Insights into the Role of the Aromatic Residue in Galactose-Binding Sites: MP2/6-311G++** Study on Galactose− and Glucose−Aromatic Residue Analogue Complexes†

Mannargudi S. Sujatha,‡ Yellamraju U. Sasidhar,‡,§ and Petety V. Balaji*,‡

School of Biosciences and Bioengineering and Department of Chemistry, Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India

Received February 18, 2005; Revised Manuscript Received April 21, 2005

ABSTRACT: The presence of an aromatic residue (Trp, Phe, Tyr) facing the nonpolar face of galactose is a common feature of galactose-specific lectins. The interactions such as those between the C−H groups of galactose and the π-electron cloud of aromatic residues have been characterized as weak hydrogen bonds between soft acids and soft bases, largely governed by dispersive and charge transfer interactions. An analysis of the binding sites of several galactose-specific lectins revealed that the spatial position−orientation of galactose relative to the binding site aromatic residue varies substantially. The effect of variations in position−orientations of galactose on the interaction energies of galactose−aromatic residue complexes has not been determined so far. In view of this, MP2/6-311G++** calculations were performed on galactose− and glucose−aromatic residue analogue complexes in eight position−orientations. The results show that the strength of the C−H···π interactions in galactose−aromatic residue complexes is comparable to that of a hydrogen bond. Rather than the type of aromatic residue, the position−orientation of the saccharide appears to be more critical in determining the strength of their interactions. Earlier studies have found the binding site aromatic residue to be critical, but its role was not clear. This study shows that the aromatic residue is important for discriminating galactose from glucose, in addition to its contribution to binding energy.

The presence of an aromatic residue (Trp, Phe, Tyr) facing the nonpolar face of galactose (Gal) is a common feature of Gal-specific lectins (*1*−*9*). Several experimental studies have underscored the critical role played by the aromatic residue in galactose-binding sites. While reengineering the saccharide specificity of mannose-binding protein (MBP-A), it was observed that the introduction of a tryptophan substantially increased the affinity for galactose (*10*). In a few other proteins, mutation of the aromatic residue led to a significant decrease/loss of affinity for galactose (*11*−*19*). A comparison of the binding sites of the two Gal-specific proteins, reengineered MBP-A (PDB code 1AFA) and tunicate C-type lectin (1TLG), shows that the binding sites of the two proteins can be superposed on each other, with the exception of a tryptophan (Figure 1). However, the altered position of tryptophan leads to a "flipping" of galactose in the binding site, resulting in differences in the interactions of galactose with the rest of the binding site.

The interactions of galactose with the aromatic residue are mediated by the nonpolar C−H atoms. The interactions between the nonpolar C−H groups and the π-electron cloud of aromatic residues have been characterized as weak hydrogen bonds between soft acids and soft bases, largely governed by dispersive and charge transfer interactions (*20*, *21*). Recently, the interaction energies of glucose− and galactose−3-methylindole complexes were found to be −5.2 and −2.7 kcal/mol, respectively, from MP2 calculations performed at the 6-31+G(d) level; the position−orientations of the saccharide relative to the aromatic residue used in this study were the same as those observed in the binding site of β-galactosidase from *Escherichia coli* (*9*).

A statistical analysis of 78 nonhomologous, pyranoside-binding proteins (excluding sialic acids and phosphorylated saccharides) showed that the pyranoses tend to bind near the aromatic ring centers; however, neither a preferential saccharide orientation nor a preferential interaction of the pyranose with the polar side chains of tryptophan and tyrosine was found in this study (*4*). An analysis of the binding sites of several Gal-specific lectins also revealed that the spatial position−orientation of galactose relative to the binding site aromatic residue varies substantially (*22*). Various experimental studies have shown that the binding site aromatic residue is critical, but its role in sugar binding is not clear. In view of this, the present study was undertaken to determine the interaction energies of galactose− and glucose (Glc)−aromatic residue analogue complexes using ab initio quantum chemical methods. The interaction energies were calculated in eight position−orientations observed in various saccharide−protein complexes (Table 1). The posi-

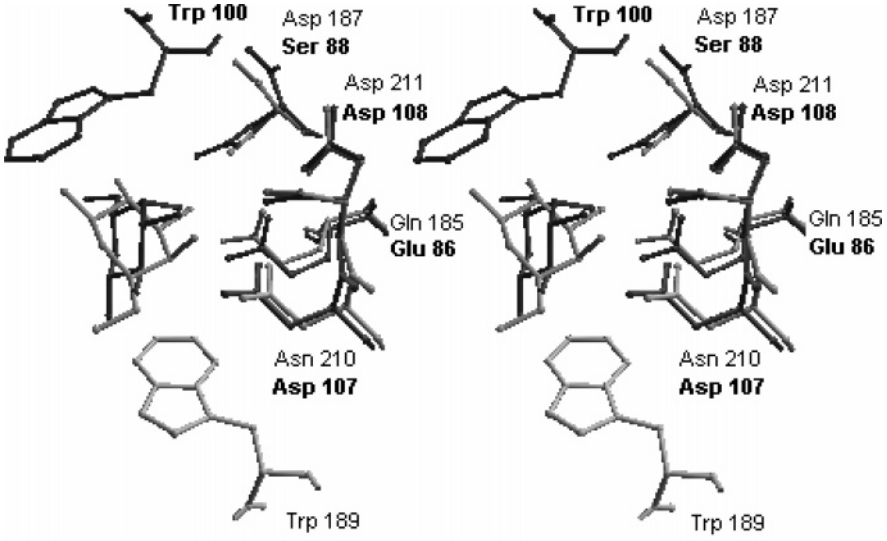
† Supported by the Council for Scientific and Industrial Research, India (to P.V.B.), Grant 37(1110)/02/EMR-II. M.S.S. is grateful to the Indian Institute of Technology Bombay for a teaching assistantship.

* To whom correspondence should be addressed: balaji@iitb.ac.in (e-mail), +91-22-25 76 77 78 (tel), +91-22-25 72 34 80 (fax).

‡ School of Biosciences and Bioengineering, Indian Institute of Technology Bombay.

§ Department of Chemistry, Indian Institute of Technology Bombay.

10.1021/bi050298b CCC: $30.25 © 2005 American Chemical Society
Published on Web 05/19/2005



1AFA ^{181}WKKD**Q**P**D**D**W**YGHGLGGG**E**DCVTIVDNGLW**NDI**^{212}
1TLG ^{82}WSPN**E**P**S**NPQ----SWQLCVQI**W**SKYNLL**DDV**^{109}

Gal	1AFA	1TLG
C2-OH	N210	-
C3-OH	E198	E86
C4-OH	D187	D107
Ring O	HOH	HOH
C6-OH	HOH	HOH
Stack	W189	W100

FIGURE 1: Superposition of the binding site residues of the reengineered MBP-A (PDB ID 1AFA; shown in black) over those of tunicate C-type lectin (1TLG; shown in gray). Both of the proteins are members of the C-type animal lectin family and are specific to galactose. The superposition was done with respect to Asp211 (in 1AFA) and Asp108 (in 1TLG). The binding sites of the two proteins are superposable except for the aromatic residue (Trp189 in 1AFA and Trp100 in 1TLG). The two proteins have no sequence similarity, and only a short stretch, containing the binding site residues, can be aligned by using a very high *E*-value of 100000. The residues, which interact with galactose, are shown in bold in the sequence alignment. The residues aligned in sequence superpose on each other: Gln185-Glu86, Asp187-Ser88, Asn210-Asp107, and Asp211-Asp108 (residues of 1AFA are given first followed by those in 1TLG). However, the sequentially and spatially equivalent residues do not have the same interactions with the bound galactose because of the flipping of the saccharide in the binding site.

Table 1: BSSE Corrected Interaction Energies for the Saccharide−Aromatic Residue Complexes

PDB ID[a]	aromatic residue	bound sugar	sugar specificity	−CH_2OH group conformation	position (r,θ,ϕ)[b]	orientation (Φ,Θ,Ψ)[b]	interaction energy (kcal/mol)[c] Gal	Glc-I	Glc-II	Glc-III
1SLT	Trp68	galactose	galactose	*gt*	5.1,150,−3	19,64,209	−3.2	23.8[d]	22.4[d]	24.9[d]
1HLC	Trp65	galactose	galactose	*gt*	4.9,143,5	37,48,199	−3.6	19.0[d]	19.2[d]	19.6[d]
2AAI	Trp37	galactose	galactose	*gg*	5.3,38,−17	61,113,12	−3.8	62.3[d]	40.5[d]	49.3[d]
1GAN	Trp69	galactose	galactose	*gt*	4.7,152,−15	29,50,191	−5.7	7.0[d]	7.9[d]	8.5[d]
7CEL	Trp376	glucose	glucose	*gg*	4.2,22,88	225,137,333	−6.6[d]	−8.1	−5.4	−5.2
1GCA	Trp183	galactose	galactose, glucose	*gt*	4.4,162,45	191,41,178	−7.5	−8.2	−4.6	−4.9
1AX1	Phe131	galactose	galactose	*gt*	4.1,8,−102	302,137,5	−4.2	0.1[d]	2.7[d]	3.3[d]
1BZW	Tyr125	galactose	galactose	*gt*	4.3,8,−51	316,122,13	−4.7	17.4[d]	18.9[d]	18.3[d]

[a] From the Protein Data Bank (*52*). The names of the protein and the resolution at which the structure were determined are as follows: 1SLT, S-lectin (1.9 Å); 1HLC, S-lac lectin (2.9 Å); 2AAI, ricin B chain (2.5 Å); 1GAN, toad ovary galectin (2.23 Å); 7CEL, β1→4-glucan cellobiohydrolase (1.9 Å); 1GCA, periplasmic glucose/galactose receptor (1.7 Å); 1AX1, *Erythrina corallodendron* lectin (1.95 Å); and 1BZW, peanut lectin (2.7 Å). [b] The bound galactose can be at different positions relative to the binding site aromatic residue; at any position, the saccharide can be in different orientations. The position of the bound sugar relative to the binding site aromatic residue is specified by polar coordinates of the centroid of the pyranose ring. The orientation of the bound sugar relative to the binding site aromatic residue is specified in terms of the Euler rigid body rotation angles (*25*). Both the position and orientation have been defined in a frame of reference defined within the aromatic residue (*22*). [c] The interaction energy is for the complex of the specified sugar with the aromatic residue analogue. The BSSE uncorrected interaction energy values for galactose and glucose are given in Table S1. [d] These are hypothetical complexes, i.e., complexes of glucose in position−orientations observed in galactose-specific proteins and vice versa.

tion−orientations have been selected from six galactose-specific proteins, one glucose-specific protein, and one protein which can bind to either galactose or glucose. The complexes of glucose in position−orientations observed in galactose-specific proteins are hypothetical complexes since glucose does not bind to these proteins; similarly, the complex of galactose in the position−orientation observed in glucose-specific protein is also a hypothetical complex (Table 1). The results show that the interaction energies of saccharide−aromatic residue analogue complexes are com-

parable to that of a hydrogen bond. It is also found that the aromatic residue plays a role in discriminating galactose from glucose in galactose-specific proteins.

METHODS

Software and Hardware. All calculations were performed using the software GAMESS (*23*). Geometry optimizations were performed on either an Intel Xeon 2.4 GHz dual processor system or a 16 node P4-1.6 GHz Linux cluster or a 34 dual-processor node Intel Xeon 3.2 GHz Linux cluster. MP2 calculations were performed on PARAM Padma, a parallel supercomputer designed and developed by the Center for Development of Advanced Computing, India's national initiative in high-performance computing (*24*).

Reason for Choosing the Analogues of Aromatic Residues for Interaction Energy Calculations. The main chain atoms of the aromatic residue and the galactose-binding pocket occur on different sides of the aromatic ring (*22*). In view of this, it is assumed that the main chain atoms make little or no contribution to the interaction energy and the side chain analogues of the aromatic residues, viz., 3-methylindole (for tryptophan), toluene (for phenylalanine), and *p*-hydroxytoluene (for tyrosine), were considered for the interaction energy calculations.

Choice of Position−Orientation and Saccharide Conformation. An infinite number of positions and, at each such position, an infinite number of orientations of the saccharide relative to the aromatic residue are theoretically conceivable. However, it is practically not possible to calculate the interaction energies in all such position−orientations. Hence, only the position−orientations observed in eight saccharide−protein complexes were used for interaction energy calculations in the present study. The β-anomer of the saccharide with its pyranose ring in the $^4C_1$ conformation was considered. The exocyclic $-CH_2OH$ group was considered to be in the same conformation as observed in the corresponding protein−saccharide complex.

Galactose interacts with the aromatic residue either through the set of atoms C3−H, C4−H, C5−H, and C6−H or through the set of atoms C1−H, C3−H, and C5−H. The positions of these methine hydrogen atoms are fixed for the $^4C_1$ conformation of the pyranose ring. The conformations of the hydroxyl groups of galactose were set arbitrarily. However, in the case of glucose, the equatorial orientation of the hydroxyl group at C4 causes this hydroxyl group to be close to the aromatic residue in position−orientations observed for galactose in Gal-specific proteins. Hence, three different conformations of the hydroxyl hydrogen atoms (Figure 2) were considered for calculating the interaction energies of the complexes of glucose.

Specifying the Position−Orientation of Saccharide Relative to the Aromatic Residue. The position of the saccharide was specified by the polar coordinates. At any given position, the orientation of the saccharide was specified by the Euler rigid body rotation angles (*25*). Both the position and orientation were specified in a frame of reference defined within the aromatic residue as described previously (*22*). The stereoviews of the 8 galactose− and 24 glucose−aromatic analogue complexes used for single point energy calculations are shown in Figure S1.

Geometry Optimization and Interaction Energy Calculation. The geometry optimized at the UHF/6-31G** basis set

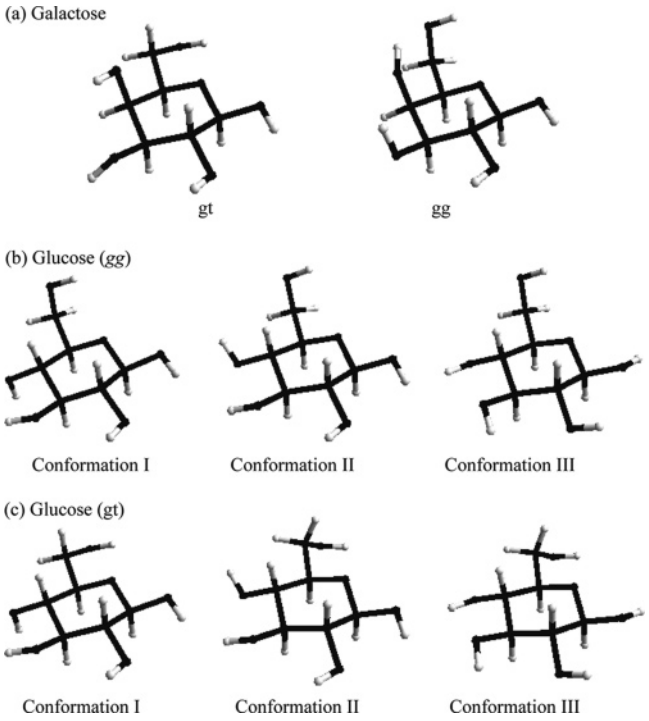


FIGURE 2: Ball-and-stick representation showing the geometry-optimized (RHF/6-311G++** level) conformations of galactose (top row) and glucose (middle and lower rows). The conformation of the $-CH_2OH$ group is specified as *gt* or *gg*. In the *gg* conformation, the $-CH_2OH$ group is gauche to both the ring oxygen and C4 atoms. In the *gt* conformation, the $-CH_2OH$ group is gauche to the ring oxygen atom and trans to the C4 atom.

level (*26*) was taken as the starting point for further optimizations at higher basis set levels i.e., RHF/6-311G** followed by RHF/6-311G++**. The HONDO/Rys integral calculation routine was used to achieve faster convergence. The convergence criterion used for optimization at the 6-311G++** level is that the largest component of the energy gradient should be less than 5×10^{-6} Hartree/Bohr. The geometry obtained from optimization at the 6-311G++** basis set level was used to generate saccharide−aromatic residue analogue complexes with defined position−orientations and to perform single point energy calculations at the MP2/6-311G++** level. The interaction energy was calculated as $E_{A-B} - (E_{A,gB} + E_{B,gA})$, where E_{A-B}, $E_{A,gB}$, and $E_{B,gA}$ are energies of the saccharide (molecule A)−aromatic residue analogue (molecule B) complex, of the saccharide in the presence of the ghost aromatic analogue, and of the aromatic analogue in the presence of the ghost saccharide, respectively. The ghost molecule is generated by assigning zero nuclear charge to all of its atoms.

Galactose was optimized by setting the $-CH_2OH$ group to either the *gt* or *gg* conformation; the *tg* conformation was not considered since it is not observed in any of the eight protein−saccharide complexes considered in this study (Table 1). Glucose also was separately optimized with the $-CH_2OH$ group in either the *gt* or *gg* conformation.

The protein structures under consideration have been determined at different resolutions (see footnote *a* to Table 1), because of which the internal geometries of the aromatic residue and of the bound saccharide are not same and are also not optimal. In view of this, the aromatic residue analogue and the saccharide have been separately optimized, and these optimized structures have been taken for single

Table 2: Interaction Energies Calculated at the MP2/6-31+G* Level for Saccharide−Aromatic Residue Complexes before and after Optimization

PDB ID	complex	no. of steps[a]	interaction energy (kcal/mol)[b]		RMSD (Å)[c]		
			initial	final	saccharide	3-methylindole	complex
			Optimization at MP2/6-31+G*				
1GAN	galactose/3-methylindole	42	−4.8	−6.1	0.05	0.05	0.06
			Optimization at B3LYP/6-31+G*				
1GAN	galactose/3-methylindole	57	−4.8	−6.0	0.06	0.07	0.10
7CEL	glucose/3-methylindole	63	−7.1	−8.3	0.07	0.05	0.10
2AAI	glucose/3-methylindole	53	66.4	−4.5	0.10[d]	0.10[d]	1.71[d]

[a] The total number of steps in the optimization run. [b] The interaction energy is for the complex of the specified sugar with the aromatic residue analogue. The BSSE corrected interaction energy values are reported. The initial and final interaction energies were calculated at MP2/6-31+G* for the initial and the optimized structures. [c] Root mean square deviation between the initial and the final optimized structures. The structures were superposed using the atoms of the aromatic residue analogue as reference atoms for calculating the RMSD for the complex. [d] The small RMSD values indicate that the geometries of the saccharide and the aromatic residue before and after optimization are very nearly the same. The large RMSD for the complex (1.7 Å) is primarily because of the change in the position−orientation of the saccharide relative to the aromatic residue.

point energy calculations at the MP2/6-311++G** level of theory. This will give the interaction energy of the complex for the specified position−orientation and specified conformation of the $-CH_2OH$ group.

Optimization of the Saccharide−Aromatic Analogue Complex. The individually optimized saccharide and the aromatic residue analogues were brought together in the desired position−orientation to form a complex. Optimization of such a complex was performed at either the MP2/6-31+G* level or the B3LYP/6-31+G* level of theory. HONDO/Rys polynomial code was used for calculating the integrals. The linear dependence threshold defined by QMTTOL in GAMESS was set to 1.0×10^{-6} (for B3LYP) or 3.0×10^{-6} (for MP2). The OPTTOL parameter, which specifies the termination criterion for optimization, was set to 5×10^{-4} (for B3LYP) or 6.5×10^{-4} (for MP2). Single point energy of the minimized complex at the MP2/6-31+G* level of theory was used to calculate the interaction energy between the saccharide and aromatic residue.

RESULTS AND DISCUSSION

Optimization of the Saccharide−Aromatic Residue Complex at the B3LYP/6-31+G Level of Theory.* Optimization of the galactose−3-methylindole complex in 1GAN position−orientation and of the glucose−3-methylindole complex in 7CEL position−orientation at the B3LYP/6-31+G* level of theory resulted in a decrease of the interaction energy by 1.2 kcal/mol (Table 2). The RMS deviation between the initial and optimized structures is very small, indicating the absence of any significant changes in the geometry/relative position−orientation due to optimization (Table 2; Figure 3A,B). The galactose−3-methylindole complex in 1GAN position−orientation was also optimized at the MP2/6-31+G* level of theory, and the optimized structure and the interaction energies are very similar to those obtained by optimizing at the B3LYP/6-31+G* level of theory (Table 2).

The interaction energies of various saccharide−aromatic residue analogue complexes in position−orientations observed in β-galactosidase−saccharide complexes have been calculated at different levels of theory (*9*). In this study also it was observed that optimization did not result in any significant changes in the geometry of the interacting moieties; optimization was performed at the HF/MINI level with a less stringent termination criterion than has been used

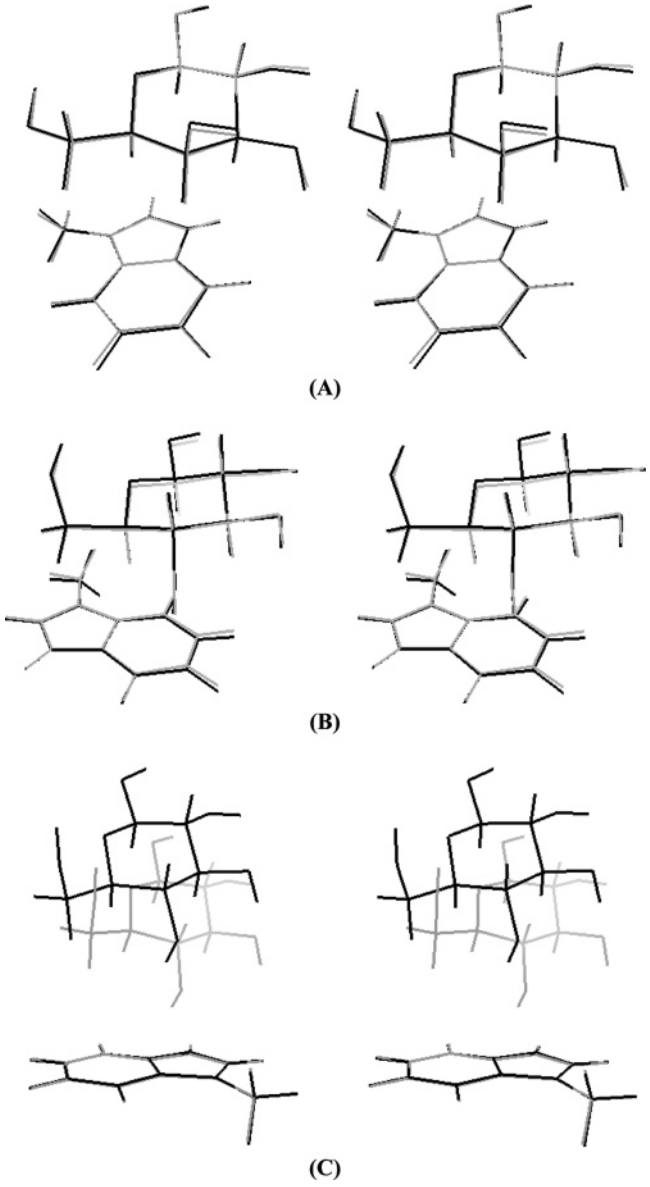


FIGURE 3: Stereoviews of the superposed saccharide−aromatic residue analogue complexes before (shown in gray) and after (shown in black) optimization at the MP2/6-31+G* level of theory: (A) galactose−3-methylindole complex in 1GAN position−orientation; (B) glucose−3-methylindole complex in 7CEL position−orientation; (C) glucose−3-methylindole complex in 2AAI position−orientation.

in the present study. It was also noted by these authors that the changes in the absolute interaction energies of the saccharide−3-methylindole complexes due to optimization are not dramatic (*9*).

Optimization was also performed for the hypothetical complex of glucose−3-methylindole in 2AAI position−orientation (Table 1). The 2AAI position−orientation is observed for galactose in the binding site of ricin B, a galactose-specific protein. In this complex, glucose is placed in the same position−orientation as that found for galactose. The single point energy calculated at the MP2/6-31+G* level of theory for the initial structure used for optimization is 66 kcal/mol. This unrealistically high interaction energy is due to the close proximity of the C4−hydroxyl group to the aromatic ring (Table 3). The repulsion is relieved by buckling of 3-methylindole and flattening of the pyranose ring near C4 during the initial stages of optimization (Figure S2). On further optimization, the saccharide moves away from 3-methylindole with the concomitant restoration of the geometries of the two molecules (Figure 3C; Table 2). The interaction energy is stabilizing in the new position−orientation obtained after optimization. These results, however, are NOT unexpected since this is a hypothetical complex. In principle, repulsion can be relieved either by retaining the position−orientation and adjusting the conformations of the interacting hydroxyl groups or by a change in the relative position−orientation. During the optimization of the glucose−3-methylindole complex in 2AAI position−orientation, the saccharide changes its position and orientation relative to the aromatic residue (Figure 3C).

For the galactose−3-methylindole complex in 1GAN position−orientation and for the glucose−3-methylindole complex in 7CEL position−orientation, the interaction energies calculated for the initial complex at the MP2/6-311++G** level (Table 1) are comparable to that for the optimized complex calculated at the MP2/6-31+G* level (Table 2). As noted earlier, optimization of the complex did not result in any significant changes in the geometry of the molecules; only the position−orientation changed in the complex, which initially had repulsion. Optimization of the entire saccharide−aromatic residue analogue complex at either the B3LYP or MP2 level of theory is computationally very demanding. In view of these observations, the other complexes were not optimized; instead, single point energy calculations were performed at the MP2/6-311++G** level of theory.

*The Single Point Interaction Energy at the MP2/6-311++G** Level of Theory for the Gal−Aromatic Residue Analogue Complex Is Comparable to That of a Hydrogen Bond.* The Gal−aromatic residue analogue interactions, mediated by multiple C−H groups, are negative and stabilizing in all of the eight position−orientations. The interaction energies range from −3.2 to −7.5 kcal/mol (Table 1) and are thus comparable to the energy of a hydrogen bond (*27*−*32*). The variations in the interaction energies of the various complexes can be attributed to the differences in the number and geometry (distance, angle, and dihedral angle; defined in the footnotes to Table 4) of the interactions. The interaction energies seem to correlate well with the average deviations from D_{plane}, θ, and ω individually (Table 4). However, because of the participation of multiple C−H groups in the interactions, the interaction energy relates to

Table 3: Intermolecular Atomic Distances (Å) in Aromatic Analogue−Glucose Complexes[a]

conformation I		conformation II		conformation III	
1SLT Position−Orientation (Complex with 3-Methylindole)					
Cε2···O4	2.67	Cε2···O4	2.56	Cε2···O4	2.58
Cζ2···O4	2.19	Cζ2···O4	2.11	Cζ2···O4	2.17
Cη2···O4	2.41	Cη2···O4	2.43	Cη2···O4	2.43
Cδ2···HO4	2.43	Hζ2···O4	2.42	Cζ2···HO4	1.85
Cε2···HO4	1.81	Cζ2···HO4	2.32	Cη2···HO4	2.18
Cζ2···HO4	1.76	Cη2···HO4	2.06		
Cη2···HO4	2.31				
1HLC Position−Orientation (Complex with 3-Methylindole)					
Cζ2···O4	2.51	Cζ2···O4	2.37	Cζ2···O4	2.41
Cη2···O4	2.30	Cη2···O4	2.23	Cη2···O4	2.24
Cζ3···O4	2.96	Cζ3···O4	2.95	Cζ3···O4	2.90
Hη2···O4	2.44	Hη2···O4	2.40	Hη2···O4	2.45
Cε2···HO4	2.34	Cη2···HO4	2.36	Cη2···HO4	2.25
Cζ2···HO4	1.65				
Cη2···HO4	1.66				
Cζ3···HO4	2.39				
2AAI Position−Orientation (Complex with 3-Methylindole)					
Cδ2···O4	2.62	Cδ2···O4	2.64	Cδ2···O4	2.63
Cε2···O4	2.16	Cε2···O4	2.19	Cε2···O4	2.12
Cζ2···O4	2.94	Cζ2···O4	2.95	Cζ2···O4	2.85
Nε1···O4	2.03	Nε1···O4	2.06	Nε1···O4	2.01
Cδ1···O4	2.43	Cδ1···O4	2.45	Cδ1···O4	2.48
Cγ···O4	2.77	Cγ···O4	2.79	Cγ···O4	2.84
Hε1···O4	2.44	Hε1···O4	2.47	Hε1···O4	2.40
Cε2···HO4	1.76	Cδ2···HO4	2.09	Cε2···HO4	2.38
Cζ2···HO4	2.41	Cε2···HO4	2.12	Cζ2···H62	2.49
Nε1···HO4	1.47	Nε1···HO4	2.42	Cη2···H62	2.49
Cδ1···HO4	2.38	Cγ···HO4	2.42	Nε1···HO4	1.91
Hε1···HO4	1.66			Cδ1···HO4	1.84
				Cγ···HO4	2.29
1GAN Position−Orientation (Complex with 3-Methylindole)					
Cε2···O4	2.94	Cε2···O4	2.81	Cε2···O4	2.83
Cζ2···O4	2.48	Cζ2···O4	2.37	Cζ2···O4	2.42
Cη2···O4	2.58	Cη2···O4	2.54	Cη2···O4	2.55
Cε2···HO4	2.01	Cη2···HO4	2.47	Cζ2···HO4	2.36
Cζ2···HO4	1.75			Cη2···HO4	2.48
Cη2···HO4	2.09				
7CEL Position−Orientation (Complex with 3-Methylindole)					
none		none		none	
1GCA Position−Orientation (Complex with 3-Methylindole)					
Cδ2···HO4	2.47	Cγ···O4	2.96	none	
Cδ1···HO4	2.41				
Cγ···HO4	2.32				
1AX1 Position−Orientation (Complex with Toluene)					
Cδ2···O4	2.96	Cδ2···O4	2.90	Cδ2···O4	2.87
Cγ···O4	2.72	Cγ···O4	2.62	Cγ···O4	2.66
Cδ1···O4	2.90	Cδ1···O4	2.76	Cδ1···O4	2.83
Cε1···HO4	2.47				
Cε2···HO4	2.48				
Cδ2···HO4	2.20				
Cγ···HO4	2.06				
Cδ1···HO4	2.20				
1BZW Position−Orientation (Complex with *p*-Hydroxytoluene)					
Cε1···O4	2.82	Cε1···O4	2.68	Cε1···O4	2.75
Cε2···O4	2.91	Cε2···O4	2.85	Cε2···O4	2.80
Cδ2···O4	2.59	Cδ2···O4	2.57	Cδ2···O4	2.53
Cγ···O4	2.37	Cγ···O4	2.35	Cγ···O4	2.37
Cδ1···O4	2.50	Cδ1···O4	2.39	Cδ1···O4	2.47
Cε1···HO4	2.04	Cη2···O4	2.90	Cη2···O4	2.90
Cε2···HO4	2.10	Cδ2···HO4	2.33	Cγ···HO4	2.10
Cδ2···HO4	2.09	Cγ···HO4	2.24	Cδ1···HO4	2.08
Cγ···HO4	2.07				
Cδ1···HO4	2.03				
Cζ···HO4	2.08				

[a] Only those atom pairs that are within the following distance cutoffs are listed here: both are non-hydrogen atoms, ≤3 Å; one of the atoms is hydrogen, ≤2.5 Å; and both atoms are hydrogens, 2.0 Å. No atom pair was found to be below these distance cutoffs in the complexes of galactose in any of the position−orientations.

Table 4: Key C−H···π Interactions in the Gal−Aromatic Residue Analogue Complexes[a]

C−H group	distance (Å) from atom/centroid/bond[b] nearest	second nearest	D_{plane} (Å)[c]	θ[d]	ω[e]
	1SLT Position−Orientation (−3.2 kcal/mol; Complex with 3-Methylindole)				
C4−H	2.4 (Cζ2)	2.5 (Cζ2−Cη)	2.4	16	88
C3−H	3.2 (Cϵ2−Nϵ1)	3.4 (Cϵ2−Cζ2)	2.7	57	41
	1HLC Position−Orientation (−3.6 kcal/mol; Complex with 3-Methylindole)				
C4−H	2.4 (Cζ2−Cη)	2.5 (Cη)	2.4	24	81
C5−H	2.8 (Cϵ2)	2.8 (Cϵ2−Nϵ1)	2.8	29	92
C3−H	2.8 (Cζ2)	3.0 (Cϵ2−Cζ2)	2.1	49	45
	2AAI Position−Orientation (−3.8 kcal/mol; Complex with 3-Methylindole)				
C4−H	2.3 (Cϵ2−Nϵ1)	2.4 (Nϵ1−Cδ1)	2.3	11	88
C6−H*R*	2.6 (Cζ2−Cη)	2.7 (Cη)	2.6	24	96
C3−H	3.6 (Nϵ1−Cδ1)	3.7 (Cϵ2−Nϵ1)	2.8	49	70
C5−H	3.3 (Cζ2)	3.4 (Cϵ2−Cζ2)	2.8	60	56
	1GAN Position−Orientation (−5.7 kcal/mol; Complex with 3-Methylindole)				
C3−H	2.9 (Cϵ2−Nϵ1)	3.1 (Cϵ2−Cζ2)	2.4	31	57
C4−H	2.6 (Cζ2)	2.6 (Cζ2−Cη)	2.6	26	86
C5−H	2.8 (centroid; 5-membered ring)	2.9 (Cδ2−Cϵ2)	2.8	21	89
	7CEL Position−Orientation (−6.6 kcal/mol; Complex with 3-Methylindole)				
C1−H	3.0 (Cϵ3)	3.1 (Cζ3−Cϵ3)	2.9	19	76
C3−H	2.9 (Cζ3)	2.9 (Cη−Cζ3)	2.8	18	74
C5−H	2.5 (centroid; entire ring)	2.5 (centroid; 6-membered ring)	2.5	11	91
	1GCA Position−Orientation (−7.5 kcal/mol; Complex with 3-Methylindole)				
C3−H	2.7 (Cϵ3)	2.7 (Cζ3−Cϵ3)	2.6	19	98
C4−H	3.2 (Cδ2−Cγ)	3.2 (centroid; 5-membered ring)	3.1	36	86
C5−H	2.9 (Cϵ2−Cζ2)	3.0 (Cϵ2)	2.9	17	101
	1AX1 Position−Orientation (−4.2 kcal/mol; Complex with Toluene)				
C3−H	2.6 (Cϵ1)	2.6 (Cδ1−Cϵ1)	2.4	26	66
C4−H	2.8 (Cγ−Cδ1)	2.9 (Cη)	2.8	29	92
C5−H	2.7 (Cη)	2.9 (Cη−Cϵ2)	2.7	23	97
	1BZW Position−Orientation (−4.7 kcal/mol; Complex with *p*-Hydroxytoluene)				
C4−H	2.5 (centroid of the ring)	2.6 (Cγ−Cδ1)	2.4	21	91
C5−H	3.0 (Cη)	3.0 (Cη−Cϵ2)	2.9	36	103
C6−H*S*	3.2 (Cϵ2−Cδ2)	3.3 (Cϵ2)	3.0	35	110
C3−H	2.8 (Cϵ1)	3.0 (Cδ1−Cϵ1)	2.6	73	40

[a] Only those interactions for which any one of the distances is ≤3.1 Å and $\theta < 70°$ and $\omega < 120°$ are listed here. The definitions of the various geometric parameters and the cutoff values for θ and ω are the same as specified by Nishio et al. (*21*); for distance, cutoff values of 3.1 Å have been used instead of 3.05 Å. [b] The distances between the hydrogen atom of the C−H group and all the atoms, bonds, and centroid of the aromatic ring were calculated. The nearest and second nearest are shown here. The midpoint was considered while calculating the distance of the hydrogen atom from the bond. In 3-methylindole, the centroid was calculated for the entire ring system and also for the 5- and 6-membered rings separately. The atom nomenclature is same as that used for the corresponding atoms in tryptophan, tyrosine, and phenylalanine. [c] Perpendicular distance of the hydrogen atom to the plane of the aromatic ring. The interaction energy correlates with the average deviation of D_{plane} from 2.9 Å (sum of the van der Waals radii of carbon and hydrogen atoms; ref *21*) with a correlation coefficient of 0.82. [d] Angle H−C−(nearest atom/centroid/bond of aromatic residue). The interaction energy correlates with the average deviation of θ from 0 (i.e., when C−H···nearest C atom are collinear) with a correlation coefficient of 0.88. [e] The angle between the plane of the aromatic ring and the plane formed by the hydrogen atom, nearest and second nearest bond/centroid/bond. The interaction energy correlates with the average deviation of ω from 90 with a correlation coefficient of 0.97.

the overall geometry in a complex way. The interaction is most stabilizing in the 1GCA and 7CEL position−orientations, and all of the interacting C−H atoms point directly ($\omega \approx 90$) to a carbon atom in these complexes (Table 4). Interestingly, the 1GCA position−orientation is observed in the glucose/galactose transport protein ($K_d = 0.2$ μM; *33*), and the 7CEL position−orientation is observed in an enzyme. All of the other position−orientations that have lesser stabilization energies are derived from lectins, which bind to monosaccharides in a shallow surface groove with weaker affinity (K_d values in the millimolar range; *3*).

Although the interaction energies for all the position−orientations are negative, and hence stabilizing, significant variations (−3.2 to −7.5 kcal/mol) are observed in these energies. This clearly shows that, although the aromatic residue contributes favorably to binding, the extent of contribution is highly dependent on the position−orientation of galactose relative to the aromatic residue. This, in turn, is dictated by rest of the binding site, since galactose has to optimize its interaction with respect to all of the residues in the binding site, not just the aromatic residue.

The Nature of the Aromatic Residue Seems Less Critical than the Position−Orientation. The interaction energies of the complexes of Gal with the three aromatic residue analogues are in the same range (Table 1), although it is known that the C−H···π interactions are influenced by the nature of the molecule contributing the π-electron cloud (*21*, *34*). Thus, the nature of the aromatic residue seems less critical than the position−orientation of the saccharide relative to the aromatic ring. However, tryptophan is found more frequently in Gal-binding sites than phenylalanine or tyrosine. A likely reason for this is that the larger surface area of tryptophan offers a wider range of position−orientations for galactose to optimize its interactions with the rest of the binding site without incurring any energetic penalty.

The Results Provide an Energetics-Based Explanation for the Experimental Observations on the Site-Specific Mutations of Aromatic Residues in Gal-Specific Proteins and the Role of Aromatic Amino Acids in Distinguishing between Glucose and Galactose. (1) The loss of the C−H···π interactions caused by mutating the aromatic residue reduces the enthalpy and hence the free energy of binding. The interactions of the C−H groups of Gal with nonaromatic amino acid residues are much weaker than those with the aromatic residues. The calculated strengths of the Gal−aromatic analogue interactions are in the same range as those calculated for hydrogen bonds (*20*, *27*−*32*). It has been shown that the loss of a hydrogen bond leads to significant increase in dissociation constant (*35*−*37*). In view of this, it is not surprising that mutating the aromatic residue has led to loss of binding in sugar-binding proteins (*10*, *12*−*19*). Experimentally also it has been observed that mutating the aromatic residue has a similar effect as that of mutating a hydrogen bonding residue in sugar-binding sites (*11*, *13*, *14*, *38*).

(2) Due to the apolar nature of the aromatic residue and of the *b* face of galactose, binding of galactose to the protein results in significant hydrophobic burial (the interaction energies calculated in the present study do not include this contribution). Such a hydrophobic burial also makes a favorable contribution to free energy of binding. It is to be

noted that the experimentally observed binding free energies (ΔG_{app}) are determined not only by the interaction energy (enthalpies or ΔG_{bind}) but also by other factors such as desolvation energy and conformational reorganization energy (*32*, *39*). Determining the individual contribution of each of these factors to the observed binding free energies is not possible because of which the computed interaction energies can only be qualitatively related to the results of site-specific mutation studies.

(3) The aromatic residue plays a role in selectivity also, i.e., in discriminating galactose from glucose. Comparing the interaction energies for the Glc−aromatic residue analogue complexes with the corresponding Gal complexes indicates that the glucose−aromatic residue interactions are unfavorable (positive interaction energy; Table 1) in position−orientations observed for galactose in Gal-specific proteins. The interaction energies for the complexes of glucose are favorable (i.e., negative) in 1GCA and 7CEL position−orientations. This is not surprising, since the former position−orientation is observed in glucose/galactose transporter protein, which binds both glucose and galactose in the same orientation (*33*); the latter position−orientation is observed in a glucose-specific enzyme (*40*).

Galactose has two nonpolar surface patches: the first patch is constituted by the hydrogen atoms bonded to C1, C3, and C5 atoms. The second patch, which can also be viewed as an extension of the first patch, is constituted by the hydrogen atoms bonded to C4, C5, and C6. The latter is discernible in galactose by virtue of the axial C4−OH group. The equatorially placed C4−OH group destroys this nonpolar surface patch in glucose (Figure S1). The binding site architecture in galactose-specific proteins has evolved in such a way that the rest of the binding site residues (i.e., residues excluding the aromatic residue) restrict galactose to interact with the aromatic residue through the second hydrophobic patch. In proteins such as the periplasmic glucose/galactose receptor, which bind both glucose and galactose, the binding site architecture is such that the aromatic residue interacts with the first nonpolar surface patch.

The equatorially oriented hydrogen atom bonded to Gal−C4 interacts with the aromatic residue in many position−orientations, and this is replaced by the bulkier and polar hydroxyl group in glucose. Hence, the interaction energy for the glucose−aromatic residue analogue complex is higher in many position−orientations observed for galactose in galactose−protein complexes. Among the hypothetical complexes of glucose (Table 1), the interaction energy is highest in the 2AAI position−orientation due to the close proximity of the C4−hydroxyl group to the aromatic ring (Table 3). Even though the 1AX1 position−orientation is observed for galactose in a Gal-specific protein, the interaction energy is 0.1 kcal/mol; the larger distance of the Glc−O4−H from the aromatic ring in this case alleviates the repulsion. In conformation I, the C4−O4−H atom is pointing toward the aromatic ring. The interaction energy is higher when glucose is in conformation II or III, possibly due to the repulsion between the lone pair of electrons of the O4 atom and the aromatic ring.

The Interaction Energies Calculated Using DFT and MP2 Correlate Well. An earlier study from our laboratory had reported the relative interaction energies of the complexes of galactose and of glucose with the aromatic residue analogues; the calculations were performed using density functional theory at U-B3LYP/6-31G**, and the relative interaction energies for 51 position−orientations were compared without any reference to the strength of the interactions (*26*). The interaction energies calculated using MP2 theory at the 6-311G++** level correlate well (correlation coefficient = 0.98) with those calculated for the same complexes using density functional theory (U-B3LYP/6-31G**; Figure 4). Thus the trends in the interaction energies observed in the present study seem to be applicable to the galactose− and glucose−aromatic residue complexes in the remaining 43 position−orientations also.

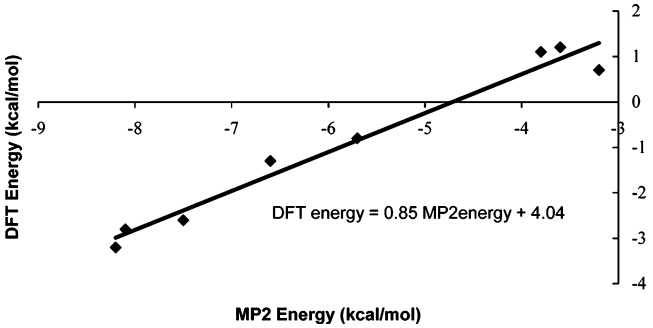


FIGURE 4: Correlation of the interaction energies calculated using MP2/6-311G++** (this study) with those obtained using DFT/6-31G** (*26*) for the galactose− and glucose−3-methylindole complexes (correlation coefficient = 0.98). Only those position−orientations for which the MP2-calculated interaction energy is negative (Table 1) have been plotted. The correlation coefficient is 0.97 if only the complexes of galactose are considered.

The interaction energies calculated using the MP2 method are more negative than those obtained using the DFT method because the former adequately assesses dispersive interactions (*41*−*45*) but the latter does not (*46*−*49*). Since the aromatic residue−sugar interactions are predominantly governed by C−H···π interactions, which are largely dispersive in nature, the stabilization from these interactions has not been completely accounted for by DFT methods.

In summary, this study shows that the role of the aromatic residue in galactose-binding sites is in detecting the subtle differences in the hydrophobic patches of galactose and glucose. The knowledge of the origin of the binding energy is crucial for understanding the balance of forces that drive binding. Hence this study will be a valuable guide for rational design of sugar-binding sites. In addition, the results can be used for protein redesign studies, as those published recently by Hellinga and co-workers (*50*, *51*).

ACKNOWLEDGMENT

We thank the Executive Director, Centre for Development of Advanced Computing (C-DAC), Bangalore, for giving permission to use the Bioinformatics Resources & Application Facility (BRAF). We also thank the scientists and staff of C-DAC, especially Dr. Rajendra R. Joshi, for help in using this facility. We thank Dr. R. B. Sunoj for helpful discussions and Amitay for help in using the Linux cluster. We thank the Gordon research group for the GAMESS software.

SUPPORTING INFORMATION AVAILABLE

Uncorrected BSSE interaction energies of galactose− and glucose−aromatic analogue complexes in various position−

orientations (Table S1), stereoviews of the complexes of galactose and glucose in conformations I, II, and III with the aromatic residue analogues in the eight position−orientations (Figure S1), and structures obtained during the intermediate stages of optimization of the glucose−3-methylindole complex in the 2AAI position−orientation showing the buckling of the aromatic ring and flattening of the pyranose ring at C4 (Figure S2). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Drickamer, K. (1997) Making a fitting choice: common aspects of sugar-binding sites in plant and animal lectins, *Structure 5*, 465−468.
2. Elgavish, S., and Shaanan, B. (1997) Lectin-carbohydrate interactions: different folds, common recognition principles, *Trends Biochem. Sci. 22*, 462−467.
3. Loris, R., Hamelryck, T., Bouckaert, J., and Wyns, L. (1998) Legume lectin structure, *Biochim. Biophys. Acta 1383*, 9−36.
4. Meyer, J. E. W., and Schulz, G. E. (1997) Energy profile of maltooligosaccharide permeation through maltoporin as derived from the structure and from a statistical analysis of saccharide−protein interactions, *Protein Sci. 6*, 1084−1091.
5. Pratap, J. V., Jeyaprakash, A. A., Rani, P. G., Sekar, K., Surolia, A., and Vijayan, M. (2002) Crystal structures of artocarpin, a Moraceae lectin with mannose specificity, and its complex with methyl-alpha-D-mannose: implications to the generation of carbohydrate specificity, *J. Mol. Biol. 317*, 237−247.
6. Quiocho, F. A., and Vyas, N. K. (1999) Atomic interactions between proteins/enzymes and carbohydrates, in *Bioorganic chemistry. Carbohydrates* (Hecht, S. M., Ed.) pp 441−457, Oxford University Press, New York.
7. Rao, V. S. R., Lam, K., and Qasba, P. K. (1998) Architecture of the sugar binding sites in carbohydrate binding proteins−a computer modeling study, *Int. J. Biol. Macromol. 23*, 295−307.
8. Rini, J. M. (1998) Lectin Structure, *Annu. Rev. Biophys. Biomol. Struct. 24*, 551−577.
9. Spiwok, V., Lipovova, P., Skalova, T., Buchtelova, E., Hasek, J., and Kralova, B. (2004) Role of CH/π interactions in substrate binding by *Escherichia coli* β-galactosidase, *Carbohydr. Res. 339*, 2275−2280.
10. Iobst, S. T., and Drickamer, K. (1994) Binding of sugar ligands to Ca(2+)-dependent animal lectins. II. Generation of high-affinity galactose binding by site-directed mutagenesis, *J. Biol. Chem. 269*, 15512−15519.
11. Adar, R., and Sharon, N. (1996) Mutational studies of the amino acid residues in the combining site of *Erythrina corallodendron* lectin, *Eur. J. Biochem. 239*, 668−674.
12. de Sousa, M., Roberts, L. M., and Lord, J. M. (1999) Restoration of lectin activity to an inactive abrin B chain by substitution and mutation of the 2 gamma subdomain, *Eur. J. Biochem. 260*, 355−361.
13. Frankel, A., Tagge, E., Chandler, J., Burbage, C., and Willingham, M. (1996) Double-site ricin B chain mutants retain galactose binding, *Protein Eng. 9*, 371−379.
14. Frankel, A. E., Burbage, C., Fu, T., Tagge, E., Chandler, J., and Willingham, M. C. (1996) Ricin toxin contains at least three galactose-binding sites located in B chain subdomains 1 alpha, 1 beta, and 2 gamma, *Biochemistry 35*, 14749−14756.
15. Hamelryck, T. W., Loris, R., Bouckaert, J., Dao-Thi, M. H., Strecker, G., Imberty, A., Fernandez, E., Wyns, L., and Etzler, M. E. (1999) Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*, *J. Mol. Biol. 286*, 1161−1177.
16. Iobst, S. T., and Drickamer, K. (1996) Selective sugar binding to the carbohydrate recognition domains of the rat hepatic and macrophage asialoglycoprotein receptors, *J. Biol. Chem. 271*, 6686−6693.
17. Lehar, S. M., Pedersen, J. T., Kamath, R. S., Swimmer, C., Goldmacher, V. S., Lambert, J. M., Blattler, W. A., and Guild, B. C. (1994) Mutational and structural analysis of the lectin activity in binding domain 2 of ricin B chain, *Protein Eng. 7*, 1261−1266.
18. Sphyris, N., Lord, J. M., Wales, R., and Roberts, L. M. (1995) Mutational analysis of the Ricinus lectin B-chains. Galactose-binding ability of the 2 gamma subdomain of *Ricinus communis* agglutinin B-chain, *J. Biol. Chem. 270*, 20292−20297.
19. Streicher, H., and Sharon, N. (2003) Recombinant plant lectins and their mutants, *Methods Enzymol. 363*, 47−77.
20. Nishio, M. (2004) CH/π hydrogen bonds in crystals, *CrystEngComm 6*, 130−158.
21. Nishio, M., Umezawa, Y., Hirota, M., and Takeuchi, Y. (1995) The CH/π interaction. Significance in molecular recognition, *Tetrahedron 51*, 8665−8671.
22. Sujatha, M. S., and Balaji, P. V. (2004) Identification of common structural features of binding sites in galactose-specific proteins, *Proteins 55*, 44−65.
23. Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., Windus, T. L., Dupuis, M., and Montgomery, J. A. (1993) General atomic and molecular electronic structure system, *J. Comput. Chem. 14*, 1347−1363 (www.msg.ameslab.gov/GAMESS/GAMESS.html).
24. www.cdacindia.com/html/parampma.asp.
25. Goldstein, H. (1980) *Classical mechanics*, Addison-Wesley, London.
26. Sujatha, M. S., Sasidhar, Y. U., and Balaji, P. V. (2004) Energetics of galactose- and glucose-aromatic amino acid interactions: implications for binding in galactose-specific proteins, *Protein Sci. 13*, 2502−2512.
27. Buemi, G., and Zuccarello, F. (2004) DFT study of the intramolecular hydrogen bonds in the amino and nitro-derivatives of malonaldehyde, *Chem. Phys. 306*, 115−129.
28. Creighton, T. E. (1993) *Proteins: structures and molecular properties*, 2nd ed., p 148, W. H. Freeman, New York.
29. Desiraju, G. R., and Steiner, T. (1999) *The weak hydrogen bond in structural chemistry and biology*, p 12, Oxford University Press, New York.
30. Fersht, A. (1999) *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*, 3rd ed., p 330, W. H. Freeman, New York.
31. Muller, A., and Leutwyler, S. (2004) Nucleobase pair analogues 2-pyridone·uracil, 2-pyridone·thymine, and 2-pyridone·5-fluorouracil: hydrogen-bond strengths and intermolecular vibrations, *J. Phys. Chem. A 108*, 6156−6164.
32. Shen, S.-Y., Yang, D.-Y., Selzle, H. L., and Schlag, E. W. (2003) Energetics of hydrogen bonds in peptides, *Proc. Natl. Acad. Sci. U.S.A. 100*, 12683−12687.
33. Zou, J. Y., Flocco, M. M., and Mowbray, S. L. (1993) The 1.7 Å refined X-ray structure of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*, *J. Mol. Biol. 233*, 739−752.
34. Desiraju, G. R., and Steiner, T. (1999) *The weak hydrogen bond in structural chemistry and biology*, Oxford University Press, New York.
35. Sahin-Toth, M., Lawrence, M. C., Nishio, T., and Kaback, H. R. (2001) The C-4 hydroxyl group of galactopyranosides is the major determinant for ligand recognition by the lactose permease of *Escherichia coli*, *Biochemistry 40*, 13015−13019.
36. Street, I. P., Armstrong, C. R., and Withers, S. G. (1986) Hydrogen bonding and specificity. Fluorodeoxy sugars as probes of hydrogen bonding in the glycogen phosphorylase-glucose complex, *Biochemistry 25*, 6021−6027.
37. Vermersch, P. S., Tesmer, J. J. G., and Quiocho, F. A. (1992) Protein−ligand energetics assessed using deoxy and fluorodeoxy sugars in equilibrium binding and high-resolution crystallographic studies, *J. Mol. Biol. 226*, 923−929.
38. Sharon, N., and Lis, H. (2002) How proteins bind carbohydrates: lessons from legume lectins, *J. Agric. Food Chem. 50*, 6586−6591.
39. Fothergill, M. D., and Fersht, A. (1991) Correlations between kinetic and X-ray analyses of engineered enzymes: crystal structures of mutants Cys→Gly-35 and Tyr→Phe-34 of tyrosyl-tRNA synthetase, *Biochemistry 30*, 5157−5164.
40. Divne, C., Stahlberg, J., Teeri, T. T., and Jones, T. A. (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*, *J. Mol. Biol. 27*5, 309−325.
41. Biot, C., Wintjens, R., and Rooman, M. (2004) Stair motifs at protein-DNA interfaces: nonadditivity of H-bond, stacking, and cation-pi interactions, *J. Am. Chem. Soc. 126*, 6220−6221.
42. Wang, Y., and Hu, X. (2002) A quantum chemistry study of binding carotenoids in the bacterial light-harvesting complexes, *J. Am. Chem. Soc. 124*, 8445−8451.

43. Wintjens, R., Lievin, J., Rooman, M., and Buisine, E. (2000) Contribution of cation-pi interactions to the stability of protein-DNA complexes, *J. Mol. Biol. 302*, 395−410.
44. Wintjens, R., Biot, C., Rooman, M., and Lievin, J. (2003) Basis set and electron correlation effects on ab initio calculations of cation-π/H-bond stair motifs, *J. Phys. Chem. A 107*, 6249−6258.
45. Tarakeshwar, P., and Kim, K. S. (2002) Comparison of the nature of π and conventional H-bonds: a theoretical investigation, *J. Mol. Struct. 615*, 227−238.
46. Elstner, M., Hobza, P., Frauenheim, T., Suhai, S., and Kaxiras, E. (2001) Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment, *J. Chem. Phys. 114*, 5149−5155.
47. Foti, M. C., DiLabio, G. A., and Ingold, K. U. (2003) Overlooked difference between hydrogen bonds of equal strength formed between catechol and an oxygen or nitrogen base. Experiments and DFT calculations, *J. Am. Chem. Soc. 125*, 14642−14647.
48. Hozba, P., and Sponer, J. (1999) Structure, energetics, and dynamics of the nucleic Acid base pairs: nonempirical ab initio calculations, *Chem. Rev. 99*, 3247−3276.
49. Sponer, J. V., Leszczynski, J., and Hobza, P. (2001) Hydrogen bonding, stacking and cation binding of DNA bases, *J. Mol. Struct.: THEOCHEM 573*, 43−53.
50. Dwyer, M. A., Looger, L. L., and Hellinga, H. W. (2004) Computational design of a biologically active enzyme, *Science 304*, 1967−1971.
51. Looger, L. L., Dwyer, M. A., Smith, J. J., and Hellinga, H. W. (2003) Computational design of receptor and sensor proteins with novel functions, *Nature 423*, 185−190.
52. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The protein data bank, *Nucleic Acids Res. 28*, 235−242.

BI050298B