J|A|C|S
ARTICLES
Published on Web 04/29/2005

Molecular Recognition of Saccharides by Proteins. Insights on the Origin of the Carbohydrate−Aromatic Interactions

María del Carmen Fernández-Alonso,[†] Francisco Javier Cañada,[†] Jesús Jiménez-Barbero,*[,†] and Gabriel Cuevas*[,‡]

Contribution from the Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, 28040 Madrid, Spain, and Instituto de Química, Universidad Nacional Autónoma de México, Apdo. Postal 70213, 04510, Coyoacán, Circuito Exterior, México DF

Received February 17, 2005; E-mail: jjbarbero@cib.csic.es; gecgb@servidor.unam.mx

Abstract: The existence of stabilizing carbohydrate−aromatic interactions is demonstrated from both the theoretical and experimental viewpoints. The geometry of experimentally based galactose-lectin complexes has been properly accounted for by using a MP2/6-31G(d,p) level of theory and by considering a counterpoise correction during optimization. In this case, the stabilizing interaction energy of the fucose−benzene complex amounts to 3.0 kcal/mol. The theoretical results obtained herein indicate that the carbohydrate−aromatic interactions are stabilizing interactions with an important dispersive component and that electronic density between the sugar hydrogens and the aromatic ring indeed exists, thus giving rise to three so-called nonconventional hydrogen bonds. Experimental evidence of the intrinsic tendency of aromatic moieties to interact with certain sugars has also been shown by simple NMR experiments in water solution. Benzene and phenol specifically interact with the clusters of C−H bonds of the alpha face of methyl β-galactoside, without requiring the well-defined three-dimensional shape provided by a protein receptor, therefore resembling the molecular recognition features that are frequently observed in many carbohydrate−protein complexes.

Introduction

Molecular recognition by specific targets is at the heart of the life processes. In recent years, it has been shown that the interactions between carbohydrates and proteins mediate a broad range of biological activities, starting from fertilization, embryogenesis, and tissue maturation, and extending to such pathological processes as tumor metastasis. The elucidation of the mechanisms that govern how oligosaccharides are accommodated in the binding sites of lectins, antibodies, and enzymes is currently a topic of major interest.[1] It is now recognized that the single molecule sugar−protein interactions are weak in nature and that multivalency is a key feature for the molecular recognition process to take place.[2] The physicochemical nature of sugar−protein interaction has been a matter of debate for years.[3] Because of the amphipatic character of the oligosaccharide, different kinds of forces may be involved in its recognition by receptors. The presence of the hydroxyl groups obviously makes possible their involvement in intermolecular hydrogen bonds to side chains of polar amino acids.[4] Moreover, it has also been postulated that water provides the driving force for the intermolecular interaction.[5] Nevertheless, not only polar forces are involved in carbohydrate recognition.[6] NMR and X-ray diffraction[7] data have shown that depending on the stereochemistry of the saccharide, the presence of many rather nonpolar C−H groups indeed constitute patches that interact with the aromatic residues of protein side chains. However, to the best of our knowledge, the detailed nature of these carbohydrate−aromatic interactions has not been properly and deeply investigated.[8] On this basis, we herein present the first in-depth study of the physicochemical origin of the intermolecular interaction between sugar and aromatic rings, using the complex between fucose (Fuc) and benzene as the model structure and employing different levels of theory. It is demonstrated that the complex is stable, that the CH/π interaction is indeed stabilizing, and that it features an important dispersion component. Finally, experimental evidences that support a preferred orientation of aromatic rings when interacting with saccharides are provided by NMR data on sugar−benzene and sugar−phenol interactions in water solution.

[†] Consejo Superior de Investigaciones Científicas.
[‡] Universidad Nacional Autónoma de México.

(1) Gabius, H. J.; Siebert, H. C.; André, S.; Jiménez-Barbero, J.; Rüdiger, H. *ChemBioChem* **2004**, *5*, 740−764.
(2) (a) Kitov, P. I.; Bundle, D. R. *J. Am. Chem. Soc.* **2003**, *125*, 16271−16284. (b) Thobhani, S.; Ember, B.; Siriwardena, A.; Boons, G.-J. *J. Am. Chem. Soc.* **2003**, *125*, 7154−7155.
(3) Sharon, N.; Lis, H. *Adv. Exp. Med. Bull.* **2001**, *491*, 1−16.
(4) Lis, H.; Sharon, N. *Chem. Rev.* **1998**, *98*, 637−674.
(5) Lemieux, R. U. *Acc. Chem. Res.* **1996**, *29*, 373−380.
(6) Sörme, P.; Arnoux, P.; Kahl-Knutsson, B.; Leffler, H.; Rini, J. M.; Nilsson, U. J. *J. Am. Chem. Soc.* **2005**, *127*, 1737−1743.
(7) (a) *NMR Spectroscopy of Glycoconjugates*; Jiménez-Barbero, J., Peters, T., Eds.; Wiley-VCH: Weinheim, Germany, 2003. (b) Wormald, M. R.; Petrescu, A. J.; Pao, Y. L.; Glithero, A.; Elliott, T.; Dwek, R. A. *Chem. Rev.* **2002**, *102*, 371−386.
(8) (a) Palma, R.; Himmel, R. E.; Brady, J. W. *J. Phys. Chem. B* **2000**, *104*, 7228−7234. (b) Spiwok, V.; Lipovova, P.; Skalova, T.; Buchtelova, E.; Hasek, J.; Kralova, B. *Carbohydr. Res.* **2004**, *339*, 2275−2280. (c) Sujatha, M. S.; Sasidhar, Y. U.; Balaji, P. V. *Prot. Sci.* **2004**, *13*, 2502−2514.

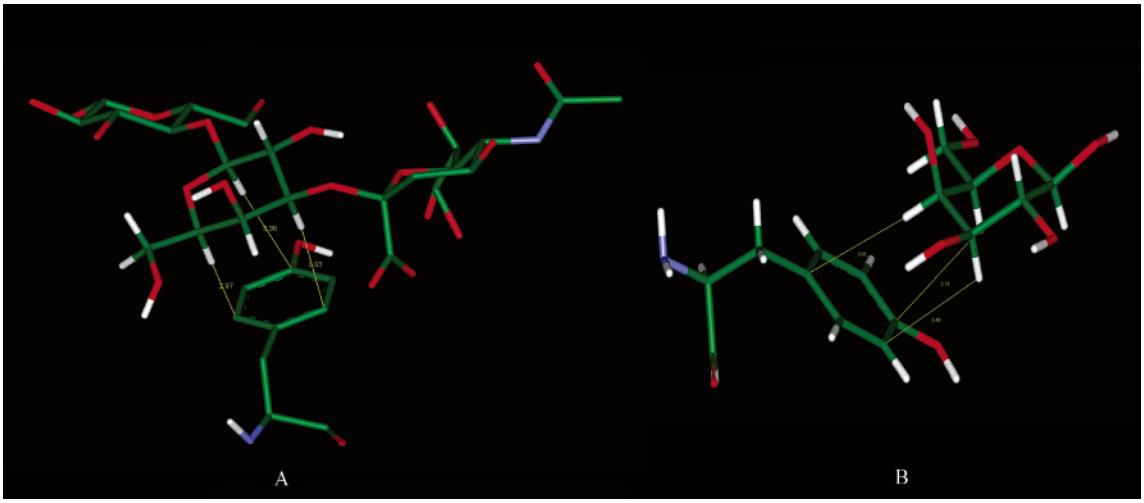

Figure 1. Examples of structures from the lectins database:[10] (A) influenza virus hemagglutinin complexed with 2,3-sialyllactose (1,3,5-type complex), PDB code 1HGG;[10c] (B) lectin II from Ulex europaeus (UEA-2) complexed with galactose (3,4,5-type complex), PDB code 1DZQ.[10d]

Experimental Section

NMR Experiments. ^{1}H NMR monodimensional spectra were recorded for 1−10 mM solutions of methyl β-galactoside and methyl α-mannoside in D_2O solution at 298 K on a Bruker Avance 500 MHz spectrometer. Sixty-four scans were taken with a 3-s relaxation delay. In all cases, TSP was used as external reference. Experiments for both sugars, as well as for a 1:1 mixture of both, were recorded both in the absence and presence of 10 mM benzene. Additional spectra were recorded for the three samples (methyl β-galactoside, methyl α-mannoside, and its 1:1 mixture in D_2O solution) with increasing concentrations of phenol (up to 200 mM).

Theoretical Calculations

The calculations were performed on HP Cluster Superdome and Compaq HPC 320 computers, at the Supercomputing Center of Galicia, Spain (CESGA).

Full geometry optimization was performed with Gaussian98[9a] and Gaussian03[9b] for the Fuc−benzene complex at four different levels of theory, using density functional theory (DFT) and Møller−Plesset calculations. In particular, the levels were (A) MP2/6-31G(d,p), (B) B3LYP/6-31G(d,p), (C) MP2/6-31G(d,p), corrected by the "counterpoise" (CP) method, and (D) MP2/6-31G(d,p)//B3LYP/ 6-31G(d,p). Vibrational frequency calculations were performed to characterize the nature of the stationary points. As input structure, the coordinates obtained from a crystallographic sugar−protein complex featuring a galactose/tyrosine interaction were employed[10] (1HGG pdb code[10c]) and modified as required. Fucose is a 6-deoxy galactose, which simplifies the calculations by removing the additional conformers originated by rotation of the hydroxymethyl group, while the aromatic ring of benzene was obtained from the phenol ring of tyrosine. For the calculations employing the MP2/6-31G(d,p)//B3LYP/6-31G(d,p) level of theory, the resulting structure from the B3LYP/6-31G(d,p) optimization was employed, and a single point (SP) calculation was performed.

According to notation, the benzene ring carbon atoms are denoted by primes (C1′−C6′), while the sugar atoms keep their regular numbering (see Figure 2).

In addition, potential energy curves were calculated for the complex, at all levels of theory, to verify the stabilizing (or not) nature of the interaction. Thus, SP calculations were performed with Gaussian 98, varying the distance between a given CH bond and the closest corresponding carbon atom. To precisely determine the interaction energy, the basis set superposition error correction (BSSE)[11] was calculated. The counterpoise method proposed by Boys and Bernardi[12] was used, and thus the proper correction by the change of geometry of the components of the complex was considered. The method proposed by Xantheas,[13] that permits optimization of the geometry by considering the BSSE, as implemented in Gaussian03, was also employed.

Finally, the atoms in molecules theory (AIM)[14] was also used to study the topological properties of the electronic density of the supramolecule. With AIM, it is possible to define structure and stability in terms of the gradient field vector, $\nabla\rho$, associated with a scalar field, such as ρ. Thus, critical points, where $\nabla\rho = 0$, can be deduced. There is a set of trajectories associated with each critical point that define an interatomic surface, which separates the basins of neighboring atoms. Bond, ring, and cage critical points were deduced for the supramolecule at each level of theory, using the AIMPAC software.[15]

Results

The pdb coordinates[10] for a galactose/tyrosine interaction, which does exist for influenza virus hemagglutinin interacting with 2,3-sialyllactose (1HGG pdb code[10c]), were isolated and modified as required to produce the target complex (fucose/benzene). The corresponding coordinates were employed as input structure for the calculations. Even when the approximation of the aromatic molecule can occur in an axial-oriented manner to produce a complex, in which the three hydrogen atoms at 1,3,5-syn-triaxial relative positions (A, Figure 1) simultaneously interact with the aromatic ring, a detailed

(9) (a) Frisch, M. J. et al.; *Gaussian 98*, Revision A.11.3; Gaussian, Inc.: Pittsburgh, PA, 1998. (b) Frisch, M. J.; et al.; *Gaussian 03*, Revision A.1; Gaussian, Inc.: Pittsburgh, PA, 2003.
(10) (a) For the 3D-lectin database, see Lectines (http://webenligne.cermav.cnrs.fr/lectines/), accessed Apr. 25, 2005. (b) For the Protein Data Bank, see Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242. (c) Sauter, N. K.; Hanson, J. E.; Glick, G. D.; Brown, J. H.; Crowther, R. L.; Park, S. J.; Skehel, J. J.; Wiley, D. C. *Biochemistry* **1992**, *31*, 9609−9621. (d) Loris, R.; De Greve, H.; Dao-Thi, M.-H.; Messens, J.; Imberty, A.; Wyns, L. *J. Biol. Chem.* **2000**, *301*, 987−1002.
(11) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024−11031.
(12) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(13) (a) Xantheas, S. S. *J. Chem. Phys.* **1996**, *104*, 8821−8824. For recent applications of the counterpoise corrected geometry optimization, see, for instance: (b) Hunt, S. W.; Higgins, K. J.; Craddock, M. B.; Brauer, C. S.; Leopold, K. R. *J. Am. Chem. Soc.* **2003**, *125*, 13850−13860. (c) Li, R. Y.; Li, Z..; Wu, D.; Li, Y.; Chen, W.; Sun, C. C. *J. Chem. Phys.* **2004**, *121*, 8775−8781.
(14) (a) Bader, R. F. W. *Atoms in Molecules, a Quantum Theory;* Clerendon Press: Bern, Switzerland, 1990. (b) Bader, R. F. W. *Atoms in Molecules in Encyclopedia of computational Chemistry*; Schleyer, P.v.R., Ed.; John Wiley & Sons, 1998; p 64.
(15) Biegler-Koing, F. W.; Bader, R. F. W.; Tang, T. H. *J. Comput. Chem.* **1982**, *3*, 317−328.

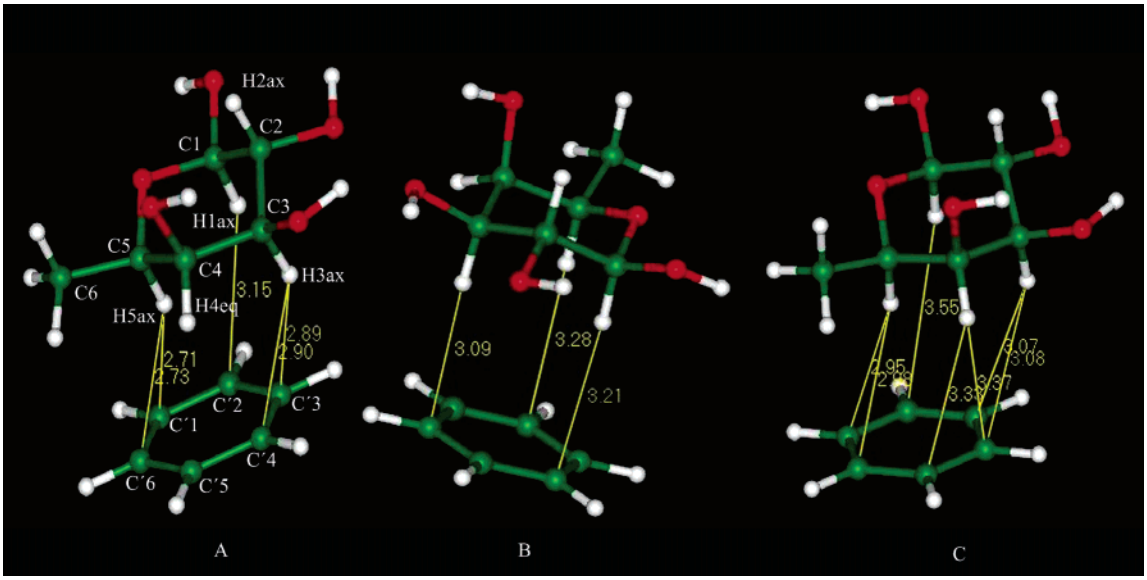

Figure 2. Stationary states of the supramolecule composed by fucose and benzene. The key distances (in Å) from CH groups of the sugar to the aromatic ring are shown. (A) MP2/6-31G(d,p); (B) B3LYP/6-31G(d,p); (C) MP2/6-31G(d,p) corrected by CP.

analysis of many reported X-ray structures for galactose-containing complexes (as that studied herein, with 6-deoxygalactose, fucose) shows that there is a general trend, where the aromatic ring preferentially interacts with the hydrogen atoms that alternate axial and equatorial positions (H-3, H-4, and H-5) in a different relative arrangement (such as B, Figure 1).[10]

In the minimum (Figure 2A) at the MP2/6-31G (d,p) level, and regarding the orientation of the hydroxyl groups of fucose, many hydrogen bonds are presented in such a way that the hydroxyl group at the anomeric position (C1) adopts an orientation that allows for the exoanomeric stabilizing interaction.[16] Thus, it is acting as a hydrogen bond acceptor from the hydroxyl group at position C2. The hydroxyl group at position C2 is acting as a hydrogen bond donor to the C1 hydroxyl and as an acceptor from that at position C3, which is equatorially oriented. The hydroxyl group at C3 is acting as a hydrogen bond donor to the C2 hydroxyl and as an acceptor from that at position C4, axially oriented. This counterclockwise arrangement of hydrogen bonds was kept constant throughout this study.

Regarding the Fuc−benzene supramolecule, the hydrogen atom H5, at C5, is located at 2.396 Å in relation to the symmetry point of the benzene ring and at a distance that oscillates between 2.712 and 2.831 Å in relation to all the benzene carbon atoms. It can be observed (A, Figure 2) that the H1 hydrogen atom is located at 3.150 and 3.203 Å from carbons C′2 and C′3, respectively, while the H3 atom is located at 2.894 Å from C′3 and at 2.896 Å from C′4. Thus, the shorter distances are observed for H3 and H5, followed by H1. Fuc H4 is further apart (located at 3.235, 3.229, and 3.973 Å from C′4, C′5, and C′6, respectively).

A second level of theory was used for the calculations. It is well-known that the hybrid method, based on the density functional theory, B3LYP[17] is not fully adequate to study long-distance interactions, since its consideration of the dispersion term is somehow deficient.[18] Therefore, the effect of the absence of this term in the determination of the geometry of the supramolecular complex can be estimated. Thus, the system was optimized at the B3LYP/6-31G(d,p) level. The calculated minimum (without imaginary frequencies) features a substantial change in the molecular geometry, as shown in B, Figure 2. The complex maintains a 1,3,5 axial-type approximation, but with much longer C−H distances between the Fuc and benzene moieties and with eclipsed rings. Each fucose hydrogen atom is oriented toward a specific carbon on the aromatic ring. Thus, the H1−C′2 distance is 3.212 Å, the H3−C′4 distance is 3.087 Å, and the H5−C′6 distance is 3.280 Å. The absence of the dispersion term possibly increases the repulsion, and the monomers are more separated than in the previous complex (A, Figure 2).

The interaction energy, $\Delta E = E_{\text{complex}} - \Sigma E_{\text{monomer}}$, including the standard counterpoise correction and just considering the singlet states was then determined.[19] Table 1 shows the calculated energies for each term required in the counterpoise correction (CP), by the standard method. At the MP2 level, a stabilizing energy of 2.4 kcal/mol was obtained, while at the B3LYP/6-31G(d,p) level, this energy decreased to only 0.1 kcal/mol. The flat curve obtained at this last level of theory (see Figure 3, ●) presents a minimum that is not reliable, since its small value is in borderline with the sensitivity of the method. This result supports that the use of the hybrid functional is not adequate for the study of this type of interactions.

The effect of the molecular geometry on the stability of the supramolecule was then evaluated. Thus, the freezing of the geometry of the individual components along with a monotonic variation of the distance between them permitted to obtain the potential energy associated to the variation of the distance between the carbohydrate and the benzene ring.

Figure 3 shows the profiles of the curves obtained for each case. For the geometry originated by the B3LYP/6-31G(d,p)

(16) (a) Martínez Mayorga, K.; Cortés, F.; Leal, I.; Reyna, V.; Quintana, D.; Antúnez, S.; Cuevas, G. *Arkivoc.* **2003**, *XI*, 132−148. (b) Juaristi, E.; Cuevas, G. *Tetrahedron* **1992**, *48*, 5019−5087. (c) Juaristi, E.; Cuevas, G. *The Anomeric Effect;* CRC press: Boca Raton, FL, 1994.
(17) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(18) (a) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Comput. Chem.* **1996**, *17*, 841−850. (b) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247−3276.
(19) Hernández-Trujillo, J.; Colmenares, F.; Cuevas, G.; Costas, M. *Chem. Phys. Lett.* **1997**, *265*, 503−507.

Table 1. Interaction Energies Corrected by Basis Set Superposition Error (BSSE) at All Levels of Theory[a]

level	*B* (Hartrees)	*C* (Hartrees)	*b* (Hartrees)	*e* (Hartrees)	*f* (Hartrees)	*g* (Hartrees)	Δ*E* (BSSE) (Hartrees//kcal mol^{-1})
A	−841.77738	−841.76746	−610.26325	−231.51024	−610.26198	−231.50538	−0.00379//−2.4
B	−844.23678	−844.23445	−611.97672	−232.25989	−611.97622	−232.25819	−0.00013//−0.1
C	−844.51554	−844.51344	−612.18947	−232.32397	−612.18985	−232.32393	−0.00244//−1.5
D	−841.77365	−841.76666	−610.26196	−231.50762	−610.26135	−231.50534	−0.00410//−2.6

[a] The results for the different levels of theory are shown in rows A−D: A, MP2/6-31G(d,p); B, B3LYP/6-31G(d,p); C, B3LYP/6-311++G(2d,2p); D, MP2/631G(d,p)//B3LYP/ 6-31G(d,p). Columns b−g show the different components for the correction by the standard counterpoise method; *b*, total energy of the complex; *c*, sum of the free molecules energies; *d*, donor massage; *e*, acceptor massage; *f*, correction by donor geometry; *g*, correction by acceptor geometry.

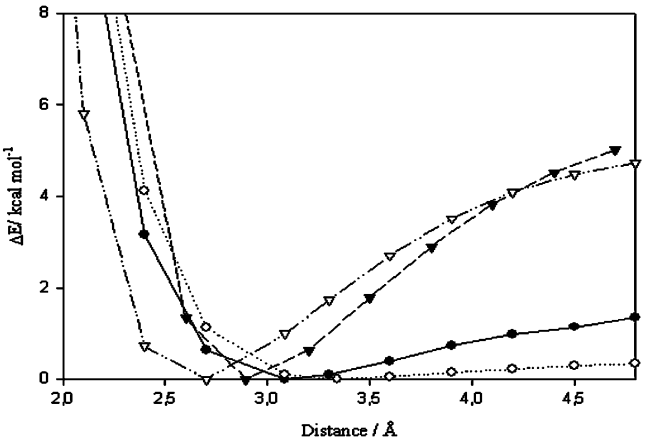

Figure 3. Distance dependence of the interaction energy. ▼ MP2/6-31G-(d,p); ● B3LYP/6-31G(d,p); ○ B3LYP/6-311++G(2d,2p); ▽ MP2/6-31G-(d,p)//B3LYP/6-31G(d,p). The energy differences were estimated with respect to the minimum at the different levels of theory which are given in Table 1.

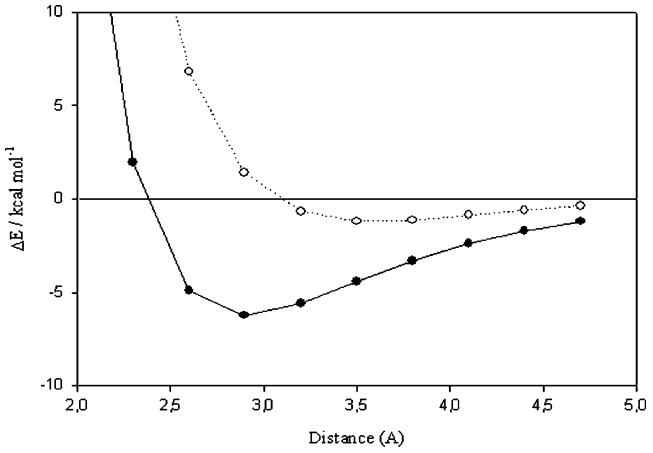

Figure 4. Interaction energies for the fucose−benzene complex at different C−H distances. ○ SCF energy; ● MP2/6-31G(d,p). Interactions energies were calculated with respect to the sum of the energy of the isolated monomers.

method (Figure 3, ●), the distance between Fuc H3 and the C′3 carbon of the benzene ring was systematically modified. In this case, a minimum at an interatomic distance of 3.1 Å was obtained, associated to a very flat curve, which showed very low sensitivity to distance increasement.

The increase of the quality of the basis set for the calculation (using B3LYP/6-311++G(2d,2p) did not improve the description of the system (Figure 3, ○). Indeed, when diffuse functions were included in the B3LYP protocol, the obtained minimum even showed a longer H3−C′3 distance, 3.3 Å. Again, a flat potential energy curve was obtained, in which the interaction energy did not substantially increase for longer interaction distances (less than 0.5 kcal/mol, when the intermolecular distance increased up to 4.3 Å). These observations fully demonstrate that this functional is not adequate for the evaluation of this type of interaction.

The potential energy curve generated at the MP2/6-31G(d,p)// B3LYP/6-31G(d,p) level (Figure 3, ▽), was calculated using the geometry shown in Figure 2B, with an SP calculation. In this case, a minimum was reached for an interaction distance of 2.7 Å, with a relevant increase of the stabilizing interaction energy (∼4 kcal/mol), when the distance was increased by 1.8 Å. The interaction energy determined at this level, with the CP correction, is 2.6 kcal/mol (Table 1D). It is interesting to note the difference (and similarity) with the potential energy obtained using the MP2/6-31G (d,p) method (Figure 3, ▼), for which the minimum is reached for a H3−C′4 interatomic distance of 2.9 Å.

The geometry at B3LYP/6-31G (d,p) level (Figure 2B) is not a minimum at the MP2 level, and when it was taken as the starting point for a full geometry optimization, it converges to that shown in Figure 2A. The values of the energies corrected with the CP method at the four levels of theory are included in Table 1. The highest interaction energy (−2.6 kcal mol^{-1}) is obtained at the MP2/6-31G(d,p)//B3LYP/6-31G(d,p) level (method D, Table 1). This energy is higher than that corresponding to MP2/6-31G(d,p), −2.4 kcal mol^{-1} (method A, Table 1). When the quality of the basis set for the B3LYP calculation is increased (B3LYP6-311++G(2d,2p), method C, Table 1), the corresponding interaction energy falls down to −1.5 kcal mol^{-1}, although is much higher than that corresponding to B3LYP/6-31G(d,p), of −0.1 kcal mol^{-1} (method B, Table 1).

The total interaction energy is composed by two contributions, the "self-consistent field", SCF, (ΔE_{SCF}) energy and the electronic correlation energy (ΔE_{CORR}). The corresponding values ΔE_{SCF} and ΔE_{CORR} are presented in Tables S1− S3 in the Supporting Information. For comparative purposes, the corresponding SCF interaction energies calculated at the MP2/6-31G(d,p) level are presented in Figure 4. As mentioned above, the minimum in the potential energy curve at the MP2 level is located at a 2.9 Å distance between the sugar and the aromatic ring. In the case of SCF component, the minimum occurs at an intermolecular distance, which is indeed very large (3.5 Å). It is also interesting to observe the similarity of the SCF results with those obtained with the B3LYP functional, since the partial electronic correlation is considered by this last method, although in a very rough manner. On the other hand, if the interaction had only an electrostatic origin, the SCF results should have provided a proper description of the complex.

Therefore, the inclusion of the correlation energy, that is, the dispersion interaction between uncorrelated monomers, is a key factor to produce a well-determined minimum in the interaction energy curve. On the other side, the main contribution to the

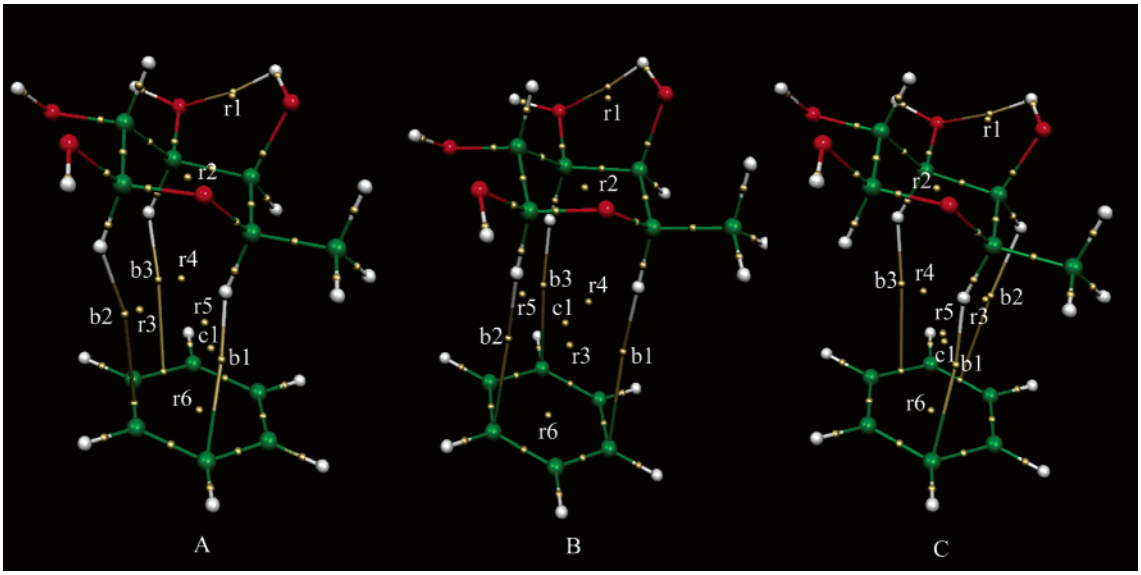

Figure 5. Critical points of the supramolecular arrangements. (A) MP2/6-31G(d,p); (B) B3LYP/6-31G(d,p); (C) MP2/6-31G(d,p) CP-corrected geometry. According to AIM, b notes bond critical points; r, ring critical points; and c, cage critical points.

total energy of the complex comes from the SCF term. (See Scheme S1 in the Supporting Information.)

These results indeed establish the importance of considering the dispersion energy for evaluating the energy for complex formation involved in aromatic−sugar interactions, and also at a more general level, pose some questions regarding the freezing of the molecular geometry during the optimization process. A methodology that allows for the optimization of the molecular geometry, including the CP correction, has been proposed.[13] Figure 2C shows the obtained geometry with this procedure when applied at the MP2/6-31G(d,p) level. Two major changes are observed: the aromatic ring is displaced toward the C4 equatorial position of the sugar, and although the distances between the interaction CH−C centers are increased, the stabilizing interaction energy further increases to −3.0 kcal mol^{-1}. In the obtained geometry (Figure 2C), the H5 hydrogen atom is displaced outside the ring toward C′1 and stays close to the benzene C′6 at approximately 2.9 Å. The H4eq hydrogen atom sits between C′4 and C′5 carbons (3.371 and 3.328 Å, respectively). H3 stays at 3.072 and 3.080 Å, respectively, from C′3 and C′4, while H1 hydrogen is now further apart (at more than 3.54 Å).

Since this structure could correspond to a new stationary state originated by a different approximation of benzene to Fuc, the search for a new stationary state using the MP2 method, without the inclusion of the counterpoise correction during the optimization, was performed again. Under these conditions, the structure of Figure 2A was reached again. This fact demonstrates the fundamental importance of including the counterpoise correction during the optimization process, since it allows reproducing the approximation of a Fuc or Gal-type pyranose ring to an aromatic moiety. In structure 2C, H-3, H-4, and H-5 of the sugar face the benzene aromatic ring, as experimentally observed in the crystal structures of galactose−lectin complexes (Figure 1B).[20]

Complex 2C (Figure 2C) shows one small imaginary frequency, where the benzene ring rotates almost freely in front of the fucose fragment. The potential energy surface is probably very flat and the imaginary frequency could be an artifact result because of numerical errors. Therefore, it is safe to assume that the geometry presented in the figure, and the energy of the structure, is quite similar to the minimum of the potential energy surface.

Since it is not easily visualized whether the interaction between the molecules is manifested by the distance between the atoms involved, the topology of the electronic density was studied, using Bader's topological theory of atoms in molecules (AIM).[14] Three of the stationary states, as shown in Figure 5, were analyzed.

The B3LYP/6-31G(d,p) geometry (B, Figure 5) allows to establish the formation of the three CH−benzene ring bonds that close three eight-membered rings (characterized by bond and ring critical points, respectively, as expected), thus generating a cage critical point. The geometry obtained at the MP2/6-31G(d,p) level (A, Figure 5) shows three bond critical points, somehow distorted: The bond formed by H5 is directed toward the carbon atom (C′1) of the aromatic ring, while H1 and H3 are pointing toward the C−C bonds (C′2−C′3 and C′3−C′4, respectively). The analysis of the electronic density generated by the geometry obtained with the counterpoise correction during the optimization process (C, Figure 5) also shows the presence of bond, ring, and cage critical points. Also, it rigorously establishes that there is no connectivity of 1,3,5-type, as in the previous A and B cases, but that the CH−π interaction gives a 3,4,5-type arrangement. Thus, now H5 hydrogen is directed toward the aromatic ring carbon C′1 while the hydrogen atoms on positions 3 and 4 are directed toward the C−C bonds. This sugar−aromatic geometry is a recurrent arrangement for galactose in many carbohydrate−protein complexes deposited in the Protein Data Bank (PDB).[10b] However, although a complete search was not pursued in the pdb database for protein−galactose complexes,[10a] it was also possible to find the alternate 1,3,5 axial-type interaction in some cases, as for the tetanus toxin (lectin-type) complexed with lactose[20] and for

(20) (a) 1,3,5-type interaction: Emsley, P.; Fotinov, C.; Black, I.; Fairweather, N. F.; Charles, I. G.; Watts, C.; Hewitt, E.; Isaacs, N. W. *J. Biol. Chem.* **2000**, 8889−8894. (b) 3,4,5-type interaction: Elgavish, S.; Shaanan, B. *J. Mol. Biol.* **1998**, *277*, 917−932. For more examples, see refs 8c and 10.

influenza virus hemagglutinin complexed to 2,3-sialyllactose (Figure 1A).[10c]

The properties of the critical points of interest, at the different levels of theory, are included in Table S4 of the Supporting Information.

The results obtained herein indicate that the carbohydrate−aromatic interactions are stabilizing interactions with an important dispersive component and that electronic density indeed exists between the sugar hydrogens and the aromatic ring, thus giving rise to three so-called nonconventional hydrogen bonds.

Experimental NMR Demostration

Carbohydrates do indeed interact with aromatic moieties.[1−4] NMR and X-ray experimental data[1−7] have demonstrated that, in many occasions, when sugars are accommodated in the binding sites of lectins, the aromatic rings within the side chains of Tyr, Phe, and Trp provide key points for interaction with the pyranose rings, especially of the galactose/fucose family,[8c] but also of the glucose one.[8,21] As mentioned above, all these sugars provide a cluster of at least three C−H vectors pointing toward the same spatial direction. Of course, these contacts take place within the context of the corresponding three-dimensional structure of the receptor−ligand complex and, in most of the cases, hydrogen bonds as well as van der Waals contacts with aliphatic side chains are also present in the recognition sites of lectins, which in turn are surrounded by water molecules. Some features of the molecular recognition process of sugars by antibodies have also been explored in the gas phase.[22] Nevertheless, whether these sugar−aromatic interactions may also take place for the isolated components in solution remains an open question. The calculations described above have shown that the fucose−benzene interaction is indeed stabilizing in vacuo, which requires a given geometry to take place in an efficient manner.

In contrast with the experimental evidences for galactose/fucose recognition, the interaction of protein receptors with α-mannose moieties mainly concerns sugar−protein hydrogen bonds or even metallic cations, such as calcium (as in the C-type receptors).[23] The stereochemistry of the regular α-mannopyranosyl rings does not permit the presence of three C−H vectors pointing toward the same orientation, since there are oxygen atoms occupying either the axial orientation at position 1 or the equatorial one at position 4.

Thus, on this basis, we hypothesized on the possibility of detecting specific carbohydrate−aromatic interactions in water, just using a simple sugar, such as methyl β-galactoside (the analogue of the fucose moiety employed for the calculations), and a simple aromatic ring, such as benzene, and using NMR to verify the presence of a complex. Benzene is water-soluble up to 10−20 mM concentration, and thus, two regular monodimensional ^{1}H NMR spectra were recorded for a 1 mM solution of methyl β-galactoside in D_2O alone and in the presence of 10 mM benzene. Interestingly, some changes in the chemical shifts of the sugar protons were observed and, remarkably, the protons on one face of the pyranose ring (H-1, H-3, H-4, H-5, and $H6_{proR}$) shifted high field upon addition of benzene (see Table S6, Supporting Information), while H-2 and the OCH_3 protons, which are located at the opposite face of the six-membered ring, slightly shifted downfield. $H6_{proS}$ did not shift at all. Although the measured changes were small (up to 1 Hz), they were reproducible and could indicate the presence of a specific interaction of benzene approaching the galactose ring by the alpha face (see Figure S1, Supporting Information), thus shielding the cluster provided by H-1, H-3, and H-4 and H-3, H-4, and H-5 hydrogens, according to the geometries described in the calculations section. In contrast, the behavior of the protons at the other side of the ring (H-2 and the OCH_3) was completely different and they moved downfield (see below).

As comparison, the same experiment was performed but using methyl α-mannoside, since, as mentioned above, this sugar cannot provide three C−H vectors pointing toward the same spatial direction. In this case, no shielding of the ring protons at either face was observed. Only the OCH_3 protons moved 0.1 Hz at higher field. Opposite to methyl β-galactoside, H-3, H-4, and H-5 were shifted at lower field up to 0.5 Hz (see Table S7 and Figure S1, in the Supporting Information). This behavior can be explained by a nonspecific interaction of benzene with methyl α-mannoside, with the external current field of the aromatic rings providing a small, but measurable, downfield shift of the ring protons. No sugar−aromatic stacking takes place.

Since it was not possible to increase the amount of benzene dissolved in the NMR tube, the same experiments were performed by adding phenol to a 10 mM methyl β-galactoside solution (or methyl α-mannoside, as control). Thus, ^{1}H NMR spectra were recorded for the corresponding sugar−phenol mixtures, using different concentrations of phenol (up to 200 mM). It is remarkable that the different proton signals kept the behavior observed for the benzene addition, but much larger shifts were observed. The higher concentration of phenol produced the higher changes in the sugar protons (Figure 6). Again, the protons at the alpha face of the galactoside, namely, H-1, H-3, H-4, and H-5, as well as $H6_R$, shifted at high field upon addition of phenol (Table 2), while the protons at the beta face of the ring, H-2, $H6_S$, and OCH_3, slightly shifted downfield.

On the other hand, the addition of 200 mM phenol to the 10 mM solution of methyl α-mannoside produced rather distinct results, indicating that no specific sugar−aromatic stacking takes place for phenol and this sugar. Only H-1 and the OCH_3 protons moved at higher field, while H-2, H-3, H-4, H-5, and both H-6 were shifted at lower field (Table S7, Figure S2, Supporting Information).

As further control, the addition of 200 mM phenol to a 10 mM 1:1 mixture of both sugars produced the same results as for the individual sugars (see Figure S3, Supporting Information). Upfield shifting of the key protons of the galactose moiety was observed, together with the simultaneous distinct behavior of those of the mannose monosaccharide.

The slight variation of pH produced by the addition of phenol or nonspecific interactions with phenol may account for the observed downfield shielding of the mannoside protons. In contrast, the observed shielding of the protons at the alpha face of the galactoside moiety is in agreement with a specific interaction of the aromatic ring of phenol (and benzene) with

(21) Asensio, J. L.; Cañada, F. J.; Siebert, H.-C.; Laynez, J.; Poveda, A.; Nieto, P.; Soedjanaamadja, U. M.; Gabius, H. J.; Jimenez-Barbero, J. *Chem. Biol.* **2000**, *7*, 529−543.
(22) (a) Kitova, E. N.; Bundle, D. R.; Klassen, J. S. *J. Am. Chem. Soc.* **2002**, *124*, 5902−5913. (b) Kitova E. N.; Bundle D. R.; Klassen J. S. *J. Am. Chem. Soc.* **2002**, *124*, 9340−9341. (c) Kitova, E. N.; Wang, W.; Bundle, D. R.; Klassen J. S. *J. Am. Chem. Soc.* **2002**, *124*, 13980−13981.
(23) McGreal, E. P.; Miller J. L.; Gordon, S. *Curr. Opin. Immunol.* **2005**, *17*, 18−24.

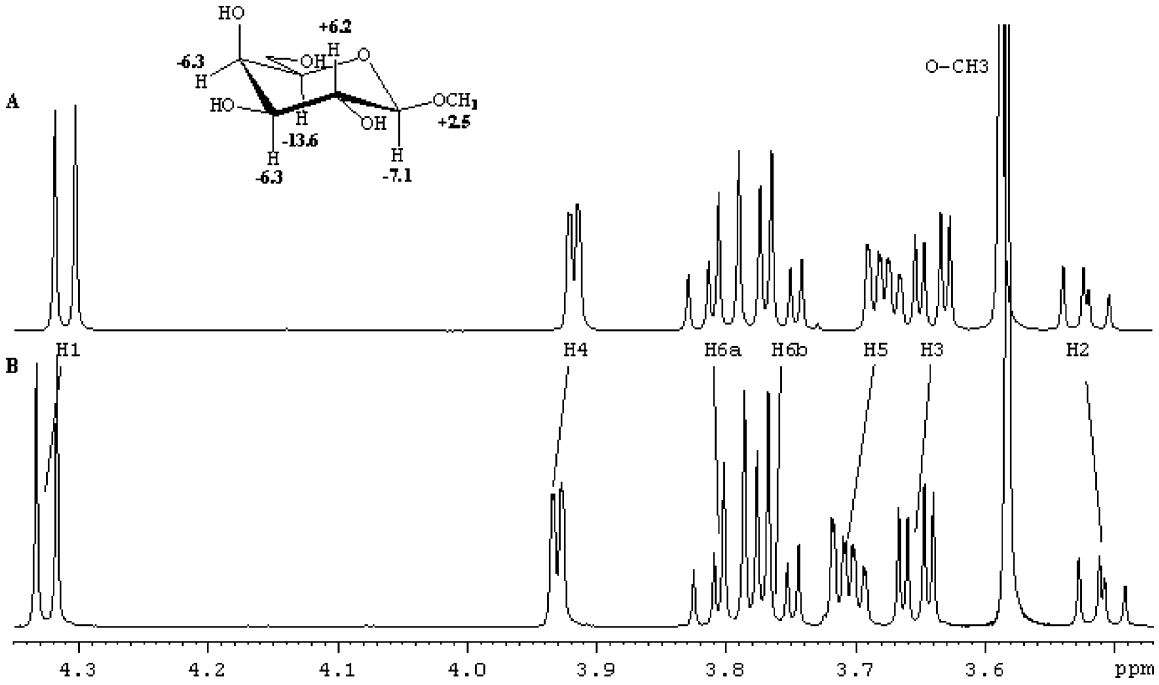

Figure 6. 500 MHz ^{1}H NMR spectra for methyl β-galactoside in D_2O at 298 K (B) and upon addition of 200 mM of phenol (A). The shifting of the protons at the bottom face of the pyranoid ring (see inset) is remarkable. TSP (0 ppm) was used as external standard.

Table 2. Chemical Shifts (δ, ppm) for Methyl β-Galactoside in D_2O Solution at 298 K and 500 MHz [a]

β-MeGal	δ (ppm)	$\Delta\delta$ (Hz) phenol 0.1 M	$\Delta\delta$ (Hz) phenol 0.2 M
O-CH3	3,583	1,1	2,5
H-1	4,325	−4.0	−7,1
H-2	3,509	3.0	6,2
H-3	3,654	−3,7	−6,3
H-4	3,93	−3,6	−6,3
H-5	3,705	−7,3	−13,6
H-6_S	3,806	0,9	2,1
H-6_R	3,76	−0,8	−1,1

[a] The observed variations (Hz) upon addition of two different phenol concentrations are also given. The shielding of H-1, H-3, H-4, and H-5 upon addition of the aromatic ring is evident. TSP (0 ppm) was used as external standard.

the C−H vectors at this face. (Figure S4, Supporting Information). Thus, these experiments provide evidence that support the existence of a specific interaction of aromatic rings with certain sugars, depending on the orientation of the hydroxyl groups and therefore on the proper clustering of at least three C−H vectors. It seems that there is an intrinsic tendency of aromatic moieties to provide stabilizing interactions with clusters of C−H vectors in sugars and that these interactions may take place without the requirement of additional protein−sugar hydrogen bonds, as calculated in the previous section and as frequently observed in carbohydrate-binding proteins. Of course, dealing with individual sugars and aromatics, the interaction is weak but definitively measurable with a binding constant in the vicinity of 10 M^{-1}, as determined by qualitative titration NMR, that corresponds to a free energy of binding of ca. 1.5 kcal/mol.

Conclusions

This study has unequivocally shown the existence of stabilizing carbohydrate−aromatic interactions from both the theoretical and experimental viewpoints. From the theoretical side, the importance of including the counterpoise correction in the optimization process has been shown. At least an MP2 level of theory should be used to properly evaluate the intermolecular dispersion energy. Under these conditions, the geometry of experimentally based galactose−lectin complexes is properly accounted for. In this case, the interaction energy of the fucose−benzene is 3.0 kcal/mol, somehow higher than that established for single carbohydrate−aromatic interactions, for instance, in the hevein/chitooligosaccharide complexes (ca. 1.5 kcal/mol).[21] However, entropy factors and the role of water may influence the actual value for the aromatic−sugar interaction. Nevertheless, the calculated energy shows it is on the same order of magnitude. These numbers show the efficiency of the CH/π interaction in the stabilization process of the sugar−aromatic supramolecule. Since three CH/π interactions participate in the formation of the complex, it can be estimated that the stabilization produced by each interaction is ∼1 kcal/mol. Moreover, the theoretical results obtained herein indicate that the carbohydrate−aromatic interactions are stabilizing interactions with an important dispersive component and that electronic density indeed exists between the sugar hydrogens and the aromatic ring, thus giving rise to three so-called nonconventional hydrogen bonds.

Finally, experimental evidence of the intrinsic tendency of aromatic moieties to interact with certain sugars has also been shown by simple NMR experiments in water solution. Benzene and phenol specifically interact with the clusters of C−H bonds of the alpha face of methyl β-galactoside, without requiring the well-defined three-dimensional shape provided by a protein receptor, thus resembling the molecular recognition features that are frequently observed in many carbohydrate−protein complexes.

Acknowledgment. This paper is dedicated to Prof. Dr. José Barluenga on the occasion of his 65th birthday. We are grateful to Prof. J. Hernández Trujillo for useful comments. We are grateful to the Dirección General de Servicios de Cómputo

Académico, Universidad Nacional Autónoma de México DGSCA, UNAM, to Consejo Nacional de Ciencia y Tecnología (CONACYT) for financial support via grant 40390-Q. We thank Ministerio de Educación y Ciencia of Spain for funding (BQU2003-C03550-03-01) and the Centro de Supercomputación de Galicia (CESGA) for supercomputing time. M.C.F.A. also thanks an FPU fellowship. The cooperation program between CSIC and UNAM is also acknowledged.

Supporting Information Available: Geometric coordinates of the different arrangements of the complex at the different levels of theory. Table S1 with the MP2/6-31G(d,p) energy. Table S2 with the SCF component of the MP2/6-31G(d,p) energy. Table S3 with the component of correlation of the MP2/6-31G(d,p) energy. Table S4 with the properties of the most important critical points for the complexes. Table S5 with the geometric parameters of the complex at the different levels of theory. Scheme S1 with the components of the total interaction energy at the MP2/6-31G(d,p) level. Table S6 with the NMR parameters for free methyl β-galactoside and in the presence of benzene and phenol. Table S7 with the NMR parameters for free methyl α-mannoside and in the presence of benzene and phenol. Figures S1−S4 with the ^{1}H NMR spectrum for free methyl α-mannoside and methyl β-galactoside, its mixture, and in the presence of benzene and phenol. Complete ref 9a and 9b. This material is available free of charge via the Internet at http://pubs.acs.org.

JA051020+