Volume 1 | Number 1 | Jan 2013 | Pages 1–100

Chemical Science

www.rsc.org/chemicalscience

ROYAL SOCIETY OF CHEMISTRY

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemicalscience

A Thorough Experimental Study of CH/π Interactions in Water: Quantitative Structure-stability Relationships for Carbohydrate/Aromatic Complexes

Ester Jiménez-Moreno, [‡a] Gonzalo Jiménez-Osés, [‡b,c] Ana M. Gómez,[a] Andrés G. Santana,[a] Francisco Corzana,[b] Agatha Bastida,[a] Jesus Jiménez-Barbero,[d,e,f] Juan Luis Asensio[a] *

[a]Instituto de Química Orgánica (IQOG-CSIC), Juan de la Cierva 3, 28006 Madrid, Spain. Fax: +34-91-5644853; Tel: +34-91-5622900.

[b]Dept. Química and Centro de Investigación en Síntesis Química, Universidad de La Rioja, Logroño, Spain.

[c]Institute of Biocomputation and Physics of Complex Systems (BIFI) University of Zaragoza, BIFI-IQFR (CSIC), Zaragoza, Spain.

[d]Centro de Investigaciones Biológicas (CIB-CSIC), Madrid, Spain.

[e]Center for Cooperative Research in Biosciences (CIC-bioGUNE), Derio-Bizkaia, Spain.

[f]Basque Foundation for Science, Ikerbasque, Bilbao, Spain.

[‡] EJM and GJO have contributed equally to this work

* e-mail: juanluis.asensio@csic.es, phone: +34 915622900, Fax: +34 915644853

Abstract

CH/π interactions play a key role in a large variety of molecular recognition processes of biological relevance. However, their origin and structural determinants in water remain poorly understood. In order to improve our comprehension of these important interaction modes, we have performed a quantitative experimental analysis of a large data set comprising 117 chemically diverse carbohydrate/aromatic stacking complexes, prepared through a dynamic combinatorial approach recently developed by our group. The obtained free energies provide a detailed picture of the structure/stability relationships that govern the association process, opening the door to the rational design of improved carbohydrate-based ligands or carbohydrate receptors. Moreover, this experimental data set, supported by quantum mechanical calculations, have contributed to understand the main driving forces that promote complex formation, underlining the key role played by coulombic and solvophobic forces on the stabilization of these contacts. This represents the most quantitative and extensive experimental study reported so far for CH/π complexes in water.

Introduction

Biomolecular recognition processes are mediated by a wide range of weak intermolecular forces, that cooperatively act to provide binding affinity and specificity. Among them, CH/π bonds have attracted a considerable attention in recent years due to their importance and widespread occurrence.[1-12] These interactions can be generally described as the attractive molecular force occurring between polarized CH fragments and aromatic rings. According to theoretical studies, in the gas phase, a major dispersive component, together with a smaller, but still significant, electrostatic contribution (which increases with the polarization of the involved CH groups) determine the stability of the resulting complexes.[2,3] However, the structural determinants and driving forces behind CH/π bonds in water have proved more difficult to dissect and remain poorly understood.

High-dielectric solvents are expected to attenuate coulombic interactions between the CH moieties and the aromatic systems. Intriguingly, we have shown in a recent study that, even in water, CH/π contacts are strongly stabilized by CH polarization, suggesting that electrostatic forces remain remarkably relevant.[5] Since this contribution increases in low dielectric environments, CH/π complexes might be expected to be stronger in organic solvents. However, it is well established that the addition of organic co-solvents usually has a strongly unfavourable influence on the interaction.[2,4d,6a,7a] This phenomenon has traditionally been considered an evidence for the relevance of the hydrophobic effect in the stabilization of the complex. However, it might also be explained by competitive CH/π contacts between the substrates and co-solvent molecules, as pointed out by Nishio and co-workers.[2]

Regarding the van der Waals contribution, while certainly dominant in the gas phase, its significance and relative weight in solution remains unclear. Indeed, recent experimental measurements of alkyl–alkyl interactions in solution seem to be consistent with a large cancellation of dispersion forces, due to competitive interactions with the solvent.[13] This attenuation is expected

to be highly dependent on the molecular properties of the involved species and its magnitude for CH/π bonds is presently unknown.

Finally, from a design perspective, a quantitative description of the most relevant features of these complexes and their influence on the binding free energies would be greatly acknowledged. Such an analysis should explicitly consider a wide range of variations in both the CH/π donor and acceptor molecules (see below).

In summary, after more than twenty years of scientific research on this topic, a detailed and quantitative picture of the origin and structure/stability relationships that govern these important interaction modes in water is still missing. Such description might have a significant impact in different fields, such as chemical biology, supramolecular chemistry, and medicinal chemistry, facilitating the design of improved ligands, receptors, foldamers, or new materials.

In order to improve our understanding of CH/π complexes in water, we have carefully studied a paradigmatic example of biological relevance: carbohydrate/aromatic stacking (Figure 1).[6-12] This particular interaction mode plays a key role in the molecular recognition of glycans by antibodies, enzymes, lectins, and nucleic acids and typically involves the simultaneous formation of 2-3 intermolecular bonds between pyranose CH groups and receptor aromatic units. It should be noted that carbohydrate/aromatic contacts can involve the α and/or β faces of the pyranose ring (herein referred as α- and β-type complexes, respectively) and that different axial/equatorial orientations are also feasible for the polar OH moieties. Moreover, the aromatic systems can also exhibit a wide range of chemical properties (electron density, charge distribution, solubility), conferring a significant structural diversity to the stacking complexes (Figure 1).

We have employed a dynamic combinatorial protocol recently developed by our group[5,8a] to analyze a large data set of 117 chemically diverse carbohydrate/aromatic complexes. This represents the most quantitative and extensive experimental study reported so far for CH/π complexes in water. The obtained experimental data, in combination with quantum mechanics calculations, provide a detailed picture of the structure/stability relationships that govern these complexes, and contribute to clarify their origin.
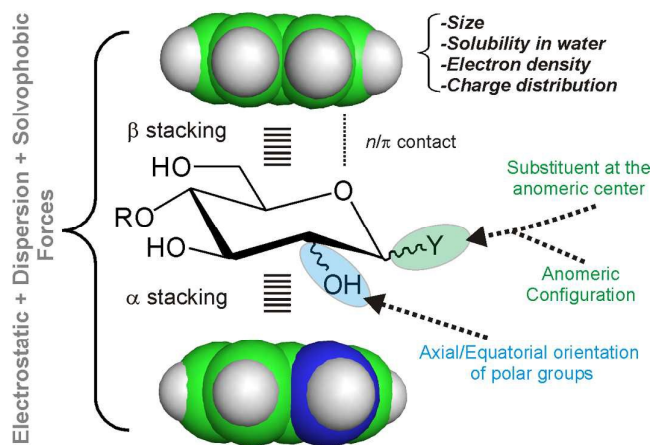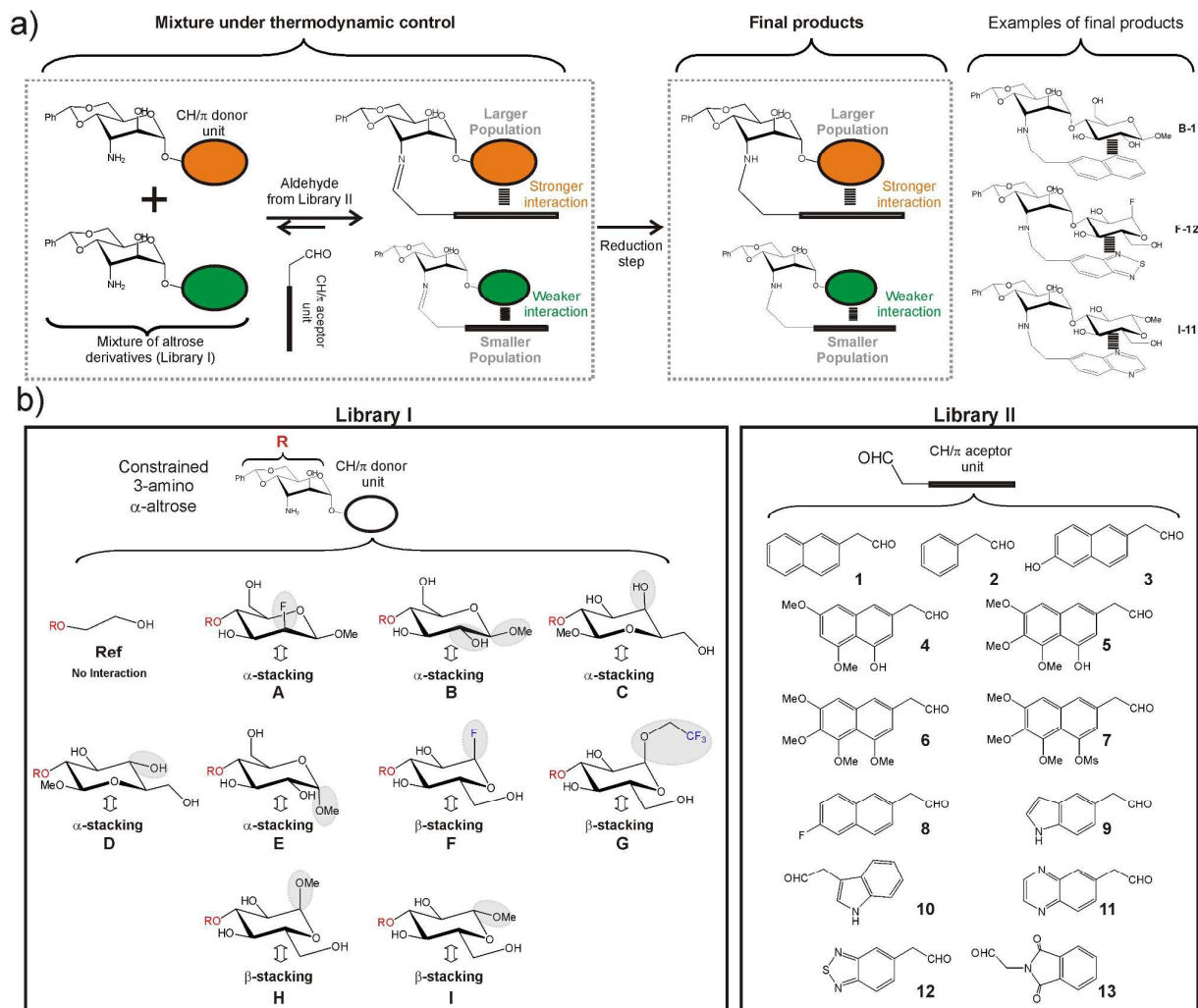
FIGURE 1.-Schematic representation of the main structural features that determine the stacking free energy in carbohydrate/aromatic CH/π complexes.

Results and Discussion

A.- Brief description of the dynamic combinatorial approach and designed libraries.- The employed methodology has been previously described and is briefly outlined in Figure 2a.[8a] First, libraries I and II were designed and synthesized (Figure 2b and S1-S2). Library I included nine disaccharide derivatives (**A-I**) formed by a common 3-deoxy-3-amino α-altrose scaffold linked to alternative CH/π donor units. Similarly, library II included thirteen acetaldehyde derivatives (**1-13**) equipped with different CH/π acceptor aromatic systems. It should be noted that for any given disaccharide/aldehyde pair, one particular imine/hemiaminal adduct could be formed. Interestingly, according to previous studies, these species are stabilized by well-defined intra-molecular CH/π contacts, involving the aromatic system and only one (α or β) face of the donor fragment (Figure 2a, 2b and S3). Thus, disaccharides with α(1-2) (**C**, **D**) or α(1-4) (**A**, **B** and **E**) bonds form α-type complexes upon reaction with the aromatic aldehyde, while those with α(1-3) linkages (**F**, **G**, **H** and **I**) establish β-type complexes. Our library also included CH/π donor units with different axial/equatorial orientations for the key OH/F positions (Figure 2b). All these modifications were selected to evaluate specific structural features of the interaction modes and their influence on stability. These include:

a) The relative orientation of the pyranose and aromatic units (compound **B** vs **D**).

b) The equatorial or axial orientation of those polar groups located on non-interacting CH positions

(compounds **A** vs **B** and **C** vs **D**).

c) The presence of axial polar groups on the interacting pyranose face (compounds **B** vs **E**).

d) The participation of the α or β pyranose face in the stacking interaction (**C** vs **H** and **D** vs **I**).

FIGURE 2.- a) Schematic representation of the employed dynamic combinatorial strategy (see the main text). Examples of the products obtained upon reductive amination of disaccharides **A-I** with aldehydes **1-13** are shown on the right. b) Libraries of disaccharide (Library I, compounds **A-I**) and aldehyde (Library II, compounds **1-13**) derivatives synthesized for our analyses. For Library I diverse functions and/or stereochemistries were considered for both anomeric and non-anomeric positions. The most relevant variations at these key sites are highlighted with grey ellipses (see the main text). The interacting face (α or β) of the different CH/π donor units is indicated with a double-headed arrow.

e) The configuration of the anomeric centre in β-type complexes (**H** vs **I**).

f) The presence of electron-withdrawing substituents at the anomeric centre in β-type complexes (**H**

vs **F** or **G**).

Finally, a reference compound (**Ref**) lacking a CH/π donor pyranoside was also prepared.

Regarding library II, aromatic systems with different sizes, orientations, electronic properties and water solubility were explicitly considered (Figure 2b).

Equimolecular mixtures of two or more model disaccharides in buffered water solutions were treated with sub-stoichiometric amounts of a given aldehyde to produce a dynamic mixture of the intermediate imines/hemiaminals (Figure 2a). After equilibration, chemical reduction of the transient species with sodium cyanoborohydride yields a non-equimolecular mixture of secondary amines whose relative populations were evaluated by either 2D or 1D NMR spectroscopy. These populations have been shown to reflect those of the intermediate species[5] and, therefore, were employed to determine the stability differences ($\Delta\Delta G$) among alternative stacking modes. Moreover, by including the reference altrose derivative **Ref** (for which no stacking is feasible) in the initial mixture, we could also determine net interaction energies for the different complexes (ΔG).

This experimental approach outperforms more conventional methodologies, commonly based on binding studies with synthetic receptors[4a,6] or wild-type/mutated proteins.[7a,e,10] First, it allows the analysis and quantification of CH/π bonds in the absence of additional interfering interactions (such as hydrogen bonds, van der Waals contacts, etc). Second, it displays an unprecedented sensitivity. Our experience showed that down to 10% population differences among final products can be detected by NMR.[8a] At 277 K (the temperature selected for our experiments), this permits the quantification of $\Delta\Delta G$ values as small as 0.05 kcal/mol. This point is of capital importance when dealing with very weak forces, such as CH/π interactions. Finally, most experimental analysis of carbohydrate/aromatic complexes reported so far has been restricted to a few simple carbohydrate/aromatic pairs, with special attention to those systems commonly found in nature. In contrast, our method is highly versatile and allows incorporating a large variety of CH/π donor and acceptor fragments.

B.- Experimental evaluation of stacking free energies.- In order to estimate the CH/π donor properties of the different pyranose units, we first carried out combinatorial experiments with relatively complex mixtures, formed by up to 5-8 selected disaccharides and one single aldehyde.

Relative populations of the final products were determined by integrating key representative signals in 2D-HSQC experiments.

One example is represented in Figure 3. In this particular case, an equimolecular mixture of disaccharides **B**, **C**, **D**, **E**, **H** and **I**, together with the reference compound (**Ref**) were treated with aldehyde **9** (see the experimental section). In the presence of an excess of aldehyde, the reductive amination reaction yields an equimolecular mixture of the final disaccharide products that can be easily identified by 2D-HSQC experiments (Figure 3 upper panel: the altrose anomeric region).
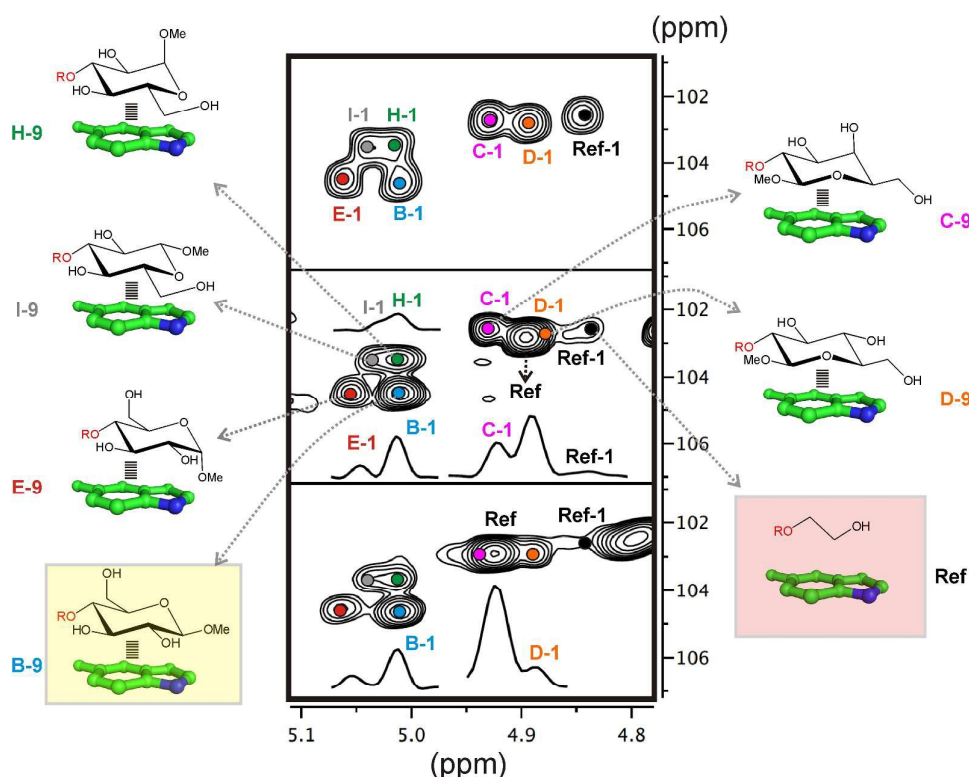
FIGURE 3.- Dynamic combinatorial experiments followed by 2D-HSQC NMR (see the main text). Middle and lower panel show NMR data acquired at slightly different pH values to relieve the signal overlapping.

On the contrary, using a sub-stoichiometric amount of aldehyde **9**, the effective competition among the alternative CH/π donor units is established. Under these circumstances (middle and lower panels), the experiment yields unequal populations of the different disaccharide products, thus providing a qualitative indication of the stability differences among the carbohydrate/aromatic complexes. Thus, the data showed that glucose unit **B** forms the most stable complex with the indole ring (complex **B-9**), while the weakest interaction corresponds to **I**. As expected, the reaction

product for the reference compound (**Ref-9**), with no carbohydrate/aromatic stacking, is barely detectable in solution.

Although the HSQC data provided valuable qualitative information, a more precise integration of the NMR signals required 1D 1H-NMR data collected with long acquisition delays. Therefore, simpler experiments, involving a single pair of altrose derivatives and one aldehyde at a time were performed. Specifically, we carried out competitions within all possible **Ref**/disaccharide pairs and the different aldehydes to obtain net interaction free energies (ΔG) for the alternative stacking modes. Moreover, we also determined the relative stabilities of all the complexes with respect to those formed by **D** ($\Delta\Delta G$), performing reactions with all possible **D**/disaccharide/aldehyde combinations. Finally, many other crosschecks were carried out with selected aldehydes and disaccharide pairs (Figure S4).

The preferred geometry of the CH/π complexes was deduced from the pattern of chemical shift perturbations promoted at the interacting pyranoses (Figures S3 and S5-S7). The interaction geometries were also estimated by molecular dynamic simulations in explicit solvent (MD, Figure S8).

The obtained results, expressed as net stacking energies (ΔG_{exp}(kcal/mol)), are schematically represented in Figure 4 and collected in Table S1. According to these data, the stability of carbohydrate/aromatic complexes in water is in the 0-1.5 kcal/mol range. These values are in close agreement with those reported in the literature for natural systems,[7a] and imply that every single CH/bond contributes by up to 0.5-0.7 kcal/mol to the interaction free energy.

C.- Structure/stability relationships: Influence of the pyranose and aromatic chemical nature.- The results indicate that both, the pyranose and the aromatic unit, have a significant influence on the stability of the stacking complexes. Several trends are apparent (see Figures 4-6):

a) The larger the number of CH/π contacts, the stronger the complex. This simple rule explains why the aromatic system **2** establishes weaker interactions than the more extended naphthyl unit present in **1**. It also explains why derivative **B** forms more stable complexes than **C** or **D** (in the former case, the glucose hydroxymethyl group also contacts the aromatic ring. See Figure S6).

9

b) An axial-oriented OMe group on the interacting face of the pyranose can be highly disruptive to complex formation. This effect is exemplified by the different stability exhibited by complexes **B-3** and **E-3** (0.64 kcal/mol larger for the former. See Table S1 and Figure 5a). The energy penalty detected in **E-3** results from the steric conflict between the axial polar group and the aromatic system and, consequently, depends on the size and orientation of this ring. In agreement with this view, for a differently oriented indole ring the complex destabilization decreases down to 0.38 kcal/mol (**B-10** vs **E-10**. Figure 5a) while the smaller phenyl ring becomes virtually insensitive to the pyranose configuration (**B-2** vs **E-2**).
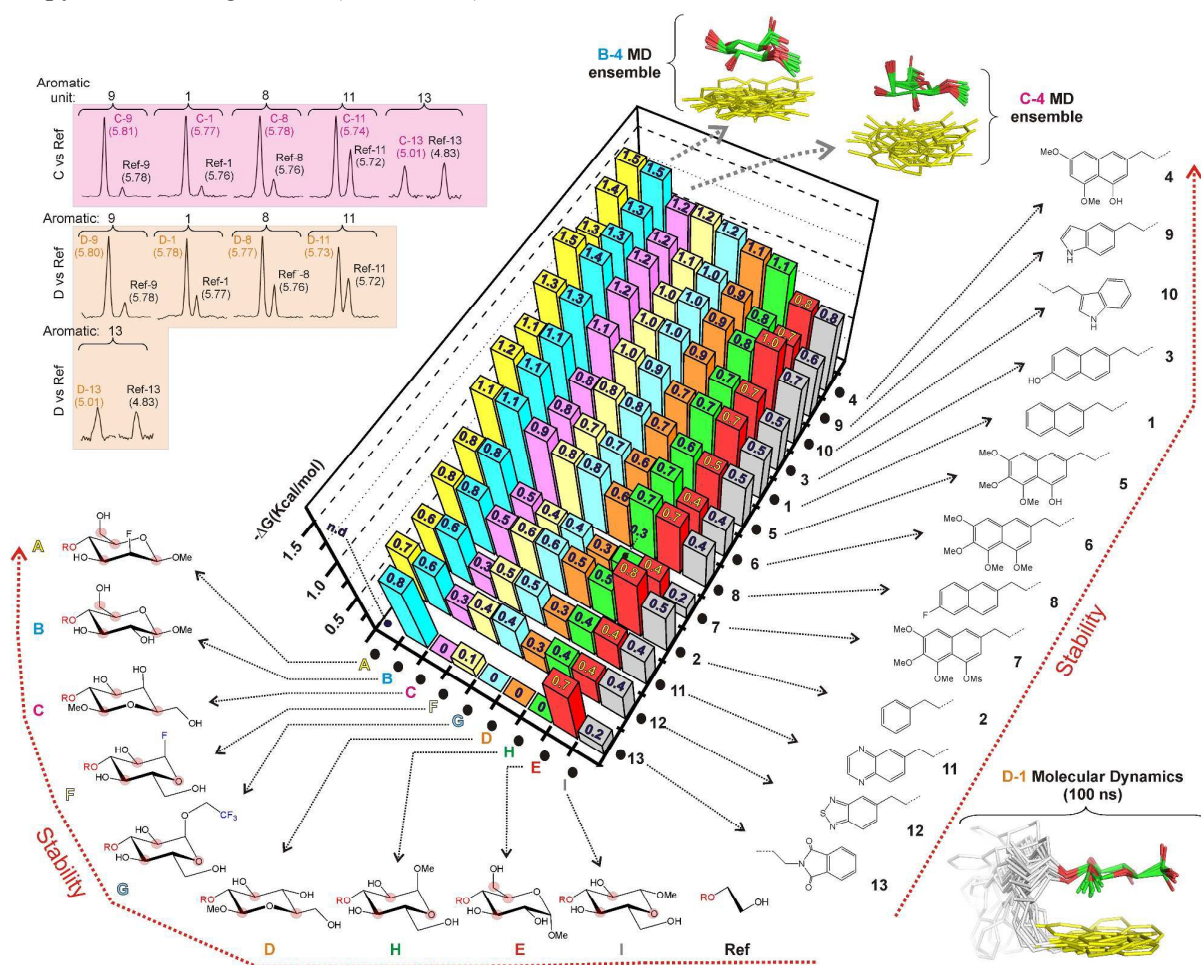
FIGURE 4.- Net stabilities (ΔG, kcal/mol) for the CH/π complexes established between donors **A-I** and the aromatic rings present in **1-13** measured from NMR pairwise competition experiments. Examples of experimental data sets are shown in the upper-left corner. The structural features of the different complexes were estimated from 100 ns MD simulations. Conformational ensembles obtained for products **B-4**, **C-4** and **D-1** through the MD simulations are also represented.

c) For non-interacting pyranose CH groups, axial polar substituents are generally preferred over equatorial ones. Complexes formed by pairs **C/D**, **A/B** or **H/I** provide an example of this rule. A

simple explanation for this behaviour is that equatorial polar moieties might be involved in re-pulsive interactions with the aromatic unit, and/or partially desolvated upon complex formation. As expected, the magnitude of the energy penalty associated to the axial/equatorial inversion depends on the precise geometrical features of the complex, varying from 0.3 kcal/mol (see **C-3** vs **D-3** in Figure 5b) to <0.1 kcal/mol (see **A-4** vs **B-4** in Figure 5b. The structural differences among these complexes, inferred from $\Delta\delta$ values, are outlined in Figure S7).

d) For glycosides with all-equatorial substituents, the α-face is, in most cases, preferred for interaction over the β-face. This selectivity is exemplified by the extra stability displayed by **D** complexes with respect to those formed by **I** (Figure 4 and Table S1). It should be noted that, in the former case, complex formation is only mediated by CH/π bonds. In contrast, for β-type interactions, the endocyclic oxygen becomes also involved in a n/π bond with the aromatic unit. According to our data, this later contact provides little contribution to the binding strength, being somewhat unfavourable for electron-rich aromatic systems.
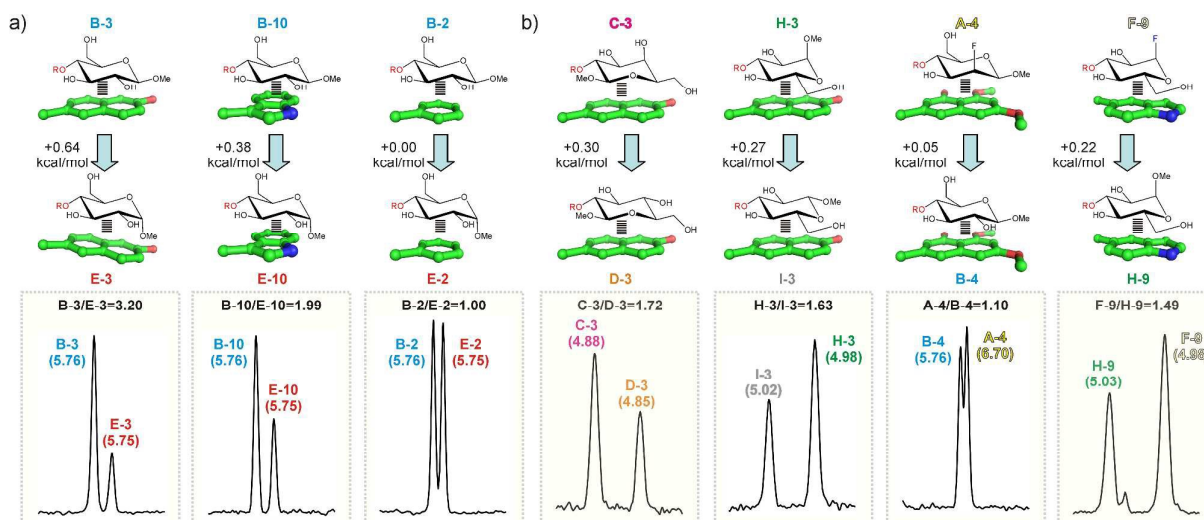
FIGURE 5.- Pairwise competition experiments between selected disaccharide pairs and aldehydes (a and b. See the main text). NMR signals for the altrose anomeric proton in the final products are shown in the bottom panel. Chemical shifts (ppm) are indicated above the signals together with the measured product ratios (black). The estimated free energy differences between the alternative CH/π complexes are represented in the upper panel.

e) Interactions mediated by the pyranose β face are sensitive to the electron density at the endocyclic oxygen, which, in turn can be modulated by attaching electron-withdrawing substituents

to the anomeric centre. This effect is clearly illustrated by the larger stability displayed by **F** or **G** complexes with respect to those formed by **H** (see for example **F-9** vs **H-9** in Figure 5b).

f) The stability of the CH/π bonds established by a simple aromatic ring (for example the naphthyl unit present in **1**) can be enhanced by the incorporation of oxygen substituents (Table S1, Figures 4 and S9). In agreement with this view, aromatic binding strengths follow the order **4>3>1**, as expected for increasing electron-rich substrates. However, this simple strategy presents limitations. It can be observed that those complexes formed by densely oxygenated aromatic systems such as **5-7** display reduced ΔG_{exp} values. Interestingly, quantum mechanical calculations (Figure S9) show that, in contrast to **4**, the OR groups present in **5-7** are not co-planar with the naphthyl unit and multiple rotamers are possible. Inspection of molecular models suggests that, for some of these conformational states, aromatic OR substituents severely restrict the internal mobility of the complex, imposing a significant entropic penalty on the recognition process (this hypothesis is supported by the theoretical evaluation of binding enthalpies for the different aromatic units. See below). Alternatively, they could participate in repulsive steric contacts with the pyranose unit.

g) The obtained data strongly suggest that electrostatics have a large influence on the aromatic binding ability. To illustrate this point, we calculated the electrostatic potential surfaces of different aromatic systems (densely oxygenated and conformationallly flexible systems **5-7** were not included in this analysis). Selected surfaces are represented in Figure 6a, together with the electric potential values at 2 Å above the centre of the rings (denoted as V_1 and V_2), which is approximately the equilibrium distance at which the CH interacts with the aromatic ring in the complexes (see quantum mechanical geometries below). It can be observed that, for any given disaccharide (as **D** and **I**. See the Figure 6b), ΔG_{exp} values correlate linearly with the sum of the aromatic electric potentials V_1+V_2. Indeed, according to these data, this simple parameter represents a reliable predictor of the ability of the aromatic unit to participate in CH/π interactions. Interestingly, the observed linear correlations exhibit distinct slopes for β-type and α-type complexes, implying that electron-rich aromatic aldehydes (as **9** or **4**) exhibit some selectivity for the α face of the pyranose unit, while electron-poor units (as **12** or **13**) are slightly β-selective (Figure 6b). This result was

12

further corroborated by pairwise competition experiments using a selected disaccharide pair (**D** and **I**. See Figure 6c) and both, electron-rich and electron-poor aldehydes (**9** and **12**). A simple explanation for this behaviour is that, for β-type complexes, electron-poor aromatic rings promote more stabilizing donor-acceptor n/π bonds with the pyranose endocyclic oxygen. Alternatively, for α-complexes, acceptor-donor CH/π bonds are dominant.
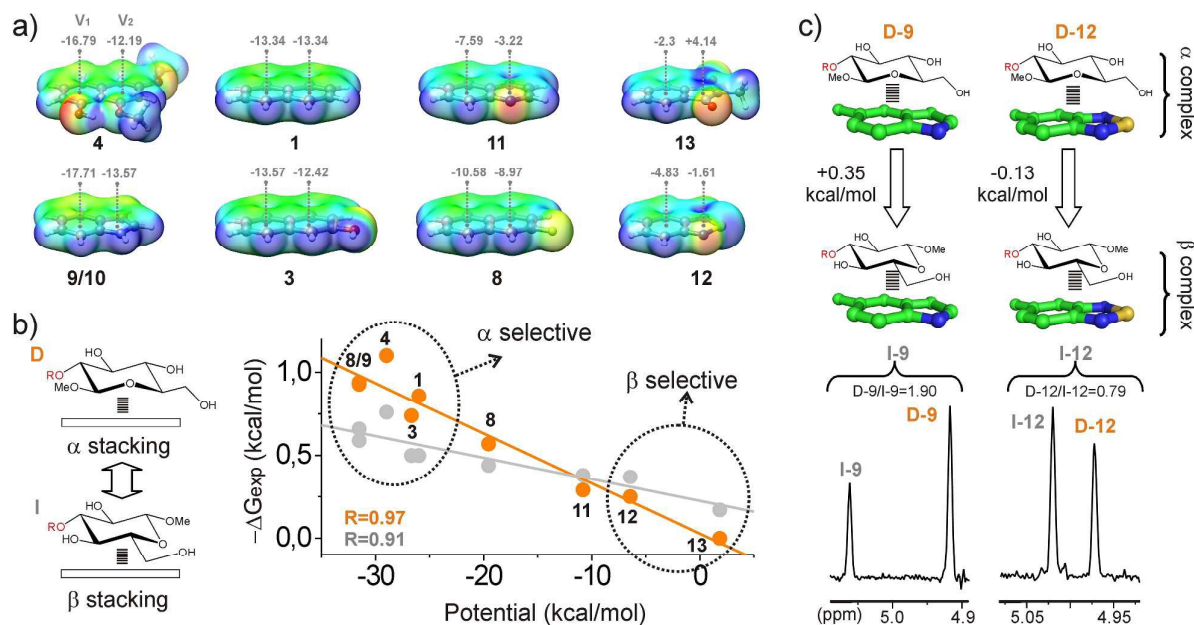
FIGURE 6.- a) Electrostatic potential surfaces calculated with M06-2X/TXVP level for selected aromatic systems, together with the electronic potential values (kcal/mol) at 2 Å above the centre of the rings (V_1 and V_2). b) Representation of ΔG_{exp} versus the aromatic potentials (V_1+V_2) for altrose derivatives **D** and **I**. Linear fits and correlation coefficients, (R) are shown. c) Pairwise competition experiments performed with disaccharide pair **D/I** and both, electron rich and poor aldehydes (**9** and **12**). NMR signals for the altrose anomeric proton in the final products are shown in the bottom panel, together with the measured product ratios (black). Free energy differences between the alternative CH/π complexes are represented in the upper panel.

D.- Dissecting the driving forces for the formation of CH/π complexes in water: Electrostatic vs dispersion components.- According to current knowledge, CH/π interactions in the gas phase are dominated by dispersive forces. Intriguingly, we have shown that, in water, the aromatic binding strengths are strongly correlated with the electrostatic potentials (Figure 6a,b). To get insights into the different contributions that stabilize CH/π complexes in this polar enviroment, a theoretical analysis was performed.

Quantum mechanical methods allowed us to assess the magnitude of the CH/π contacts established by the evaluated aromatic systems. Interaction enthalpies (ΔH_{QM}) were calculated at the M06-
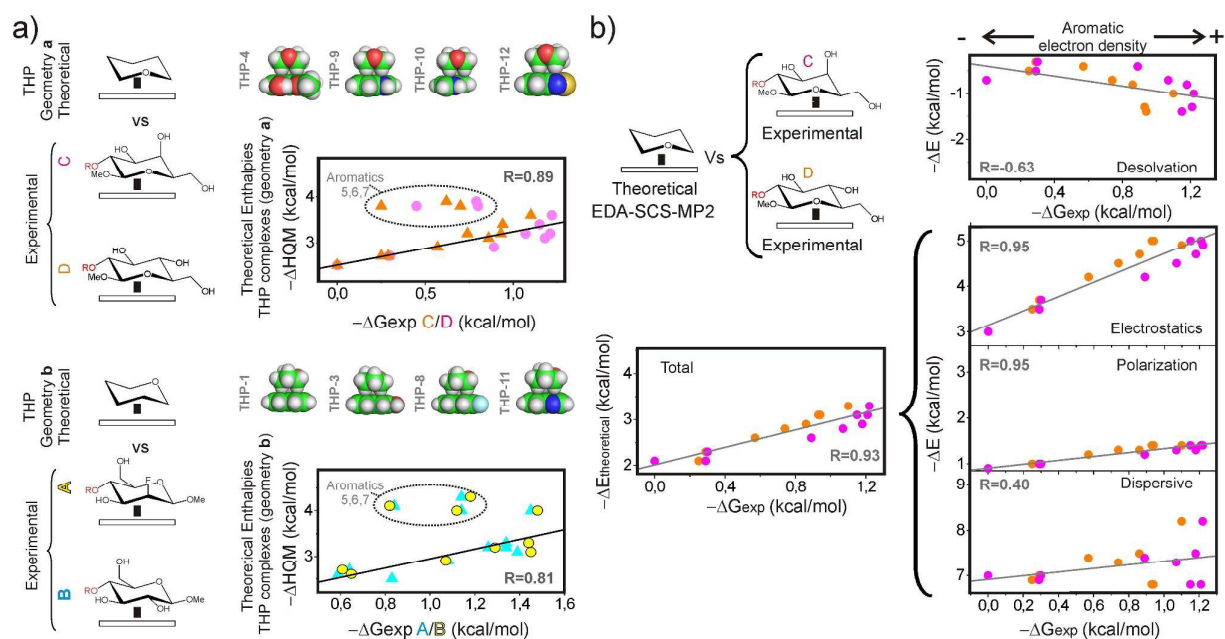
2X/TZVP level[14] in water (IEF-PCM method[15]) for simplified inter-molecular complexes formed by tetrahydropyrane (THP, as a pyranose model) and the aromatic units present in **1-13** (Figure 7a, Table S2, see experimental section for details). For each interacting pair, two orientations of the THP unit were considered (herein referred as a and b) to simulate the alternative geometries of the CH/π contacts established by derivatives **C/D** and **A/B/E**, respectively, upon reductive amination with the aldehydes. A representation of the calculated ΔH_{QM} values versus the free energies (ΔG_{exp}) measured for **C** and **D** (geometry a. Upper panel) and **A** and **B** (geometry b. Lower panel), together with the optimized geometries obtained for some of the complexes is shown in Figure 7a. It can be observed that the most significant deviations between both data sets correspond to the densely oxygenated aromatic systems **5**-**7**. As previously mentioned, complex formation by these units could display a larger entropic cost, which explains the lack of correspondence between ΔG_{exp} and ΔH_{QM} values. For the rest of the aromatic systems, ΔG_{exp} correlates linearly with ΔH_{QM} with good correlation coefficients (see Figure 7a).

The optimized geometries for the THP/aromatic complexes were employed to dissect the different contributions to the stability (Figure 7b and Table S2). To this aim, single-point interaction energies calculated at the SCS-MP2[16]/6-311G(2d,p) level in the gas phase were partitioned into electrostatic, polarization, exchange, repulsion and dispersion contributions, through the Localized Molecular Orbital Energy Decomposition Analysis (LMO-EDA).[17] In addition, the penalties associated to the THP and aromatic desolvation were evaluated at the same theory level (it should be noted that these theoretical values do not account for solvophobic contributions. See the experimental section for details).

These various contributions were represented versus the experimental ΔG_{exp} values measured in solution. The purpose of this comparison was to identify the main components of the interaction energies responsible for the experimentally observed trends. As an example, a representation of the experimental ΔG_{exp} values measured for **C** and **D** versus the interaction energies (both the total values and their contributions) calculated with SCS-MP2/6-311G(2d,p) level are shown in Figure 7b. Linear fits and correlation coefficients are also represented in all cases (see also Figure S10). As

observed in these plots, the predicted interaction energies show a good correspondence with the experimental ΔG_{exp} values. Regarding the energy contributions, different trends are apparent:

a) First, a weak and negative correlation is deduced for the energy penalty associated to the desolvation processes that accompany binding. This observation can be rationalized by considering that the stronger complexes usually involve more polar aromatic units, which in turn are more strongly solvated by water.

FIGURE 7.- a) Interaction enthalpies in water were calculated by quantum mechanics (M06-2X/TXVP level) for simplified inter-molecular complexes formed by tetrahydropyrane and the aromatic units present in **1-13** assuming two alternative geometries (herein referred as a and b. See the main text). These data are represented against the experimental free energy values (ΔG_{exp}) measured for derivatives **C** and **D** (geometry a. Up) and **A** and **B** (geometry b. Down). Linear fits and correlation coefficients (R) are indicated. b) Experimental ΔG_{exp} values measured for derivatives **C** (orange) and **D** (magenta) are represented against the gas phase interaction energies and the different contributions calculated with SCS-MP2/6-311G(2d,p) level for the THP/aromatic complexes (see the main text). Linear fits and correlation coefficients are shown in all cases.

b) Second, dispersion forces prevail among favourable contributions in the gas phase. However, this term show a poorer correlation with the solution experimental free energies (ΔG_{exp}) than the electrostatic and polarization terms. Of note, some of the most stable complexes (as **9** and **10**, involving an indole unit) display weaker dispersive forces. In addition, this contribution exhibits little variations for most of the complexes.

15

c) Finally, electrostatic and polarization contributions display a fairly good positive correlation with the experimental stabilities and show larger variations between strong and weak complexes.

In summary, although dominant in the gas phase, dispersion forces do not seem to account for the observed ΔG_{exp} variations in solution. On the contrary, these can be adequately justified in terms of electrostatic and polar interactions. All these observations strongly suggest that the experimental trends are, to a significant extent, dictated by electrostatics. Therefore, this term should be considered as an important determinant for the formation of CH/π bonds in water. In summary, our results indicate that, even in high dielectric environments, coulombic forces represent a key modulating influence for these complexes and, consequently, can be employed to predict and fine-tune the relative binding strengths.

E.- Dissecting the driving forces for the formation of CH/π complexes in water: Evidence for the relevance of solvophobic contributions.- The role of solvophovic forces on the stabilization of CH/π bonds in water was also evaluated. To this aim, we measured the effect of adding organic co-solvents on the strength of a selected complex (**C-3**). Figure 8 shows some of the pairwise competition experiments carried out with pair **C/Ref** and aldehyde **3** in the presence of increasing fractions (in the 0-30% range) of various organic co-solvents (indicated in black).

The data show that the addition of organic co-solvents is, in all cases, highly disruptive to the formation of the CH/π complex, with notable differences in the magnitude of the destabilization. Thus, 30% methanol decreases the free energy of association (ΔG_{exp}) from 1.18 down to 0.75 kcal/mol. In contrast, identical fraction of THF renders the interaction almost undetectable ($\Delta G_{exp} < 0.1$ kcal/mol). Overall, the disruptive power of the solvents follows the order MeOH<EtOH<DMSO<acetonitrile<THF. This observation is fully consistent with the solvent effects in carbohydrate binding by synthetic receptors, recently reported by Davis and co-workers.[6a] It should be noted that this experimental trend does not correlate with the dipolar moment ($\mu_{DMSO}=3.96 > \mu_{acetonitrile}=3.92 > \mu_{H2O}=1.85 > \mu_{THF}=1.75 > \mu_{MeOH}=1.70 > \mu_{EtOH}=1.69$) or dielectric constant ($\varepsilon_{H2O}=80.0 > \varepsilon_{DMSO}=46.7 > \varepsilon_{acetonitrile}=37.5 > \varepsilon_{MeOH}=33.0 > \varepsilon_{EtOH}=24.5 > \varepsilon_{THF}=7.5$) of the different solvent molecules. Moreover, it can not be easily rationalized only by invoking

16

competitive solute/solvent interactions (in fact, the relative stabilities of the alternative solvent/naphthalene complexes, estimated through quantum mechanics, does not provide support for this hypothesis. See Figure S11).

In order to demonstrate the relevance of the hydrophobic effect as a driving force for complex formation, the relationship between the carbohydrate/aromatic interaction energies (ΔG_{exp}) and the cohesive energy densities of the solvent media was evaluated (Figure 8).[13] It can be observed that the stronger interactions correspond to higher cohesive energy densities, while a decrease in this parameter usually implies a destabilization of the complex. In fact, both ΔG_{exp} values and cohesive energy densities exhibit an excellent linear correlation (correlation coefficient R=0.93).
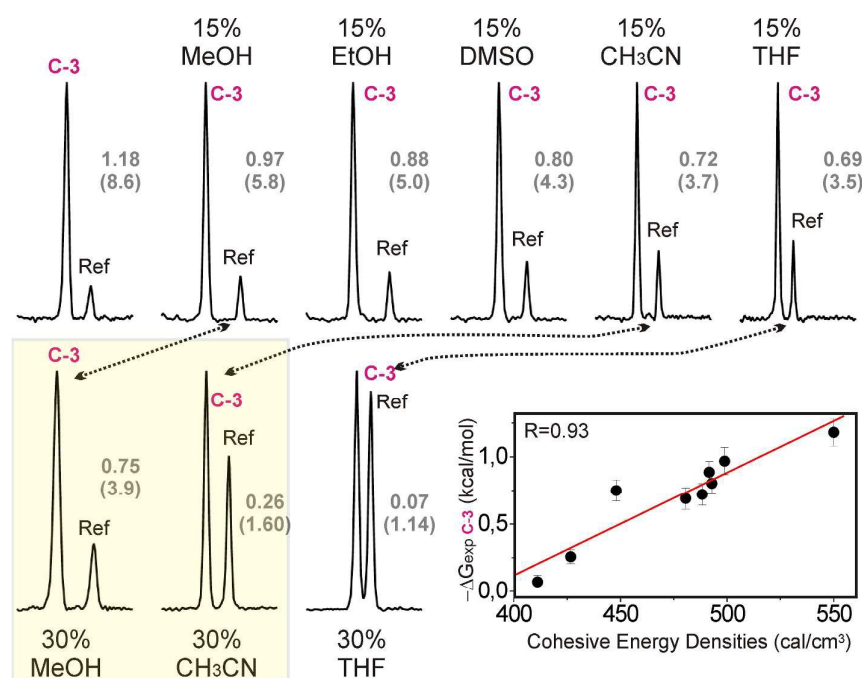
FIGURE 8.- Pairwise competition experiments performed with pair **C/Ref** and aldehyde **3** in the presence of organic co-solvents (indicated in black). NMR signals for the altrose anomeric proton in the final products are shown. Product ratios (in brackets) together with the **C-3** complex stabilities under the different solvent conditions are indicated in grey. Differences in the destabilization induced by methanol and acetonitrile are highlighted with a yellow square. A representation of the measured stabilities versus the cohesive energy densities of the different mixtures is shown in the bottom-right corner. Linear fit and correlation coefficient (R) are shown.

According to these data, the destabilizing influences of the organic co-solvents cannot be explained just in terms of solute/solvent competitive contacts. On the contrary, it seems to rely on a delicate balance between solute/solvent and solvent/solvent interactions. More specifically, the co-solvent ability to reduce the cohesive energy density of the mixture is a key feature to this process. Indeed

the key influence of this parameter on the solvation of non-polar solutes has been recently demonstrated.[18] Our results confirm that solvophobic forces play an important role in the stabilization of CH/π bonds in water.

Conclusions

An extensive analysis of CH/π interactions in water has been performed. The employed dynamic combinatorial methodology has allowed the accurate determination of the binding free energies for 117 isolated carbohydrate/aromatic complexes, presenting a large diversity of chemical features at both the CH/π donor and acceptor units. These data have revealed the most relevant structure-stability relationships that govern complex formation, opening the door to the rational design of improved carbohydrate-based ligands or carbohydrate receptors.

From a more fundamental perspective, our analysis has provided insights into the origin of these contacts. In particular, our results highlight the key role of electrostatics on the stabilization of CH/π complexes in water. Indeed, most of the free energy variations herein described can be rationalized in terms of simple coulombic interactions between the pyranose and the aromatic units. Thus, in most cases, attractive or repulsive polar interactions can be invoked to explain how the stability of the complexes is modulated by the orientation of the pyranose electron-withdrawing substituents. In addition, it also clarifies the observed facial selectivity and how this parameter is affected by the electron-rich or electron-poor character of the aromatic unit. Finally, the relative binding strengths of the different aromatic systems (in the 0-1.5 kcal/mol range) can be anticipated from their electronic properties.

Regarding dispersion forces, they have been traditionally considered to be dominant in the gas phase. However, according to recent studies this consideration might not hold for a physiological environment.[13] While their relative weight in solution cannot be inferred from this work, our results suggest that this term exhibits smaller variations for systems of similar sizes and, therefore can not be employed to effectively tune the binding free energies.

Finally, we have demonstrated that the strength of the CH/π bonds not depends only on the donor/acceptor contacts. Indeed, solute/solvent and, most importantly, solvent-solvent interactions have to be taken into account. Our results indicate that solvophobic interactions provide a significant contribution to the complex stability.

Overall, the data presented herein provide key clues, at the general and specific levels, to understand and modulate biomolecular interaction processes in which aromatics are involved, with particular interest for the molecular recognition of glycans.

Electronic Supplementary Information available

Synthesis of disaccharides and aldehydes, NMR competition experiments and theoretical calculations. Figures S1-S11. Tables S1 and S2.

Methods

Experimental procedures for the synthesis of our libraries are described in the Supplementary Material.

Dynamic combinatorial experiments.- As a first step, reductive amination reactions were performed in NMR tubes with single altrosamine derivatives. Thus, a mixture of the saccharide (~1 mM) and aldehyde (~2 mM) were dissolved in D_2O (10 mM phosphate, pH 6.2) and treated with sodium cyanoborohydride at room temperature. After reaction completion, final products were assigned employing a combination of 2D-COSY, TOCSY and HSQC experiments. From these data, we determined the chemical shift perturbations promoted by the aromatic units on the CH/π

donor pyranoses. The obtained Δδ (ppm) values were taken as an indicative of the geometry of the different stacking complexes (Figures S3 and S5-S7).

Next, equimolecular mixtures of 6-8 altrosamine derivatives (~1 mM each) in D_2O (10 mM phosphate, pH 6.2) were treated with a sub-stoichiometric amount of a given arylacetaldehyde (compounds **1**-**13** in Figure 2b). The aldehyde concentration was carefully adjusted to achieve maximum conversions of the altrosamine derivatives not superior to 40% (usually in the 400-800 μM range depending on the number of components present in the mixture). All dynamic combinatorial assays were carried out at low temperature in order to maximize the stability of the CH/π complexes and consequently, the sensitivity of the experiments. Thus, mixtures were left to equilibrate for 2 hours at 277 K, and then we added sodium cyanoborohydride (5 mM). Reactions were kept overnight at 277 K. After reaction completion, we performed a qualitative quantification of the reaction products by focusing our attention on the altrosamine anomeric signals in 2D-HSQC experiments acquired at 318 K.

In order to evaluate more precisely the stability of the different CH/π complexes considered in our study, we performed pairwise competition experiments following the same protocol. The quantification of the relative populations of the final products was accomplished by integrating key NMR signals in 1D experiments acquired at 318 K with long relaxation delays (d1>5). Overlapping signals were deconvoluted before integration employing the line-fitting module implemented in MestreNova software. The obtained values reflect the population of the corresponding imine/hemiaminal intermediates and therefore, were employed to derive the relative stability (ΔΔG) of the alternative stacking modes according to $\Delta\Delta G = R \cdot 277 \cdot Ln(P_1/P_2)$ (where P_1/P_2 represents the population ratio and 277 is the reaction and equilibration temperatures in K). This protocol has been recently validated by our group employing structurally related allosamine disaccharides (for which imine/hemiaminal intermediates could be detected and their populations measured).[5]

We performed pairwise competitions with all possible **Ref**/disaccharide pairs so that the net free energy (ΔG) of the stacking complexes could be determined. Selected data points were derived from at least two independent reactions in order to estimate errors. In these cases, product

populations were also measured from NMR spectra acquired with distinct relaxation delays. Experimental errors were found to be dependent on the population ratios (P_1/P_2). Errors upper bounds were estimated as ±10% for $1<P_1/P_2<3$, ±15% for $3<P_1/P_2<6$ and ±20% for $6<P_1/P_2$. At 277 K these values translate into energy ranges of ±0.05, ±0.08, ±0.11 kcal/mol, respectively (Table S1).

In addition to the net interaction energies (ΔG), we also determined free energy differences ($\Delta\Delta G$) of all the complexes with respect to those formed by **D**, by performing the corresponding pairwise competitions. Finally, many other crosschecks were carried out with selected aldehydes and disaccharide pairs (Figure S4). The obtained $\Delta\Delta G$ values were found to be fully consistent with the net interaction energies (ΔG) and errors represented in Table S1.

Quantum Mechanical Calculations.- Full geometry optimizations for the different THF/aromatic complexes were carried out with Gaussian 09[19] using the M06-2X functional[14] with an ultrafine integration grid and the triple-zeta TZVP basis set.[20] Two interaction modes, named as a and b, were assumed in each case to simulate the alternative geometries of complexes formed by derivatives **C**/**D** and **A**/**B**/**E**, respectively, upon the reductive amination reaction. Solvent effects were included in the optimization through the integral equation formalism variant of the polarizable continuum model (IEF-PCM),[15] using UFF radii and parameters for water, as implemented in Gaussian 09. Frequency analyses were carried out at the same level used in the geometry optimizations to evaluate the zero-point vibrational energy and thermal corrections at 298 K necessary to derive enthalpies and free energies. The nature of the stationary points was determined according to the appropriate number of negative eigenvalues of the Hessian matrix.

Gas-phase single-point energies at the spin-component scaled MP2 (SCS-MP2) level[16] with the triple-zeta 6-311G(2d,p) basis set, and also at the M06-2X/TZVP level (see Figure S10), were calculated on the optimized geometries using Gamess (v.2014).[21] Basis set superposition errors (BSSE) were corrected by the Boys–Bernardi counterpoise method.[22] Solvation free energies (ΔG_{solv}) were calculated at the ab initio level using the IEF-PCM method using the parameters for water as implemented in Gamess. Various contributions to the interaction energies (electrostatic,

exchange, repulsion, polarization and dispersion) were estimated through the Localized Molecular

Orbital Energy Decomposition Analysis (LMO-EDA) method[17] as implemented in Gamess.

*Molecular dynamics calculations.-*The conformational and dynamical properties of selected

complexes (Figure S8) were estimated through Molecular Dynamics (MD) simulations performed

with the AMBER 12 package.[23] 100 ns trajectories were collected in the presence of explicit TIP3P

water,[24] periodic boundary conditions and Ewald sums for the treatment of long-range electrostatic

interactions,[25] following a protocol identical to that previously described.[26] In all cases, RESP

atomic charges were derived by applying the RESP module of AMBER to the HF/6-31G(d) ESP

charges calculated with Gaussian 09. The ffSB14 force field[27] was implemented with

GLYCAM06[28] and GAFF[29] parameters to accurately simulate the conformational behavior of these

molecules. The timestep was 1 fs in all the simulations.

References

1   E. A. Meyer, R. K. Castellano and F. Diederich, *Angew. Chem. Int. Ed.*, 2003, **42**, 1210.

2   (a) M. Nishio, M. Hirota and M. Umezawa, *The CH/π interaction. Evidence, Nature, and Consequences*, Wiley-VCH, New York, 1998; (b) M. Nishio, *Chem. Chem., Phys.*, 2011, **13**, 13873; (c) M. Nishio, Y. Umezawa, J. Fantini, M.S. Weiss and P. Chakrabarti, *Phys. Chem. Chem. Phys.*, 2014, **16**, 12648.

3   (a) S. Tsuzuki and A. Fujii, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2584; (b) A. Fujii, K. Shibasaki, T. Kazam, R. Itaya, N. Mikami and S. Tsuzuki, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2836.

4   (a) R. Carrillo, M. López-Rodriguez, V.S. Martín and T. Martín, *Angew. Chem. Int. Ed.*, 2009, **48**, 7803; (b) A. E. Aliev, J. R. T. Arendorf, I. Pavlakos, R. B. Moreno, M. J. Porter, H. S. Rzepa and W. B. Motherwell, *Angew. Chem. Int. Ed.*, 2015, **54**, 551; (c) A. Scarso, L. Trembleau and J. Rebeck, *J. Am. Chem. Soc.*, 2004, **126**, 13512; (d) B. Bhayana and C. S. Wilcox, *Angew. Chem. Int. Ed.*, 2007, **46**, 6833; (e) M. J. Plevin, D. L. Bryce and J. Boisbouvier, *Nat. Chem.*, 2010, **2**, 466; (f) M. Harigai, M. Kataoka, and Y. Imamoto, *J. Am. Chem. Soc.*, 2006, **128**, 10646.

5   E. Jimenez-Moreno, A. M. Gomez, A. Bastida, F. Corzana, G. Jiménez-Osés, J. Jiménez-Barbero and J.L. Asensio, *Angew. Chem. Int. Ed.*, 2015, **54**, 4344.

6   (a) E. Klein, Y. Ferrand, N. P. Barwell and A. P. Davis, *Angew. Chem. Int. Ed.*, 2008, **47**, 2693; (b) Y. Ferrand, M. P. Crump and A. P. Davis, *Science,* 2007, **318**, 619; (c) C. Ke, H. Destecroix, M. P. Crump and A. P. Davis, *Nat. Chem.* 2012, **4**, 718; (d) Y. Ferrand, E. Klein, N. P. Barwell, M. P. Crump, J. Jiménez-Barbero, C. Vicent, G.-J. Boons, S. Ingale and A. P. Davis, *Angew. Chem. Int. Ed.*, 2009, **48**, 1775; (e) N. P. Barwell and A. P. Davis, *J. Org. Chem.*, 2011, **76**, 6548.

7   (a) J. L. Asensio, A. Ardá, F. J. Cañada and J. Jiménez-Barbero, *Acc. Chem. Res.*, 2013, **46**, 946; (b) G. Terraneo, D. Potenza, A. Canales, J. Jiménez-Barbero, K. K. Baldridge and A. Bernardi, *J. Am. Chem. Soc.*, 2007, **129**, 2890; (c) M. C. Fernández-Alonso, F. J. Cañada, J. Jimenez-Barbero and G. Cuevas, *J. Am. Chem. Soc.*, 2005, **127**, 7379; (d) K. Ramírez-Gualito, R. Alonso-Ríos, B. Quiroz-García, A. Rojas-Aguilar, D. Díaz, J. Jiménez-Barbero, G. Cuevas, *J. Am. Chem. Soc.*, 2009, **131**, 18129; (e) M. Chavez, C. Andreu, P. Vidal, N. Aboitiz, F. Freire, P. Groves, J. L. Asensio, G. Asensio, M. Muraki, F. J. Cañada and J. Jimenez-Barbero, *Chem. Eur. J.*, 2005, **11**, 7060.

8   (a) A. G. Santana, E. Jiménez-Moreno, A.M. Gómez, F. Corzana, C. González, G. Jiménez-Osés, J. Jiménez-Barbero, J. L. Asensio, *J. Am. Chem. Soc.*, 2013, **135**, 3347; (b) T. Vacas, F. Corzana, G. Jiménez-Osés, C. González, A.M. Gómez, A. Bastida, J. Revuelta and J. L. Asensio, *J. Am. Chem. Soc.*, 2010, **132**, 12074.

9   (a) J. Screen, E. C. Stanca-Kaposta, D. P. Gamblin, B. Liu, N. A. Macleod, L. C. Snoek, B. G. Davis and J. P. Simons, *Angew. Chem. Int. Ed.*, 2007, **46**, 3644; (b) E. J. Cocinero, P. Carcabal, T. D. Vaden, B. G. Davis and J. P. Simons, *J. Am. Chem. Soc.*, 2011, **133**, 4548; (c) Z. Su, E. J. Cocinero, E. C. Stanca-Kaposta, B. G. Davis and J. P. Simons, *Chem. Phys. Lett.*, 2009, **471**, 17.

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

11   (a) Z. R. Laughrey, S. E. Kiehna, A. J. Riemen and M. L. Waters, *J. Am. Chem. Soc.*, 2008, **130**, 14625; (b) S. E. Kiehna, Z. R. Laughrey and M. L. Waters, *Chem. Commun.*, 2007, **39**, 4026.

12   R. Lucas, I. Gomez-Pinto, A. Aviño, J. J. Reina, R. Eritja, C. González and J. C. Morales, *J. Am. Chem. Soc.*, 2011, **133**, 1909.

13   (a) L. Yang, C. Adam, G. S. Nichol and S. L. Cockroft, *Nat. Chem.*, 2013, **5**, 1006; (b) C. Adam, L. Yang and S. L. Cockroft, *Angew. Chem. Int. Ed.*, 2015, **54**, 1164.

14   Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215.

15   G. Scalmani and M. J. Frisch, *J. Chem. Phys.*, 2010, 132.

16   S. J. Grimme, *Chem. Phys.*, 2003, **118**, 9095.

17   P. Su and H. Li, *J. Chem. Phys.*, 2009, **131**, 1.

18   S. Otto, *Chem. Sci.,* 2013, **4**, 2953.

19   J. A. Pople, et al. GAUSSIAN 09, Revision D.01, Gaussian, Inc., Wallingford CT, 2009.

20   (a) A. Schaefer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571; (b) A. Schaefer, C. Huber, R. Ahlrichs, *J. Chem. Phys.*, 1994, **100**, 5829.

21   (a) M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347; (b) M. S. Gordon and M. W. Schmidt, in Theory and Applications of Computational Chemistry, the first forty years. Chapter 41, pp 1167–1189, C. E. Dykstra, G. Frenking, K. S. Kim and G. E. Scuseria, editors. Elsevier, Amsterdam, 2005.

22   S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553.

23 D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2012), AMBER 12, University of California, San Francisco.

24 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L Klein, *J. Chem. Phys*., 1983, **79**, 926.

25 T. Darden, D. York and L. Pedersen, *J. Chem. Phys*., 1993, **98**, 10089.

26 D. Madariaga, N. Martínez-Sáez, V. J. Somovilla, L. García-García, M. A. Berbis, J. Valero-Gónzalez, S. Martín-Santamaría, R. Hurtado-Guerrero, J. L. Asensio, J. Jiménez-Barbero A. Avenoza, J. H. Busto, F. Corzana and J. M. Peregrina, *Chem. Eur. J*., 2014, **20**, 12616.

27 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712.

28 K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods, *J. Comput. Chem*., 2008, **29**, 622.

29 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput Chem*., 2004, **25**, 1157.