficient of determination (R^2) value

Complex	Hydrogen						
	H1	H2	H3	H4	H5	H6a	H6b
α-Gal-Trp	16.3 (0.9952)	17.2 (0.9930)	25.2 (0.9948)	33.8 (0.9957)	27.6 (0.9967)	22.9 (0.9947)	
β-Gal-Trp	30.0 (0.9967)	17.2 (0.9979)	38.5 (0.9972)	37.4 (0.9965)	44.4 (0.9961)	23.4 (0.9986)	
α-Gal-Phe	5.3 (0.9942)	4.9 (0.9934)	6.5 (0.9962)	6.9 (0.9940)	7.6 (0.9964)	5.9 (0.9980)	6.2 (0.9955)
β-Gal-Phe	6.8 (0.9972)	4.5 (0.9855)	8.0 (0.9979)	8.0 (0.9973)	8.3 (0.9917)	6.7 (0.9962)	6.6 (0.9939)
α-Gal-His	5.4 (0.9773)	4.7 (0.9714)	5.6 (0.9752)	6.7 (0.9785)	6.6 (0.9879)	6.1 (0.9799)	6.4 (0.9811)
β-Gal-His	5.6 (0.9739)	4.2 (0.9749)	6.1 (0.9764)	6.5 (0.9714)	7.6 (0.9988)	6.4 (0.9919)	6.4 (0.9921)

Gal: D-galactose; Trp: L-tryptophan; Phe: L-phenylalanine; His: L-histidine.

face of the carbohydrate, but with a preference for one of the faces. That is, the hydrogen atoms on the face of the carbohydrate, in which the complex with the amino acid is more stable, suffer a greater shielding effect than the hydrogen atoms on the opposite side of the carbohydrate.

The H5 of α-Glc with Phe experienced the highest perturbation in ^{1}H chemical shift with an increase in the amino acid concentration. H6a and H6b also experienced pronounced shielding for this anomer. The other hydrogen atom of the bottom face of the carbohydrate, H3, showed the smallest shielding effect with a concentration of 32 mmol L^{-1} of L-phenylalanine. The hydrogen atoms of the upper face of α-Glc (H1, H2, and H4) showed an intermediate shielding effect, with H2 experiencing a greater ^{1}H chemical shielding perturbation and H1 the smallest perturbation. For β-Glc with Phe, the H1 and H5 of the bottom face experienced the larger shielding effect, followed by the methylene hydrogen atoms. H2 and H4 of the upper face, together with H3 of the bottom face, showed the smallest shielding effect.

With His, it was not observed a significant facial selectivity for α-Glc. For β-Glc, the hydrogen atoms of the bottom face presented a higher shielding effect than the hydrogen atoms of the upper face.

In general, the shielding effects in Glc were smaller with His than with Phe. This in turn produced a smaller shielding effect than with Trp. For α-Gal with Trp, the greater shielding effects were observed for H3–H6, with the greatest shielding effect for H4, while H1 and H2 showed the smallest shielding effects. The greater shielding effect for β-Gal was observed for H5, while H2 showed the smallest shielding effect. H1 showed a greater shielding effect in β-Gal than in α-Gal. The greater shielding effect observed for H3–H5 is consistent with the model in which the hydrogen atoms of the Gal are oriented toward the shielding cone generated by the indole group of Trp.

Both anomers of Gal showed a higher shielding effect on the hydrogen atoms of the bottom face with Phe. H1 showed a more pronounced shielding effect in β-Gal for those at an axial position and oriented in the bottom face compared to α-Gal, with H1 at an equatorial position at the upper face. The shielding effect order for Gal with His was similar to that observed with the other amino acids in which the hydrogen atoms of the upper face experienced a smaller effect than the hydrogen atoms of the bottom face. For Gal, it was also observed that the shielding effect decreased from Trp to Phe and to His. H4 and H5 showed

the highest shielding effect, and H2 of the upper face of the carbohydrate showed the smallest shielding effect.

The chemical shift perturbation observed for the studied complexes was parallel to the interactions of carbohydrates with proteins, as reported by Hudson *et al.*[18] The authors analyzed the position of aromatic residues of the amino acids around monosaccharides in the Protein Data Bank (PDB) and observed that there was a statistical preference between the faces of the carbohydrate, mainly for β-Gal and β-Glc.

A series of experiments with glycine (Tables S4 and S8, SI section) was performed to compare the effect of the absence of the aromatic group on the amino acid. These experiments showed very different results, with a smaller chemical shift perturbation than that induced by the aromatic amino acids. The linear regression of the chemical shift perturbation showed a maximum angular coefficient of 2.5 for Gly.

Theoretical calculations

The complexes formed by carbohydrates and amino acids were optimized at the M06-2X/6-31++G(d,p)[30] level in water with the continuum solvation model IEF-PCM.[32] Different orientations between the carbohydrate and amino acid were employed, always seeking an interaction of the amino acid aromatic group with the C–H bonds of the carbohydrate. That is, complex orientations with only hydrogen bonds between the amino or carboxylic group of the amino acid with the hydroxyl groups of the carbohydrate were not investigated. The three different conformations relative to the hydroxymethyl group of each carbohydrate anomer (*gauche-gauche*, *gauche-trans* and *trans-gauche*, dihedral angle of C6–OH with C5–O5 and C5–C4, respectively) (Figures S1, S2, SI section) were obtained after a conformational search involving a 2D potential energy surface with posterior optimization of the minimum structures found at the M06-2X/6-31++G(d,p) level in water (IEF-PCM, with the description of the molecular cavity by the radii of Bondi). See the SI section for more details of the carbohydrate conformational search. The same initial conformation of the amino acids was employed for all calculations performed.

D-Glucose-L-tryptophan

In the complexes between Glc and Trp the amino acid was positioned with its indole group over the bottom or upper face of the carbohydrate at the conformation $^4C_1$. The most stable complexes for both faces (bottom and upper) of both anomers (α and β) for each carbohydrate conformation (*gauche-gauche*, *gauche-trans* and *trans-gauche*) are shown in Table 3 (Table S11, SI section). Figure 3 shows the most stable complexes for anomers α and β with Trp at the bottom and upper face of the carbohydrate.

Table 3. Relative energy for the complexes between Glc and Trp (with ZPE and BSSE corrections) at the M06-2X/6-31++G(d,p) level in water (IEF-PCM/Bondi). Bottom and upper are based on the monosaccharide ring

entry	Complex	ΔE / (kcal mol^{-1})
1	α-Glc-gg_Trp_bottom	0.60
2	α-Glc-gg_Trp_upper	0.41
3	α-Glc-gt_Trp_bottom	0.33
4	α-Glc-gt_Trp_upper	0.00
5	α-Glc-tg_Trp_bottom	1.88
6	α-Glc-tg_Trp_upper	1.48
7	β-Glc-gg_Trp_bottom	0.19
8	β-Glc-gg_Trp_upper	0.71
9	β-Glc-gt_Trp_bottom	0.18
10	β-Glc-gt_Trp_upper	0.57
11	β-Glc-tg_Trp_bottom	0.94
12	β-Glc-tg_Trp_upper	1.56

Glc: D-glucose; Trp: L-tryptophan.

It was observed that α-Glc with Trp positioned at the upper face of the carbohydrate was more stable in all conformations, while for β-Glc the complexes were more stable with the amino acid positioned at the bottom face. Analyses of the aromatic residues in proteins around
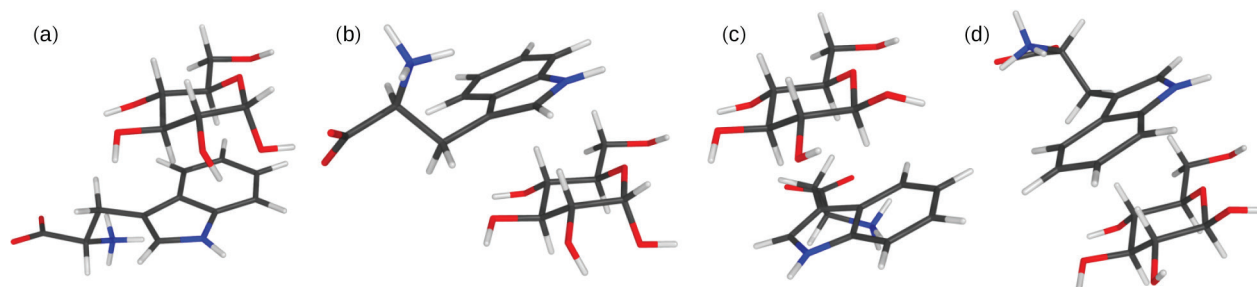

Figure 3. Most stable complexes between Glc and Trp. (a) α-Glc-gt_Trp_bottom; (b) α-Glc-gt_Trp_upper; (c) β-Glc-gt_Trp_bottom and (d) β-Glc-gt_Trp_upper. Bottom and upper are based on the monosaccharide ring.

α-Glc reported by Hudson *et al.*[18] also showed a preference for the upper face of the carbohydrate, while for β-Glc this preference was for the bottom face, with a 2.7-fold preference for aromatic residues over aliphatic residues.

The complex with the *gauche-trans* conformation of the carbohydrate was more stable for both anomers, followed by the *gauche-gauche* conformation. The *trans-gauche* conformation was the least stable for both anomers on the Glc-Trp complexes. For the α anomer, the α-Glc-gt_Trp_upper complex is the most stable, while for the β anomer the energy of β-Glc-gg_Trp_bottom and β-Glc-gt_Trp_bottom are equivalent (0.19 and 0.18 kcal mol^{-1}, respectively).

Surfaces representing the NCI[27] were calculated for the complexes between Glc and Trp in order to investigate what determines the observed facial preference. Figures 4 and 5 show the non-covalent interaction surfaces for the *gauche-trans* conformation of α and β anomers, respectively (see Figure S7, SI section, for the plot of

the reduced density gradient *versus* the electron density multiplied by the sign of λ_2; animated versions of the NCI surfaces are also available). The red surface is related to steric interactions, while the blue surface is related to strong stabilizing interactions such as hydrogen bonding, and the green surface is related to van der Waals interactions.

The α-Glc-gt_Trp_bottom complex (Trp was over the bottom face) showed very pronounced van der Waals interactions between H3 and H5 of the carbohydrate with the indole group of Trp. The hydrogen atoms were positioned approximately at the center of each one of the two rings that constitute the indole. The surface of non-covalent interactions from the amino acid side showed mostly steric interactions, while attractive van der Waals interactions predominated on the carbohydrate side of the NCI's surface, with repulsive interactions located in alignment with the carbon-carbon bonds of the indole. The carbohydrate was not completely aligned with the indole

Figure 4. Surface of non-covalent interactions for the complex of α-Glc-gt with Trp at the (a) bottom face and (b) upper face of the carbohydrate. Isosurface of 0.5 au. Color scale blue-green-red (–0.04 to 0.02 au) according to the sign of ($\lambda_2\rho$).
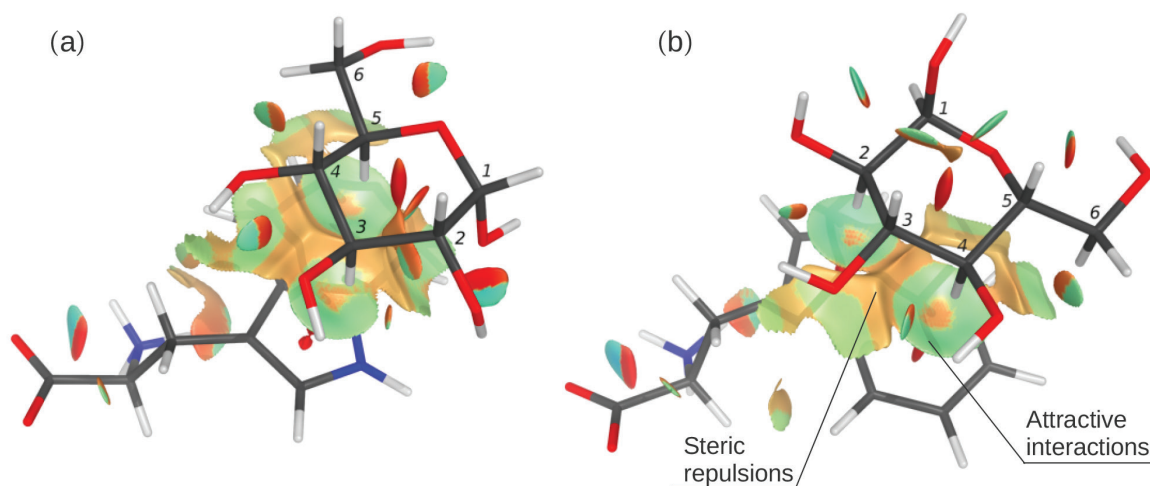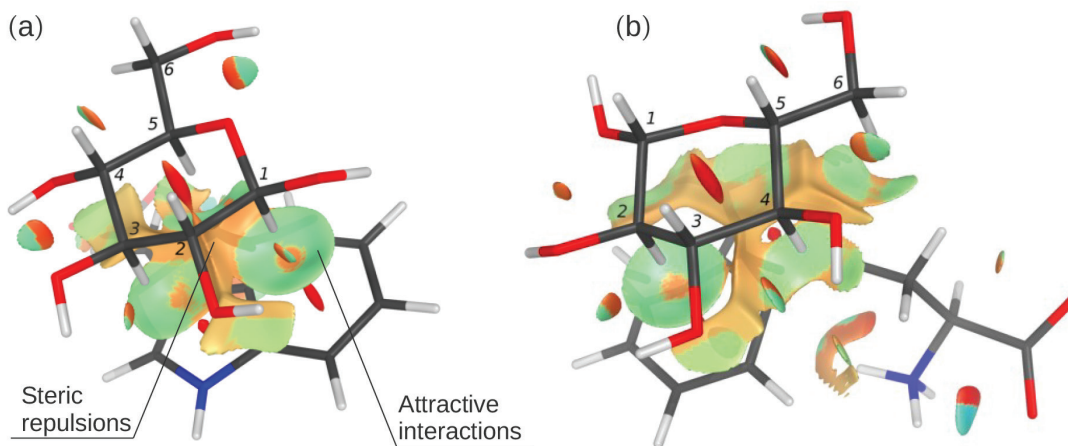

Figure 5. Surface of non-covalent interactions for the complex of β-Glc-gt with Trp at the (a) bottom face and (b) upper face of the carbohydrate. Isosurface of 0.5 au. Color scale blue-green-red (–0.04 to 0.02 au) according to the sign of ($\lambda_2\rho$).

group of the amino acid in this complex. The region around the anomeric hydroxyl group of the carbohydrate was further away from the amino acid than the region around the C4 carbon.

For the α-Glc-gt_Trp_upper complex (Trp was over the upper face), H2 and H4 of the carbohydrate were oriented in the direction of the center of the two rings that constitute the indole and presented more pronounced van der Waals interactions. Other featured interactions involved H6b. O3 and O5 also showed van der Waals interactions with the indole group. In this conformation, H1 in the equatorial position is turned away from the indole and did not show intermolecular interactions. Again, the NCI's surface showed steric interactions concentrated over the carbon-carbon bonds of the amino acid. In this complex, the carbohydrate was more aligned with the amino acid aromatic group than the α-Glc *gauche-trans* complex, with Trp at the bottom face, mainly due to interactions of H2, H4 and H6b with the amino acid.

The β-Glc-gt_Trp_bottom complex (Trp was over the bottom face) showed more pronounced van der Waals interactions for H1, H3, and H5 with H1 and H3, approximately centered in each one of the two rings that constituted the indole. Van der Waals interactions were also observed for O2 and O4. The carbohydrate was well-aligned with the amino acid and the distribution of the steric interactions on the non-covalent interaction surface was similar to that observed for the complex with α-Glc.

In relation to the β-Glc-gt_Trp_upper complex (Trp was over the upper face), the most pronounced van der Waals interaction involved H2, which was oriented toward the benzene ring of the indole. H4 and H6b also presented van der Waals interactions with the amino acid, but they were not oriented over the indole. O1, O3, and O5 also showed van der Waals interactions according to the NCI surface and showed steric interactions with the same pattern of the other Glc-Trp complexes.

The distance between the carbohydrate and the amino acid had a direct relationship with the intermolecular interactions. In the SI section, videos are provided to illustrate this relationship by showing how the NCI surfaces varied as a function of the intermolecular distance between carbohydrate and amino acid. When the intermolecular distance increased, smaller NCI surfaces were observed, leading to a slight increase in the relative energy. When the intermolecular distance diminished, the NCI surfaces became larger, but with a higher steric component, leading to a higher relative energy. The distance in the complexes studied was a balance between stabilizing and steric intermolecular interactions (see "Effect of the intermolecular distance between carbohydrate and amino

acid on the EDA" sub-section in the Supplementary Information for an EDA of the distance relationship between carbohydrate and amino acid).

NBO[26] theory was also employed to evaluate the stereoelectronic effects involved in the complex conformations. For Glc-Trp complexes, the stabilizing intermolecular orbital interactions (Table 4) were higher for the complex with Trp at the upper face of α-Glc and with Trp at the bottom face of β-Glc. The higher stabilizing energies involved the amino acid orbitals as donor orbitals and the carbohydrate σ^*_{C-H} as acceptors. The relative high energy of intermolecular orbital interactions with the carbohydrate orbitals acting as donors observed for the *gauche-gauche* conformation of both anomers of Glc with Trp at the upper face were the result of hydrogen bonding between η_{O6} and σ^*_{N-H}, observed only in these two complexes. This orbital interaction was 2.07 and 1.12 kcal mol^{-1} (Glc → Trp and Trp → Glc, respectively) for the complex with α-Glc and of 1.55 and 0.69 kcal mol^{-1} for the complex with β-Glc. Additional data involving the NBO analysis is available in the Supplementary Information (Table S12).

Table 4. Sum of the intermolecular orbital energy interactions obtained by NBO for the complexes between Glc and Trp calculated at the M06-2X/6-31++G(d,p) level

entry	Complex	Energy / (kcal mol^{-1})	
		Glc → Trp	Trp → Glc
1	α-Glc-gg_Trp_bottom	0.74	3.17
2	α-Glc-gg_Trp_upper	3.73	3.87
3	α-Glc-gt_Trp_bottom	0.69	3.05
4	α-Glc-gt_Trp_upper	0.19	4.84
5	α-Glc-tg_Trp_bottom	0.32	4.15
6	α-Glc-tg_Trp_upper	0.16	4.78
7	β-Glc-gg_Trp_bottom	0.91	5.76
8	β-Glc-gg_Trp_upper	2.74	4.23
9	β-Glc-gt_Trp_bottom	0.87	5.43
10	β-Glc-gt_Trp_upper	0.84	3.82
11	β-Glc-tg_Trp_bottom	0.39	5.30
12	β-Glc-tg_Trp_upper	0.20	5.37

Glc: D-glucose; Trp: L-tryptophan.

The stabilizing intermolecular energies between donor and acceptor orbitals overcome the destabilizing orbital interaction between the two occupied orbitals (Table 5). The higher the energy of the interaction between the two occupied orbitals, the higher the destabilizing effect. The most stable complexes are the ones where the carbohydrate and amino acid are closer to each other, consequently, presenting higher steric energy.

Table 5. Sum of the steric energy of the orbital interactions obtained by NBO for the complexes between Glc and Trp calculated at the M06-2X/6-31++G(d,p) level

| entry | Complex | Energy / (kcal mol⁻¹) | | | |
| | | Intramolecular | | Intermolecular | Σ[a] |
		Glc	Trp	Glc → Trp	
1	α-Glc-gg_Trp_bottom	406.70	493.21	10.94	910.85
2	α-Glc-gg_Trp_upper	404.22	492.52	13.84	910.57
3	α-Glc-gt_Trp_bottom	407.73	493.88	10.67	912.28
4	α-Glc-gt_Trp_upper	404.80	495.06	11.53	911.39
5	α-Glc-tg_Trp_bottom	407.91	494.53	10.47	912.91
6	α-Glc-tg_Trp_upper	406.17	492.75	12.07	910.99
7	β-Glc-gg_Trp_bottom	404.71	492.68	16.39	913.78
8	β-Glc-gg_Trp_upper	403.49	492.70	13.19	909.37
9	β-Glc-gt_Trp_bottom	405.51	493.44	14.84	913.78
10	β-Glc-gt_Trp_upper	404.26	494.36	11.38	910.00
11	β-Glc-tg_Trp_bottom	406.22	494.19	13.88	914.30
12	β-Glc-tg_Trp_upper	405.36	492.94	12.11	910.41

[a]Sum of all interactions. Glc: D-glucose; Trp: L-tryptophan.

The α-Glc-Trp complexes did not present significant differences in total steric energy in relation to which face the amino acid was positioned at. On the other hand, the β-Glc complex presented a more pronounced difference.

In general, NBO analyses showed that interactions in which the carbohydrate's orbitals acted as a donor in α-Glc were those involving lone pairs of the oxygen interacting with π* orbitals of the amino acid. For β-Glc, interactions involving the σ_{C-H} orbitals of the carbohydrate with the π* orbitals of the amino acid were also observed. The intermolecular orbital interactions observed with the amino acid orbitals acting as a donor were mainly between π orbitals of the amino acid with σ^*_{C-H} of the carbohydrate.

The theoretical calculations showed that the interaction of the carbohydrate's C–H bonds with the π system of the indole group of Trp was directly involved with the stability of the complexes. In addition, the intermolecular interactions observed were determinant in explaining the higher stability of the amino acid at one face of the carbohydrate for the Glc-Trp complexes since the more stable complex presented higher intermolecular interactions by both NCI surface and NBO analyses.

This preference of the amino acid for one of the faces of each anomer of the carbohydrate is in agreement with the shielding effect of Glc observed in the ¹H NMR chemical shift perturbation analyses with the Trp. For α-Glc, a higher stability of the complex with Trp at the upper face of the carbohydrate was observed, positioning H2 and H4 in the shielding cone of the indole group. Moreover, for β-Glc, the most stable complexes observed were those with Trp

positioned at the bottom face of Glc, directing H1, H3, and H5 to the shielding cone of the indole group.

To better understand the intermolecular interaction nature, SAPT0 calculations were performed for each complex. These calculations provide an energy decomposition analysis of the intermolecular interactions into the following components: electrostatic, exchange, induction, and dispersion (Table 6). It also gives a total energy interaction.

The total energy of the SAPT0 analyses showed that the complexes between α-Glc and Trp at the upper face of the carbohydrate presented higher stabilizing intermolecular interactions than the complexes with the amino acid at the bottom face. For β-Glc, the complexes with Trp at the bottom face presented higher total intermolecular energy interactions, with β-Glc-gg as an exception. This occurs due to a hydrogen bond interaction between O6 of the carbohydrate and N–H bond of the amino acid, increasing the electrostatic component in this complex.

It should be noted that the dispersion component was more pronounced than the electrostatic component in all complexes of Glc and Trp. On average, the electrostatic component was half the value of the dispersion component (−6.13 and −12.28 kcal mol⁻¹, respectively). The highest energy value was observed for the exchange component, with an average value of 12.70 kcal mol⁻¹, while the induction component presented the smallest average energy value (−1.74 kcal mol⁻¹).

β-Glc presented higher average energy values for each component than α-Glc. The C–H bonds of β-Glc were

Table 6. Energy decomposition analysis for the complexes between Glc and Trp through the SAPT0/6-31++G(d,p) calculation

entry	Complex	Component energy / (kcal mol^{-1})				
		Electrostatic	Exchange	Induction	Dispersion	Total
1	α-Glc-gg_Trp_bottom	−3.77	11.09	−1.27	−11.81	−5.75
2	α-Glc-gg_Trp_upper	−10.04	14.22	−2.35	−12.39	−10.56
3	α-Glc-gt_Trp_bottom	−4.04	10.91	−1.32	−11.96	−6.40
4	α-Glc-gt_Trp_upper	−5.61	12.02	−1.45	−11.74	−6.79
5	α-Glc-tg_Trp_bottom	−3.33	11.11	−1.69	−11.45	−5.36
6	α-Glc-tg_Trp_upper	−5.91	12.08	−1.50	−11.93	−7.26
7	β-Glc-gg_Trp_bottom	−6.63	14.61	−2.15	−12.34	−6.51
8	β-Glc-gg_Trp_upper	−9.39	13.91	−2.13	−12.65	−10.26
9	β-Glc-gt_Trp_bottom	−6.16	14.06	−2.00	−13.16	−7.26
10	β-Glc-gt_Trp_upper	−4.61	12.40	−1.47	−12.64	−6.32
11	β-Glc-tg_Trp_bottom	−7.00	13.53	−1.94	−12.99	−9.40
12	β-Glc-tg_Trp_upper	−6.02	12.45	−1.57	−12.28	−7.43

Glc: D-glucose; Trp: L-tryptophan.

closer in distance to the benzene ring of the indole than the bonds in α-Glc (2.57 and 2.35 Å on average for α-Glc and β-Glc, respectively). Regarding the five-membered ring of the indole, it was observed that the complexes of α-Glc presented smaller average distances between hydrogen atoms of the carbohydrates and the center of the five-membered ring than β-Glc (2.46 and 2.62 Å for α-Glc and β-Glc, respectively).

With respect to the complexes between α-Glc and Trp, it can be seen that there is a considerable variation in the total energy of SAPT0 depending on which face of the carbohydrate the amino acid was positioned. The complexes at the upper face had a more negative mean energy (−8.21 kcal mol^{-1}) than the complexes at the bottom face (−5.84 kcal mol^{-1}). This was mainly due to the difference between the average values of the electrostatic component (−3.71 and −7.19 kcal mol^{-1} for α-Glc_bottom and α-Glc_upper, respectively). The mean value of the exchange component was also higher for the complexes with the amino acid at the upper face of α-Glc, while the induction and dispersion components did not show such a pronounced variation between the two faces of α-Glc.

In general, the complexes between β-Glc and Trp with the amino acid at the bottom face present more negative total energy values with SAPT0 analysis. Again, we can observe a more pronounced difference in the electrostatic component than in the induction or dispersion components when comparing the different faces of the carbohydrate for each conformation, in a similar way to that observed in the case of the complexes with α-Glc, but this time favoring the complexes with the amino acid at the bottom face of the carbohydrate.

These results indicate that although the dispersion component has a greater influence on the interaction energy between the two molecules, it is the energy difference of the electrostatic component that is related to the preference for one of the faces of the carbohydrate.

The dipole moment for each C–H bond was calculated by NBO methodologies (Table S13, SI section) and showed that for α-Glc, both in the complexes and as an isolated carbohydrate, the C1–H, C3–H, and C4–H bonds presented the highest dipole moment. Of these three, both C1–H and C4–H are bonds at the upper face of the carbohydrate.

For β-Glc, the highest dipole moment was observed for the C–H bonds at the bottom face of the carbohydrate (C1–H, C3–H, and C5–H), both in the complexes and in the different conformations of the isolated carbohydrate.

With regard to the atomic charges of the hydrogen atoms H1–H5 of α-Glc and β-Glc complexes with Trp, no significant difference in charge was observed (Table S14, SI section). The isolated α-Glc presents a certain variation in the atomic charge for these atoms, but without any correlation to the different faces of the carbohydrate. Significant differences between the atomic charges of H1–H5 for the isolated β-Glc were not observed.

D-Glucose-L-phenylalanine

The phenyl ring of Phe is smaller than the indole ring of Trp and it does not cover the entire face of the carbohydrate as Trp does. The center of the aromatic ring of Phe was directed to each C–H bond of the carbohydrate and, after the optimizations, at least one complex with the amino acid aromatic ring centered in each hydrogen of the upper and bottom face of the carbohydrate was analyzed. The most

stable complex for each conformation of both anomers of Glc with the aromatic ring centered at H1–H6 are presented in Table 7 and can be visualized in the SI section (Figures S8-S13).

Table 7. Relative energy for the complexes between Glc and Phe (with ZPE and BSSE corrections) at the M06-2X/6-31++G(d,p) level in water (IEF-PCM/Bondi) and the sum of the intermolecular orbital interactions obtained by NBO calculated at the same level in vacuum

entry	Complex	ΔE / (kcal mol^{-1})	NBO / (kcal mol^{-1})	
			Glc $\rightarrow$ Phe	Phe $\rightarrow$ Glc
1	α-Glc-gg_Phe_H1	0.70	0.82	3.43
2	α-Glc-gg_Phe_H2	0.07	0.70	2.53
3	α-Glc-gg_Phe_H3	0.00	0.61	3.54
4	α-Glc-gg_Phe_H4	0.90	0.67	1.91
5	α-Glc-gg_Phe_H5	0.33	0.76	2.45
6	α-Glc-gg_Phe_H6	0.91	1.18	1.51
7	α-Glc-gt_Phe_H1	0.92	0.36	2.19
8	α-Glc-gt_Phe_H2	0.33	0.37	2.73
9	α-Glc-gt_Phe_H3	0.01	0.73	3.78
10	α-Glc-gt_Phe_H4	0.26	0.32	4.58
11	α-Glc-gt_Phe_H5	1.38	0.79	2.64
12	α-Glc-gt_Phe_H6	1.75	0.25	2.18
13	α-Glc-tg_Phe_H1	1.72	0.92	2.19
14	α-Glc-tg_Phe_H2	1.11	0.40	2.71
15	α-Glc-tg_Phe_H3	0.75	0.66	3.95
16	α-Glc-tg_Phe_H4	0.82	0.35	4.37
17	α-Glc-tg_Phe_H5	2.16	0.77	1.70
18	α-Glc-tg_Phe_H6	1.44	0.69	2.83
19	β-Glc-gg_Phe_H1	0.37	0.61	3.70
20	β-Glc-gg_Phe_H2	0.40	0.20	3.46
21	β-Glc-gg_Phe_H3	0.50	0.49	3.85
22	β-Glc-gg_Phe_H4	1.11	0.69	2.10
23	β-Glc-gg_Phe_H5	0.15	0.45	3.56
24	β-Glc-gg_Phe_H6	1.33	1.42	1.82
25	β-Glc-gt_Phe_H1	0.02	0.29	5.03
26	β-Glc-gt_Phe_H2	0.68	0.17	3.69
27	β-Glc-gt_Phe_H3	0.42	0.42	3.90
28	β-Glc-gt_Phe_H4	0.57	0.60	2.93
29	β-Glc-gt_Phe_H5	0.29	0.56	3.51
30	β-Glc-gt_Phe_H6	0.77	1.58	2.76
31	β-Glc-tg_Phe_H1	1.05	0.38	4.64
32	β-Glc-tg_Phe_H2	1.18	0.47	3.17
33	β-Glc-tg_Phe_H3	1.32	0.46	4.04
34	β-Glc-tg_Phe_H4	1.15	0.68	3.06
35	β-Glc-tg_Phe_H5	1.02	0.65	3.63
36	β-Glc-tg_Phe_H6	1.96	0.70	2.70

NBO: natural bond orbital; Glc: D-glucose; Phe: L-phenylalanine.

In general, the complexes of α-Glc-Phe with the *gauche-gauche* conformations were the most stable ones, followed by the *gauche-trans* conformations. The complexes with the *trans-gauche* conformation were the least stable ones. A large attractive intermolecular surface at the center of the ring of Phe, which was centered at one of the carbohydrate C–H bonds, was observed. This orientation allowed NCI surfaces to minimize the intermolecular steric interactions with the carbon skeleton of the benzene ring of the amino acid.

Intermolecular NCI surfaces involving the –CH$_2$– group of the amino acid were also observed and presented both steric and attractive features. The general steric behavior was similar to that observed for the complexes with Trp in which the steric component was mostly positioned above the carbon skeleton of the amino acid. Attractive intermolecular interactions involving the aromatic ring of Phe with other C–H bonds or hydroxyl groups of the carbohydrate were also observed. The most stable complexes were the ones that presented larger intermolecular van der Waals NCI surfaces.

For α-Glc, the complexes with the amino acid at the upper face of the carbohydrate presented great stability and similar behavior to the NMR results. The NMR data showed a pronounced chemical shift perturbation for one of the hydrogen atoms of the bottom face of α-Glc, with the other hydrogen atom showing the smallest perturbation. The hydrogen atoms of the upper face of the carbohydrate showed intermediate perturbations. However, the complexes with the amino acid ring centered at H3 were more stable in the calculations, whereas the other hydrogen atom of the bottom face of the carbohydrate (H5) showed the highest shielding effect.

The β-Glc complexes with Phe presented the following general conformation stabilization order: *gauche-trans*, *gauche-gauche*, and *trans-gauche*. The complexes with the amino acid at the bottom face of the carbohydrate were more stable, and the relative energy order was close to the observed ^{1}H chemical shift perturbation. The complexes with the aromatic ring centered at H1 and H5 were the most stable ones. The energy of the intermolecular orbital interactions was also higher for the complexes with the amino acid at the bottom face of β-Glc.

For the complexes between Glc and Phe in EDA through SAPT0 calculations (Table S17, SI section), a higher average value for the dispersion component than for the electrostatic component was observed. However, the difference between these two components was smaller than that observed for the complexes between Glc and Trp, mainly due to the smaller values of the dispersion component in the complexes with Phe.

A significant difference was not observed in the average interaction energy values calculated by SAPT0 for the complexes of α-Glc with Phe in relation to the different faces of the carbohydrate. However, for β-Glc, a significant difference favoring the amino acid at the bottom face was observed, as indicated by NMR analysis. Figures S14 and S15 (SI section) showed the individual components of EDA for the complexes of α-Glc and β-Glc, respectively. The electrostatic component presented a more pronounced variation than other components for each complex.

D-Glucose-L-histidine

The imidazole group of His is smaller than the aromatic groups of Phe and Trp and it does not cover the entire face of the carbohydrate. The behavior of the complexes between Glc and His was similar to that observed for Phe regarding the orientation of the center of the amino acid ring toward the C–H bonds of the carbohydrate at $^4C_1$ conformation. The most stable complexes for each conformation of both anomers of Glc with the ring centered at H1–H6 are presented in Table 8 and can be visualized in the Supplementary Information (Figures S16-S21).

Analysis of the behavior of the complexes with His presented an extra complicating factor: the intermolecular hydrogen bond involving the hydroxyl groups of the carbohydrate and the N–H bond of the imidazole ring or the carboxylic group.

In general, the most stable complexes presented hydrogen bonding. However, it could be noted by analyses of the NCI surfaces that the complexes with more intermolecular van der Waals interaction points were among the most stable ones, even surpassing some complexes with hydrogen bonding such as the case of α-Glc-gg. The most stable complexes with His also presented smaller steric surfaces.

The most stable complexes for α-Glc involved the amino acid at the bottom face of the carbohydrate and presented higher intermolecular orbital energy interactions mainly due to the hydrogen bonds between the lone pair of the nitrogen atoms of the imidazole ring and the hydroxyl groups of the carbohydrate ($\eta_N \rightarrow \sigma^*_{O-H}$) or between the oxygen lone pairs of the carbohydrate with the N–H bond in the imidazole ($\eta_O \rightarrow \sigma^*_{N-H}$). Hydrogen bonding involving the lone pairs of the amino acid carboxylic oxygen with the hydroxyl groups of the carbohydrate ($\eta_O \rightarrow \sigma^*_{O-H}$) were also observed for some complexes, due to the smaller size of the imidazole centered at the C–H bond of the carbohydrate, causing its carboxylic portion to get too close to the carbohydrate hydroxyl groups. Without the intermolecular hydrogen bond, the complexes with His at the upper face centered at H2 and H4 of α-Glc presented more prominent intermolecular interactions.

Table 8. Relative energy for the complexes between Glc and His (with ZPE and BSSE corrections) at the M06-2X/6-31++G(d,p) level in water (IEF-PCM/Bondi) and the sum of the intermolecular orbital interactions obtained by NBO calculated at the same level in vacuum

entry	Complex	ΔE / (kcal mol⁻¹)	NBO / (kcal mol⁻¹)	
			Glc → His	His → Glc
1	α-Glc-gg_His_H1	5.06	0.09	1.85
2	α-Glc-gg_His_H2	3.57	13.10	3.10
3	α-Glc-gg_His_H3	2.43	0.92	9.65
4	α-Glc-gg_His_H4	3.62	1.06	11.20
5	α-Glc-gg_His_H5	3.27	1.14	2.43
6	α-Glc-gg_His_H6	3.47	1.59	2.65
7	α-Glc-gt_His_H1	6.27	0.23	2.22
8	α-Glc-gt_His_H2	4.37	2.11	2.98
9	α-Glc-gt_His_H3	3.85	1.03	10.12
10	α-Glc-gt_His_H4	4.91	1.30	3.05
11	α-Glc-gt_His_H5	4.80	1.55	3.02
12	α-Glc-gt_His_H6	1.53	2.03	31.24
13	α-Glc-tg_His_H1	6.17	0.15	1.96
14	α-Glc-tg_His_H2	4.43	1.58	2.94
15	α-Glc-tg_His_H3	3.43	0.88	10.36
16	α-Glc-tg_His_H4	5.16	0.25	3.22
17	α-Glc-tg_His_H5	2.33	3.07	26.98
18	α-Glc-tg_His_H6	5.34	0.83	1.15
19	β-Glc-gg_His_H1	4.26	0.47	3.84
20	β-Glc-gg_His_H2	3.38	11.89	2.90
21	β-Glc-gg_His_H3	0.00	2.32	40.48
22	β-Glc-gg_His_H4	2.17	2.70	21.76
23	β-Glc-gg_His_H5	4.20	0.81	2.97
24	β-Glc-gg_His_H6	5.26	0.62	1.91
25	β-Glc-gt_His_H1	5.28	0.81	3.36
26	β-Glc-gt_His_H2	4.75	2.04	2.78
27	β-Glc-gt_His_H3	0.62	2.24	39.41
28	β-Glc-gt_His_H4	4.74	1.18	3.75
29	β-Glc-gt_His_H5	5.17	1.33	4.35
30	β-Glc-gt_His_H6	1.58	1.46	37.86
31	β-Glc-tg_His_H1	5.22	0.60	3.31
32	β-Glc-tg_His_H2	2.37	4.65	36.38
33	β-Glc-tg_His_H3	1.18	2.23	41.11
34	β-Glc-tg_His_H4	2.70	2.89	22.32
35	β-Glc-tg_His_H5	4.94	2.05	2.75
36	β-Glc-tg_His_H6	6.23	0.33	1.45

NBO: natural bond orbital; Glc: D-glucose; His: L-histidine.

For complexes with His, it was not possible to obtain any correlations through energy decomposition analyses obtained by SAPT0 (Table S20, SI section) due to a

greater presence of hydrogen bonding. These interactions lead to distortions in the total interaction energy, mainly due to the electrostatic component, which presents a more pronounced energy when there is intermolecular hydrogen bonding.

The complexes with His, in comparison with the complexes with Phe, presented less interactions with the amino acid's ring besides one of the C–Hs bonds of the carbohydrate oriented toward the center of the imidazole ring. However, the complexes with His presented carbohydrate interactions with the –CH– group of the amino acid that was absent in the other complexes.

The calculations did not indicate a clear preference of His for one of the faces of β-Glc, similar to the chemical shift perturbations.

D-Galactose-L-tryptophan

For complexes between Gal and Trp, it was observed that the amino acid positioned at the bottom face of both carbohydrate anomers (α and β) was more stable (Table 9). This is in accordance with what was observed in the carbohydrate-protein interactions analyzed by Hudson *et al.*[18] The *gauche-trans* conformation with the amino acid at the bottom face of the carbohydrate was the most stable conformation for both α and β anomers, but for the complexes with the amino acid at the upper face of the carbohydrate the relative energy difference between the *gauche-gauche* and *gauche-trans* conformations was

small. In all cases, the *trans-gauche* conformer was the least stable. Figure 6 shows the most stable complexes of each anomer of Gal with *gauche-trans* conformation and the amino acid positioned at both faces of the carbohydrate.

Table 9. Relative energy for the complexes between Gal and Trp (with ZPE and BSSE corrections) at the M06-2X/6-31++G(d,p) level in water (IEF-PCM/Bondi). Bottom and upper are based on the monosaccharide ring

entry	Complex	ΔE / (kcal mol^{-1})
1	α-Gal-gg_Trp_bottom	0.36
2	α-Gal-gg_Trp_upper	1.50
3	α-Gal-gt_Trp_bottom	0.00
4	α-Gal-gt_Trp_upper	1.56
5	α-Gal-tg_Trp_bottom	1.27
6	α-Gal-tg_Trp_upper	2.58
7	β-Gal-gg_Trp_bottom	0.57
8	β-Gal-gg_Trp_upper	2.11
9	β-Gal-gt_Trp_bottom	0.25
10	β-Gal-gt_Trp_upper	2.01
11	β-Gal-tg_Trp_bottom	1.38
12	β-Gal-tg_Trp_upper	3.22

Gal: D-galactose; Trp: L-tryptophan.

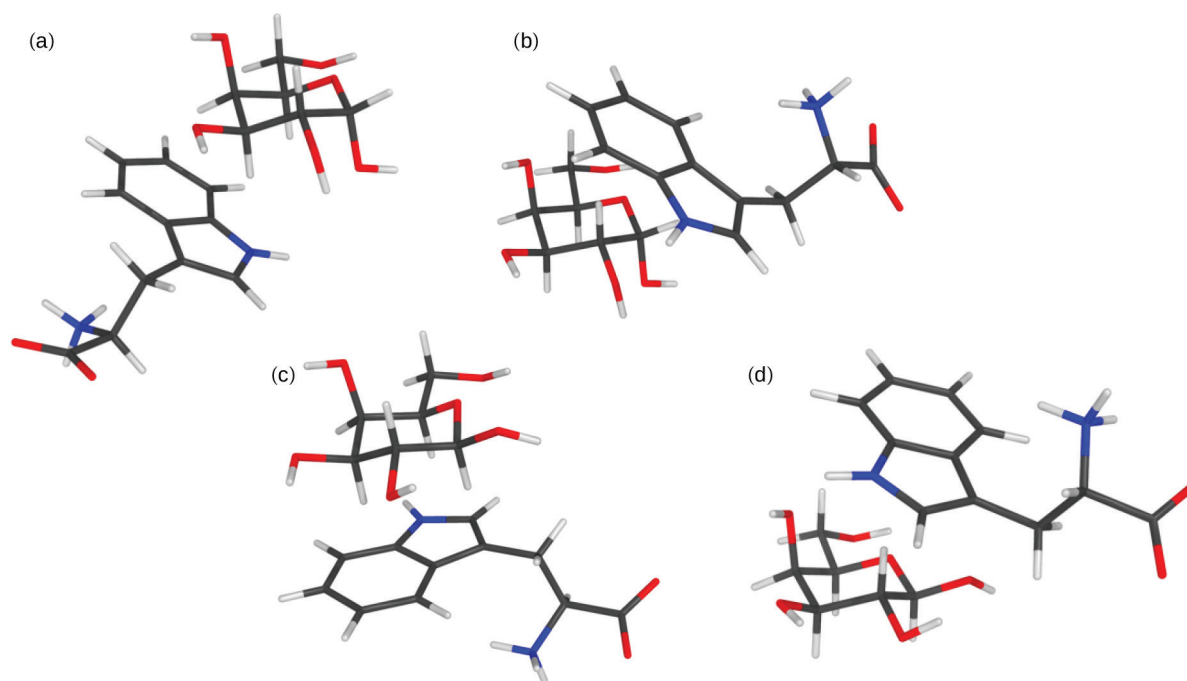It was observed that the Trp at the bottom face of α-Gal assumed a position in which the indole ring was aligned

Figure 6. Most stable complexes between Gal and Trp. (a) α-Gal-gt_Trp_bottom; (b) α-Gal-gt_Trp_upper; (c) β-Gal-gt_Trp_bottom and (d) β-Gal-gt_Trp_upper. Bottom and upper are based on the monosaccharide ring.

with H3–H5 of the carbohydrate (Figure 6a). In the case of Trp at the upper face of the carbohydrate, the alignment observed was with H1 and H2 (Figure 6b).

In the complex of β-Gal with the Trp at the bottom face, the carbohydrate and the indole group were well aligned, with the H1, H3 and H5 oriented toward the indole group (Figure 6c). However, for the complex with the amino acid at the upper face of the carbohydrate, only a small alignment with H2 and O4–H was observed.

The analyses of the non-covalent interactions for the complexes of *gauche-trans* α-Gal and *gauche-trans* β-Gal with Trp at both faces of the carbohydrate allow for a better understanding of the orientations of the amino acid in relation to the carbohydrate presented in these complexes (Figures 7 and 8 for anomers α and β, respectively; see Figure S22, SI section, for the plot of the reduced density gradient *versus* the electron density multiplied by the sign of λ_2; animated versions of the NCI surfaces are also available for better visualization).

In the case of α-Gal with Trp at the bottom face of the carbohydrate three points of van der Waals interactions involving H3, H4, and H5 of the carbohydrate with the amino acid were observed, with H3 and H4 facing toward the center of one of the rings of the indole group. Other less effective van der Waals intermolecular interactions involving H6a, O1, and O3 with the amino acid were also observed. On the other hand, the NCI surface with Trp at the upper face of α-Gal showed an intermolecular interaction involving only H1 and H2, aside from O4 in the *gauche-trans* conformer.

A large difference in the NCI surface between Trp at the bottom and upper face of the carbohydrate was observed for β-Gal. For the first case, strong intermolecular van der Waals interactions involving H1, H3, and H5 of the carbohydrate with the amino acid were observed. H3 and H5 were oriented toward the centers of the two rings of the amino acid indole group for this anomer. For the second case, the NCI surface observed was quite fragmented, with
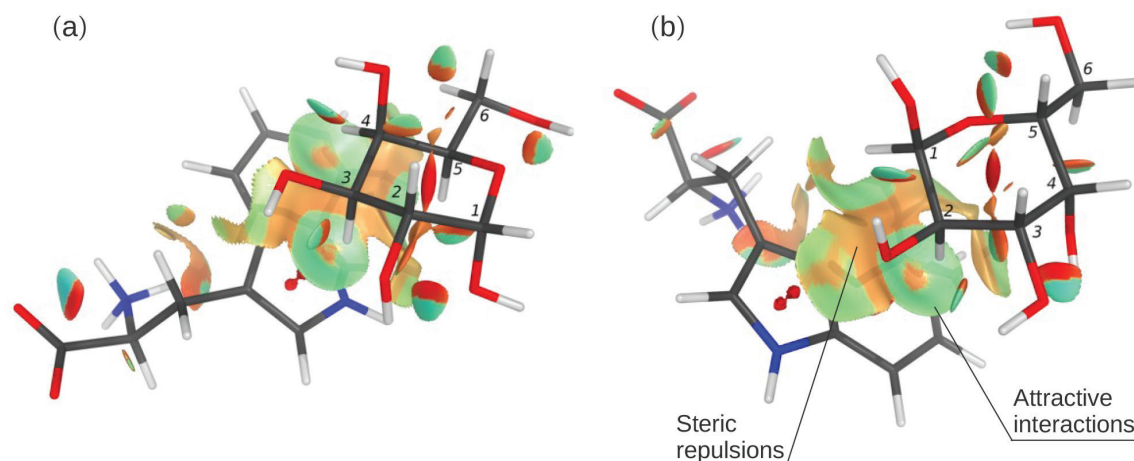

Figure 7. Surface of non-covalent interactions for the complex between α-Gal-gt and Trp. (a) Bottom face and (b) upper face. Isosurface of 0.5 au. Color scale blue-green-red (–0.04 to 0.02 au) according to the sign of ($\lambda_2\rho$).
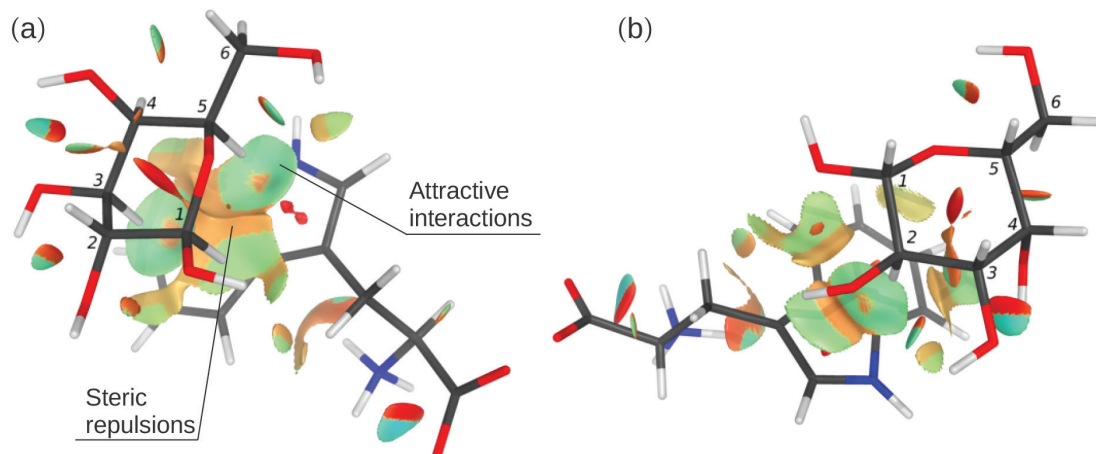

Figure 8. Surface of non-covalent interactions for the complex between β-Gal-gt and Trp. (a) Bottom face and (b) upper face. Isosurface of 0.5 au. Color scale blue-green-red (–0.04 to 0.02 au) according to the sign of ($\lambda_2\rho$).

only the H2 of the carbohydrate exhibiting intermolecular van der Waals interactions with the amino acid. The other attractive intermolecular interactions involved the oxygen atoms of the carbohydrate.

The behavior of the steric intermolecular interaction observed for the complexes between Gal and Trp was similar to that discussed for complexes between Glc and Trp. Analyses of the orbital intermolecular interactions (Table 10) showed that the complexes with Trp at the bottom face had a greater stabilizing energy than the complexes with Trp at the upper face. The most pronounced interactions were those with the amino acid orbitals acting as donor orbitals and the carbohydrate low-occupancy orbitals acting as acceptor orbitals.

The intermolecular orbital steric repulsion (Table 11) was again stronger for the more stable complexes due to the closer proximity between carbohydrate and amino acid. An important factor was the large difference in the steric energy between the position of the Trp on β-Gal. This large difference was the result of the greater distance of the amino acid to the carbohydrate in the complexes with Trp on the upper face of Gal.

The complexes of β-Gal and α-Gal-gt with Trp at the bottom face presented higher intermolecular total energies calculated by SAPT0 analyses (Table 12). For the other α-Gal conformations, the amino acid at the upper face presented a more negative total energy than the same conformation with the amino acid at the bottom face.

As observed for the complexes with Glc, the complexes between Gal and Trp presented an average value for the

Table 10. Sum of the intermolecular orbital energy interactions obtained by NBO for the complexes between Gal and Trp calculated at the M06-2X/6-31++G(d,p) level

entry	Complex	Energy / (kcal mol^{-1})	
		Gal → Trp	Trp → Gal
1	α-Gal-gg_Trp_bottom	0.39	3.93
2	α-Gal-gg_Trp_upper	0.43	2.21
3	α-Gal-gt_Trp_bottom	0.43	4.00
4	α-Gal-gt_Trp_upper	0.20	2.53
5	α-Gal-tg_Trp_bottom	0.36	3.81
6	α-Gal-tg_Trp_upper	0.46	2.23
7	β-Gal-gg_Trp_bottom	0.24	4.20
8	β-Gal-gg_Trp_upper	0.52	3.09
9	β-Gal-gt_Trp_bottom	0.30	4.40
10	β-Gal-gt_Trp_upper	0.39	3.05
11	β-Gal-tg_Trp_bottom	0.23	4.58
12	β-Gal-tg_Trp_upper	0.45	2.75

Gal: D-galactose; Trp: L-tryptophan.

electrostatic component that was smaller (2.3 times) than the average value for the dispersion component. No significant difference in energy was observed in any of the energy components when comparing the two anomers of Gal, as obtained by the SAPT0 analysis. However, when comparing the different carbohydrate faces between each conformation on both anomers, it was observed that the complexes with the amino acid at the bottom face presented the highest energies for all components.

Table 11. Sum of the steric energy of the orbital interactions obtained by NBO for the complexes between Gal and Trp calculated at the M06-2X/6-31++G(d,p) level

entry	Complex	Energy / (kcal mol^{-1})			
		Intramolecular		Intermolecular	Σ[a]
		Glc	Trp	Glc → Trp	
1	α-Gal-gg_Trp_bottom	411.00	494.94	13.70	919.63
2	α-Gal-gg_Trp_upper	411.09	495.31	11.13	917.53
3	α-Gal-gt_Trp_bottom	408.32	494.78	13.89	916.98
4	α-Gal-gt_Trp_upper	408.40	494.12	11.09	913.62
5	α-Gal-tg_Trp_bottom	406.49	494.58	13.42	914.49
6	α-Gal-tg_Trp_upper	406.60	494.96	11.16	912.72
7	β-Gal-gg_Trp_bottom	408.69	494.61	14.12	917.43
8	β-Gal-gg_Trp_upper	410.70	493.18	9.35	913.23
9	β-Gal-gt_Trp_bottom	407.09	494.58	14.10	915.77
10	β-Gal-gt_Trp_upper	407.36	493.99	8.60	909.95
11	β-Gal-tg_Trp_bottom	404.96	495.58	14.30	914.84
12	β-Gal-tg_Trp_upper	405.55	493.38	8.13	907.06

[a]Sum of all interactions. Gal: D-galactose; Trp: L-tryptophan.

Table 12. Energy decomposition analysis for the complexes between Gal and Trp through the SAPT0/6-31++G(d,p) calculation

entry	Complex	Component energy / (kcal mol^{-1})				
		Electrostatic	Exchange	Induction	Dispersion	Total
1	α-Gal-gg_Trp_bottom	−5.40	12.73	−1.54	−12.01	−6.22
2	α-Gal-gg_Trp_upper	−4.69	9.74	−1.25	−10.37	−6.56
3	α-Gal-gt_Trp_bottom	−6.19	12.75	−1.54	−12.09	−7.07
4	α-Gal-gt_Trp_upper	−4.24	9.75	−1.21	−10.19	−5.89
5	α-Gal-tg_Trp_bottom	−4.67	13.02	−1.66	−12.29	−5.60
6	α-Gal-tg_Trp_upper	−4.63	9.69	−1.22	−10.14	−6.30
7	β-Gal-gg_Trp_bottom	−5.63	13.34	−1.81	−12.35	−6.46
8	β-Gal-gg_Trp_upper	−3.29	10.29	−1.88	−10.94	−5.81
9	β-Gal-gt_Trp_bottom	−6.91	13.25	−1.88	−12.54	−8.09
10	β-Gal-gt_Trp_upper	−2.92	9.28	−1.55	−9.87	−5.06
11	β-Gal-tg_Trp_bottom	−5.29	14.06	−1.97	−12.81	−6.01
12	β-Gal-tg_Trp_upper	−3.07	8.81	−1.53	−9.75	−5.54

Gal: D-galactose; Trp: L-tryptophan.

For both α-Gal and β-Gal, the smallest C–H bond dipole moment observed was that of the C2–H bond, which is the only C–H bond on the upper face of the carbohydrate (Table S23, SI section). The atomic charges of the H1–H5 in the complexes between Gal and Trp (Table S24, SI section) did not present significant differences between each other, but H2 always presented the highest charge. In β-Gal, a considerable decrease in the H1 charge was also observed. Considering the isolated carbohydrate, the H2 of α-Gal was the one with the highest atomic charge, with a very pronounced difference for the other hydrogen atoms. In β-Gal, the difference in atomic charge between H1–H5 was smaller, but with H2 still presenting the highest atomic charge. This indicates that the dipole moment of the C–H bond is more important in explaining the difference between the faces of the carbohydrate than the atomic charge of the hydrogen atoms.

It was observed that the amino acid at the bottom face was more stable in both anomers for the complexes between Gal and Trp. This higher stability is related to the intermolecular orbital interactions involving the π orbitals of the amino acid with H3–H5 of the α anomer of the carbohydrate and involving H1, H3, and H5 of the β anomer. These hydrogen atoms were positioned in the shielding cone of the indole group and were in agreement with the ^{1}H chemical shift perturbation observed in NMR analyses.

When compared to the Glc-Trp complexes, it is important to note that α-Gal showed a higher stability with the amino acid at the bottom face due to the axial position of O4–H in Gal. This axial position allows the indole group

to interact with three hydrogen atoms of the carbohydrate (H3–H5), while this interaction is only with H2 in the upper face. This led to complexes with the amino acid directed mostly to one side of the carbohydrate in α-Gal-Trp. For α-Glc, the bottom face presented only two hydrogen atoms to interact with the indole group and the anomeric hydroxyl group was in an axial position, increasing the steric interaction of the carbohydrate on this face. A similar steric effect occurred due to O4–H hydroxyl group located in the upper face of Gal.

D-Galactose-L-phenylalanine

Data for the most stable complex for each conformation of both anomers of Gal with the aromatic ring of Phe centered at H1–H6 is presented in Table 13 and the structures can be visualized in the SI section (Figures S23-S28).

The complexes with the Gal *trans-gauche* conformation were the least stable complexes with Phe, like the complexes of Glc with Phe. The complexes with the amino acid at the bottom face of the carbohydrate were the most stable for both anomers, highlighting the ones centered at H3 and H5, in accordance with the experimental ^{1}H chemical shift perturbations. These complexes showed higher energies involved in the intermolecular orbital interactions than the complexes with the amino acid at the upper face of the carbohydrate. The analyses of the NCI surfaces were similar to those described for the Glc-Phe complexes.

For β-Gal, the complexes with the aromatic ring of Phe oriented toward H1 presented higher intermolecular orbital interactions (5.04, 4.48, and 4.62 kcal mol^{-1} for

Table 13. Relative energy for the complexes between Gal and Phe (with ZPE and BSSE corrections) at the M06-2X/6-31++G(d,p) level in water (IEF-PCM/Bondi) and sum the of the intermolecular orbital interactions obtained by NBO calculated at the same level in vacuum

entry	Complex	ΔE / (kcal mol^{-1})	NBO / (kcal mol^{-1})	
			Gal → Phe	Phe → Gal
1	α-Gal-gg_Phe_H1	1.09	0.50	2.52
2	α-Gal-gg_Phe_H2	0.67	0.43	2.53
3	α-Gal-gg_Phe_H3	0.12	0.68	2.73
4	α-Gal-gg_Phe_H4	1.13	0.46	2.51
5	α-Gal-gg_Phe_H5	0.06	0.42	2.85
6	α-Gal-gg_Phe_H6	0.89	1.71	1.61
7	α-Gal-gt_Phe_H1	1.00	0.50	2.59
8	α-Gal-gt_Phe_H2	0.67	0.32	2.46
9	α-Gal-gt_Phe_H3	0.00	0.74	2.51
10	α-Gal-gt_Phe_H4	0.36	0.47	3.14
11	α-Gal-gt_Phe_H5	0.13	0.61	2.84
12	α-Gal-gt_Phe_H6	0.84	0.40	2.20
13	α-Gal-tg_Phe_H1	2.50	0.26	2.67
14	α-Gal-tg_Phe_H2	1.83	0.81	2.12
15	α-Gal-tg_Phe_H3	1.75	0.35	5.59
16	α-Gal-tg_Phe_H4	1.73	0.57	3.11
17	α-Gal-tg_Phe_H5	0.71	1.03	3.95
18	α-Gal-tg_Phe_H6	3.18	0.16	1.19
19	β-Gal-gg_Phe_H1	1.11	0.40	4.08
20	β-Gal-gg_Phe_H2	1.22	0.53	3.43
21	β-Gal-gg_Phe_H3	0.99	0.47	3.27
22	β-Gal-gg_Phe_H4	1.00	0.47	2.65
23	β-Gal-gg_Phe_H5	0.68	0.39	3.58
24	β-Gal-gg_Phe_H6	2.16	0.37	2.53
25	β-Gal-gt_Phe_H1	0.53	0.25	4.79
26	β-Gal-gt_Phe_H2	0.89	0.47	2.99
27	β-Gal-gt_Phe_H3	0.64	0.46	3.19
28	β-Gal-gt_Phe_H4	0.55	0.51	3.26
29	β-Gal-gt_Phe_H5	0.67	0.50	3.76
30	β-Gal-gt_Phe_H6	1.83	0.32	1.99
31	β-Gal-tg_Phe_H1	2.09	0.38	4.24
32	β-Gal-tg_Phe_H2	1.85	0.41	2.71
33	β-Gal-tg_Phe_H3	1.60	0.41	2.98
34	β-Gal-tg_Phe_H4	1.91	1.19	2.10
35	β-Gal-tg_Phe_H5	1.54	0.53	2.97
36	β-Gal-tg_Phe_H6	2.34	0.80	1.92

NBO: natural bond orbital; Gal: D-galactose; Phe: L-phenylalanine.

gauche-trans, *gauche-gauche* and *trans-gauche* conformers, respectively). In this anomer, H1 was in an axial position on the bottom face where it had a good interaction with the amino acid. The same H1 in the α-Gal was in an equatorial position on the upper face of the carbohydrate and showed considerably smaller intermolecular orbital interactions (3.09, 3.02, and 2.93 kcal mol^{-1} for *gauche-trans*, *gauche-gauche* and *trans-gauche* conformers, respectively) than the complexes of β-Gal.

SAPT0 analysis (Table S27; Figures S30-S31, SI section) for the complexes between Gal and Phe did not show a clear trend towards favoring one of the faces of the carbohydrate. The energy values of the components were quite close for each face. Similarly to Glc-Phe complexes, the Gal-Phe complexes presented smaller energy values for the dispersion component than the Gal-Trp complexes.

D-Galactose-L-histidine

Table 14 presents the most stable complex for each conformation of both anomers of Gal with the His ring centered at H1–H6. These complex structures can be visualized in the SI section (Figures S31-S36).

The behavior of the NCI surfaces of the Gal-His complexes were similar to that discussed for the Glc-His complexes.

The complexes with the His ring centered at the upper face of α-Gal-gt (H1 and H2) were the most stable. These complexes presented intermolecular hydrogen bonds involving the lone pairs of the carboxylic oxygen of the amino acid with the hydroxyl group bonded to C6 ($\eta_O \rightarrow \sigma^*_{O6-H}$). Even without considering the intermolecular hydrogen bond, the complexes centered at H2 presented higher energies of intermolecular orbital interactions. However, these complexes also presented higher steric energies for each conformation of α-Gal (Table S29, SI section).

The complexes with His centered at the bottom face of the carbohydrate (H1, H3, and H5) were the most stable for β-Gal. Again, some complexes presented very pronounced intermolecular orbital interaction energies due to hydrogen bonding, such as $\eta_O \rightarrow \sigma^*_{O8-H}$ (36.21 kcal mol^{-1}) for β-Gal-tg_His_H3 and $\eta_N \rightarrow \sigma^*_{O11-H}$ (13.59 kcal mol^{-1}) for β-Gal-gg_His_H2. The stability order for β-Gal was closer to the observed ^{1}H chemical shift perturbation than for α-Gal.

The complexes between Gal and His presented a similar situation to that observed for Glc-His, where the presence of intermolecular hydrogen bonds lead to higher energy values for the electrostatic component than for the dispersion component. For both Glc and Gal, the highest energy values for the induction component among all complexes were observed in the complexes with His (Table S30, SI section).

Table 14. Relative energy for the complexes between Gal and His (with ZPE and BSSE corrections) at the M06-2X/6-31++G(d,p) level in water (IEF-PCM/Bondi) and the sum of the intermolecular orbital interactions obtained by NBO calculated at the same level in vacuum

entry	Complex	ΔE / (kcal mol^{-1})	NBO / (kcal mol^{-1})	
			Gal $\rightarrow$ His	His $\rightarrow$ Gal
1	α-Gal-gg_His_H1	3.77	1.66	1.93
2	α-Gal-gg_His_H2	3.10	1.85	2.90
3	α-Gal-gg_His_H3	3.47	0.58	3.19
4	α-Gal-gg_His_H4	3.84	0.14	2.70
5	α-Gal-gg_His_H5	3.20	0.52	3.36
6	α-Gal-gg_His_H6	3.71	0.38	2.11
7	α-Gal-gt_His_H1	0.37	2.73	32.31
8	α-Gal-gt_His_H2	0.00	2.93	21.58
9	α-Gal-gt_His_H3	3.23	0.21	2.77
10	α-Gal-gt_His_H4	3.91	1.41	2.16
11	α-Gal-gt_His_H5	3.16	0.49	3.08
12	α-Gal-gt_His_H6	0.38	2.63	30.95
13	α-Gal-tg_His_H1	4.67	1.60	1.86
14	α-Gal-tg_His_H2	4.30	1.82	2.77
15	α-Gal-tg_His_H3	4.71	0.73	3.01
16	α-Gal-tg_His_H4	4.76	1.61	1.65
17	α-Gal-tg_His_H5	3.73	0.56	2.83
18	α-Gal-tg_His_H6	5.78	0.66	1.43
19	β-Gal-gg_His_H1	4.18	0.80	7.01
20	β-Gal-gg_His_H2	3.80	0.84	20.18
21	β-Gal-gg_His_H3	4.06	0.45	2.78
22	β-Gal-gg_His_H4	3.98	0.14	2.75
23	β-Gal-gg_His_H5	4.23	0.08	2.86
24	β-Gal-gg_His_H6	4.72	0.18	2.23
25	β-Gal-gt_His_H1	3.40	0.90	3.72
26	β-Gal-gt_His_H2	3.95	1.78	3.00
27	β-Gal-gt_His_H3	3.80	0.88	2.87
28	β-Gal-gt_His_H4	4.01	0.08	2.68
29	β-Gal-gt_His_H5	3.41	2.53	2.85
30	β-Gal-gt_His_H6	4.87	0.33	1.50
31	β-Gal-tg_His_H1	4.31	0.66	7.36
32	β-Gal-tg_His_H2	4.05	1.41	2.92
33	β-Gal-tg_His_H3	0.63	2.14	40.38
34	β-Gal-tg_His_H4	4.58	13.10	1.10
35	β-Gal-tg_His_H5	4.20	1.22	3.50
36	β-Gal-tg_His_H6	5.72	0.15	1.47

NBO: natural bond orbital; Gal: D-galactose; His: L-histidine.

Conclusions

Analysis of the NMR data and molecular modelling provided a detailed explanation of the interaction of the monosaccharides Glc and Gal with amino acids containing aromatic groups (Trp, Phe and His).

Trp has the larger group and induced the higher chemical shift perturbation on the hydrogen atoms bonded to the carbon atoms of the carbohydrates. Trp also showed a clear preference for one of the faces of the carbohydrates studied. This preference can be understood through the analyses of the intermolecular interactions present in these complexes, mostly the ones involving the π orbitals of the amino acid interacting with the σ^*_{C-H} orbitals of the carbohydrates. The decomposition of the intermolecular energy showed that the dispersion component contributes more to the total energy of the intermolecular interaction than the electrostatic component, except when hydrogen bonds were present. However, the preference for one of the faces of the monosaccharides was related to the difference in the electrostatic component between the complexes at different faces of the carbohydrate. These analyses were performed employing NCI and NBO methodologies with the DFT method (M06-2X) and SAPT0 for EDA. The most stable complex of α-Glc had the amino acid at the upper face of the carbohydrate, while the most stable complex for β-Glc was the one with the amino acid at the bottom face. For α-Gal, the most stable complex showed the aromatic group of the amino acid oriented towards hydrogen atoms H3–H5, while the aromatic group was oriented at the bottom face (H1, H3, and H5) of the carbohydrate for β-Gal.

Phe presented intermediary chemical shift perturbations and a preference for one of the faces of the carbohydrate, while His presented the smallest chemical shift perturbation and did not show a clear preference for one of the carbohydrate faces. The facial selectivity was influenced by the electrostatic component mainly through C–H bond dipoles and decreased as the amino acid aromatic ring became smaller since the number of C–H bonds interacting with it decreases from Trp to Phe and to His.

Supplementary Information

Supplementary information (chemical shift perturbation data, description of the carbohydrate conformational search, more detailed complex energy tables and supplementary figures, Cartesian coordinates for all structures presented, animated versions of the NCI surface of some complexes, effect of the intermolecular distance between carbohydrate and amino acid on the EDA, details of input generation and videos showing the approximation of Trp to the upper and bottom face of Glc anomers) is available free of charge at http://jbcs.sbq.org.br.

Acknowledgments

References

1. Lis, H.; Sharon, N.; *Lectins*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003.

2. Lis, H.; Sharon, N.; *Chem. Rev.* **1998**, *98*, 637.

3. Mazik, M.; *Chem. Soc. Rev.* **2009**, *38*, 935.

4. Quiocho, F. A.; *Pure Appl. Chem.* **1989**, *61*, 1293.

5. Vyas, N.; Vyas, M.; Quiocho, F.; *Science* **1988**, *242*, 1290.

6. Toone, E. J.; *Curr. Opin. Struct. Biol.* **1994**, *4*, 719.

7. Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J.; *Acc. Chem. Res.* **2013**, *46*, 946.

8. Chen, W.; Enck, S.; Price, J. L.; Powers, D. L.; Powers, E. T.; Wong, C. H.; Dyson, H. J.; Kelly, J. W.; *J. Am. Chem. Soc.* **2013**, *135*, 9877.

9. Tsuzuki, S.; Uchimaru, T.; Mikami, M.; *J. Phys. Chem. A* **2011**, *115*, 11256.

10. Lucas, R.; Peñalver, P.; Gómez-Pinto, I.; Vengut-Climent, E.; Mtashobya, L.; Cousin, J.; Maldonado, O. S.; Perez, V.; Reynes, V.; Aviñó, A.; Eritja, R.; González, C.; Linclau, B.; Morales, J. C.; *J. Org. Chem.* **2014**, *79*, 2419.

11. Fernández-Alonso, M. C.; Cañada, F. J.; Jiménez-Barbero, J.; Cuevas, G.; *J. Am. Chem. Soc.* **2005**, *127*, 7379.

12. Ramírez-Gualito, K.; Alonso-Ríos, R.; Quiroz-Garcia, B.; Rojas-Aguilar, A.; Díaz, D.; Jiménez-Barbero, J.; Cuevas, G.; *J. Am. Chem. Soc.* **2009**, *131*, 18129.

13. Kozmon, S.; Matuška, R.; Spiwok, V.; Koča, J.; *Chem. - Eur. J.* **2011**, *17*, 5680.

14. Stanca-Kaposta, E. C.; Gamblin, D. P.; Screen, J.; Liu, B.; Snoek, L. C.; Davis, B. G.; Simons, J. P.; *Phys. Chem. Chem. Phys.* **2007**, *9*, 4444.

15. Kozmon, S.; Matuška, R.; Spiwok, V.; Koča, J.; *Phys. Chem. Chem. Phys.* **2011**, *13*, 14215.

16. Stanca-Kaposta, E. C.; Carçabal, P.; Cocinero, E. J.; Hurtado, P.; Simons, J. P.; *J. Phys. Chem. B* **2013**, *117*, 8135.

17. Kumari, M.; Balaji, P. V.; Sunoj, R. B.; *Phys. Chem. Chem. Phys.* **2011**, *13*, 6517.

18. Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher,

T.; Kiessling, L. L.; Woolfson, D. N.; *J. Am. Chem. Soc.* **2015**, *137*, 15152.

19. Barwell, N. P.; Davis, A. P.; *J. Org. Chem.* **2011**, *76*, 6548.

20. Jiménez-Moreno, E.; Jiménez-Osés, G.; Gómez, A. M.; Santana, A. G.; Corzana, F.; Bastida, A.; Jiménez-Barbero, J.; Asensio, J. L.; *Chem. Sci.* **2015**, *6*, 6076.

21. Jiménez-Moreno, E.; Gómez, A. M.; Bastida, A.; Corzana, F.; Jiménez-Oses, G.; Jiménez-Barbero, J.; Asensio, J. L.; *Angew. Chem., Int. Ed.* **2015**, *54*, 4344.

22. Sharma, R.; McNamara, J. P.; Raju, R. K.; Vincent, M. A.; Hillier, I. H.; Morgado, C. A.; *Phys. Chem. Chem. Phys.* **2008**, *10*, 2767.

23. Vandenbussche, S.; Díaz, D.; Fernández-Alonso, M. C.; Pan, W.; Vincent, S. P.; Cuevas, G.; Cañada, F. J.; Jimínez-Barbero, J.; Bartik, K.; *Chem. - Eur. J.* **2008**, *14*, 7570.

24. Xu, G.; Evans, J. S.; *J. Magn. Reson., Ser. B* **1996**, *111*, 183.

25. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; *Gaussian 09, Revision B.01*; Gaussian Inc., Wallingford, CT, 2009.

26. Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Weinhold, F.; *NBO 5.9*; Theoretical Chemistry Institute, University of Wisconsin, Madison, WI, 2009. Available at http://www.chem.wisc.edu/~nbo5, accessed in December 2018.

27. Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W.; *J. Am. Chem. Soc.* **2010**, *132*, 6498.

28. Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince III, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Remigio, R. D.; Richard, R. M.; Gonthier, J. F.; James, A. M.; McAlexander, H. R.; Kumar, A.; Saitow, M.; Wang, X.; Pritchard, B. P.; Verma, P.; Schaefer III, H. F.; Patkowski, K.; King, R. A.; Valeev, E. F.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D.; *J. Chem. Theory Comput.* **2017**, *13*, 3185.

29. Delano, W. L.; *PyMOL(TM) Molecular Graphics System*, version 1.7.0.0; Schrodinger, LLC, USA, 2015.

30. Zhao, Y.; Truhlar, D. G.; *Theor. Chem. Acc.* **2008**, *120*, 215.

31. Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A.; *J. Chem. Phys.* **1980**, *72*, 650.

32. Cancès, M. T.; Mennucci, B.; Tomasi, J.; *J. Chem. Phys.* **1997**, *107*, 3032.

33. Zhao, Y.; Truhlar, D. G.; *Acc. Chem. Res.* **2008**, *41*, 157.

34. Sameera, W. M. C.; Pantazis, D. A.; *J. Chem. Theory Comput.* **2012**, *8*, 2630.

35. Jensen, J. H.; *Molecular Modeling Basics*; CRC Press: Boca Raton, 2010.

36. Simon, S.; Duran, M.; Dannenberg, J. J.; *J. Chem. Phys.* **1996**, *105*, 11024.

37. Jeziorski, B.; Moszynski, R.; Szalewicz, K.; *Chem. Rev.* **1994**, *94*, 1887.

38. Hohenstein, E. G.; Sherrill, C. D.; *J. Chem. Phys.* **2010**, *132*, 184111.

39. Hohenstein, E. G.; Parrish, R. M.; Sherrill, C. D.; Turney, J. M.; Schaefer, H. F.; *J. Chem. Phys.* **2011**, *135*, 174107.

40. Stott, K.; Stonehouse, J.; Keeler, J.; Hwang, T. L.; Shaka, A. J.; *J. Am. Chem. Soc.* **1995**, *117*, 4199.