Organic & Biomolecular Chemistry

Cite this: *Org. Biomol. Chem.*, 2012, **10**, 4186

www.rsc.org/obc

PAPER

Conformational mapping and energetics of saccharide–aromatic residue interactions: implications for the discrimination of anomers and epimers and in protein engineering†

Manju Kumari,[a] Raghavan B. Sunoj*[b] and Petety V. Balaji*[a]

Received 24th January 2012, Accepted 22nd March 2012
DOI: 10.1039/c2ob25182e

Aromatic residues play a key role in saccharide-binding sites. Experimental studies have given an estimate of the energetics of saccharide–aromatic residue interactions. In this study, dependence of the energetics on the mutual position-orientation (PO) of saccharide and aromatic residue has been investigated by geometry optimization of a very large number (164) of complexes at MP2/6-31G(d,p) level of theory. The complexes are of Tyr and Phe analogs with α/β-D-Glc, β-D-Gal, α-D-Man and α/β-L-Fuc. A number of iso-energy POs are found for the complexes of all six saccharides. Stacking and non-stacking modes of binding are found to be of comparable strengths. In general, complexes of *p*-OHTol are stronger than those of Tol, and those dominated by OH···O interactions are more stable than ones dominated by CH···π interactions. The strengths of OH···O/π interactions, but not those of CH···π, show large variations. Even though an aromatic residue has a large variety of POs to interact with a saccharide, distinct preferences are found due to anomeric and epimeric differences. An aromatic residue can interact from either the *a*- or *b*-face of Glc, but only through the *b*-face with Gal, its C4-epimer. In contrast, stacking interaction with Man (C2-epimer of Glc) requires the participation of the $-CH_2OH$ group and free rotation of this group, as is observed in solution, precludes all modes of stacking interactions. It is also found that an aromatic residue can be strategically placed either to discriminate or to accommodate (i) anomers of Glc and of Fuc and (ii) Gal/Fuc. Thus, analysis of the optimized geometries of by far the largest number of complexes, and with six different saccharides, at this level of theory has given insights into how Nature cleverly uses aromatic residues to fine tune saccharide specificities of proteins. These are of immense utility for protein engineering and protein design studies.

Introduction

Several biochemical, spectroscopic and crystallographic studies have established the importance of aromatic residues in carbohydrate-binding sites.[1] In fact, *E. coli* maltoporin has a platform of aromatic residues on which maltose slides while moving across the membrane.[2] Analyses of the 3-D structures of binding sites in proteins showed that Trp has the highest propensity among the twenty amino acid residues to be in a carbohydrate-binding site.[3] Saccharides can interact with Trp and other aromatic residues through H-bonding interactions such as NH···O, OH···π and CH···O. In addition, a saccharide can 'stack' on an aromatic residue through CH···π interactions. In fact, stacking often seems to be the preferred mode of interaction.

A number of experimental and computational studies have investigated the interactions between saccharides and aromatic residues in general and the contribution of CH···π interaction to the affinity and specificity of binding in particular.[4–7] Changes in the chemical shifts (δ) of the pyranose ring H atoms brought about by the addition of aromatic molecules (*e.g.*, benzene and Trp) have been used to infer CH···π mediated stacking interactions.[4] From such studies, it has been deduced that (i) galactose, but not mannose, can stack against benzene ring,[4a,c] (ii) the anomers of galactose differ in the way they interact with Phe or Trp, but the anomers of glucose interact in the same way, and (iii) galactose can stack on benzene in water but not in DMSO or acetonitrile.[4b] The gas phase interactions of methylglycosides of glucose, galactose and fucose with toluene, investigated by infra red ion-dip (IRID) measurements, showed that multiple modes of interactions are possible.[7c] The interactions between a saccharide and an aromatic residue have also been investigated when

[a]*Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India. E-mail: balaji@iitb.ac.in; Fax: +91 222 572 3480; Tel: +91 222 576 7778*
[b]*Department of Chemistry, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India. E-mail: sunoj@chem.iitb.ac.in; Fax: +91 222 576 7152; Tel: +91 222 576 7173*
†Electronic supplementary information (ESI) available. See DOI: 10.1039/c2ob25182e

these moieties are part of dicarboxy cyclohexanediol[5] or a 12-residue β-hairpin.[6] In these studies, no differences were found in the interactions due to a change in the configuration at C4 atom (*i.e.*, glucose *versus* galactose); also, no differences were found in the interactions of glucose with Trp and Phe. In the case of end-protected phenylalanine, the methylglycoside derivatives of saccharides were found to interact only through the backbone –NH and –CO groups, there were no interactions through the phenyl ring.[7a,b]

The enthalpies of solvation of saccharides or their derivatives with benzene have been measured by calorimetric methods.[4a,8] The enthalpy of solvation of mannose (~ -18 kcal mol^{-1}) was found to be less than that of galactose (~ -21 kcal mol^{-1}). The difference in the solvation enthalpies (2.4 kcal mol^{-1}) has been suggested to arise due to CH···π interactions.[4a] Using NMR spectroscopy, ΔH for the binding of β-1,4-linked $(GlcNAc)_3$ to the peptide AcAMP2, or its Phe → Trp or 2-naphthyl-Ala variants, was found to range between -10.8 and -15.3 kcal mol^{-1}.[9] In addition, this study found that the association constant and binding enthalpy increase with the size of the aromatic group and that the association weakens when Phe is replaced by 4-fluoro-Phe.

Quantum chemical calculations on saccharide–aromatic residue complexes have also been employed to quantify CH···π interactions. The calculations have been performed at different levels of theory: MP2/6-31G(d,p),[4c] MP2/6-31+G(d),[10] MP2/6-311+G(d),[11] MP2/6-311+G(d,p),[12] MP2/6-311++G(d,p),[13] $CCSD(T)_{limit}$,[14] M05-2X with Dunnings basis sets,[15] DFT-D/TZV2D[16] and DFT-D BP/def-TZVPP.[17] The saccharides considered in these studies are glucose, galactose, xylose and fucose, and the aromatic residues are Trp, 3-methylindole (3-MeIn; Trp analog), *p*-hydroxytoluene (*p*-OHTol; Tyr analog), toluene (Tol; Phe analog), benzene and coronene. The interaction energy depends on the relative position-orientation (PO) of the saccharide and the aromatic residue. In some of these studies, the POs have been taken from high-resolution structures of protein–carbohydrate complexes and these POs correspond to stacking mode of the interaction. In all these studies, the interaction was found to be stabilizing, with the magnitude varying between 2 and 12 kcal mol^{-1}, except when glucose is taken in a PO observed for galactose.[13a]

Experimental studies on model systems have been able to indicate whether or not a saccharide can interact with an aromatic residue through stacking.[4] Some of the studies have also been able to find the enthalpy of interaction.[4a,8,9] However, the mode of binding cannot be deduced from these studies. The *ab initio* calculations reported so far have considered only a limited number of POs, especially those observed in protein–carbohydrate complexes.[10–17] However, it is of relevance to determine other POs in which a saccharide can stack favorably on an aromatic residue. Such data provide knowledge about the potential energy surface of this binary complex and in turn, can be used to understand such phenomena as the energetics of transport across an aromatic platform as found in maltoporin.[18] In view of this, interaction energies for the saccharide–aromatic residue complexes have been calculated in a very large number of POs obtained by Monte Carlo sampling (Table 1). The six diastereomeric hexoses α-D-glucose, β-D-glucose, β-D-galactose, α-D-mannose, α-L-fucose and β-L-fucose have been considered to delineate differences, if any, in the patterns of interactions due to changes in the configurations of carbon atoms *i.e.*, anomers (α *versus* β of glucose and fucose), and C2 (glucose and mannose) and C4 (glucose and galactose) epimers (Fig. 1). In the present study, interaction energies are reported with *p*-OHTol (Y-series of complexes) and Tol (F-series of complexes). These data have been compared with those for the complexes of these six hexopyranoses with 3-MeIn (W-series of complexes) reported recently.[19]

Table 1 The number of complexes optimized

	Aromatic residue	
Saccharide	*p*-OHTol	Tol
α-D-Glucose	22	21[a]
β-D-Glucose	18	9
β-D-Galactose	12	10
α-D-Mannose	12	10
α-L-Fucose	12	12[a]
β-L-Fucose	13	13[a]
Total	**89**	**75**

[a] In these cases, two different initial geometries converged to the same geometry after optimization. Hence, data are given for only 72 complexes.

Based on the observation that the main chain atoms of the aromatic residue and bound galactose are on opposite sides of the aromatic ring in galactose-specific proteins, it has been suggested that the contribution of the main chain atoms to the interaction energy will be negligible.[20] Hence, *p*-OHTol and Tol were used as analogs of tyrosine and phenylalanine, respectively (Fig. 1). The saccharides are in the pyranose form and in their respective preferred chair conformations: $^4C_1$ for D-glucose, D-galactose and D-mannose, and $^1C_4$ for L-fucose. The initial structure used for conformational sampling of α-D-glucose–*p*-OHTol complex was taken from a protein–carbohydrate complex. The α-D-glucose–Tol complex was generated by replacing the –OH group of *p*-OHTol by a hydrogen atom. The other complexes were generated from these two by suitably changing the configuration of the appropriate carbon atom(s) of α-D-glucose. All the 12 different aromatic–saccharide complexes were subjected to stochastic/Monte Carlo sampling. The initial geometries for geometry optimization were chosen such that the intra-molecular interactions are primarily of CH···π type. The conformation of the –OH and $-CH_2OH$ groups were chosen such as to maximize intra-molecular H-bonding with the objective that OH···π interactions are minimized. This was because the primary objective of this study is to explore the stacking interactions between saccharides and aromatic residues.

Results

Stability of the complexes

The presence of the –OH group makes the Y-series complexes stronger than those of F-series for a given saccharide. Between 9 and 22 geometries were optimized for each saccharide and aromatic residue analog pair (Table 1). Details of the various interactions in the binary complexes of the Y- and F-series are

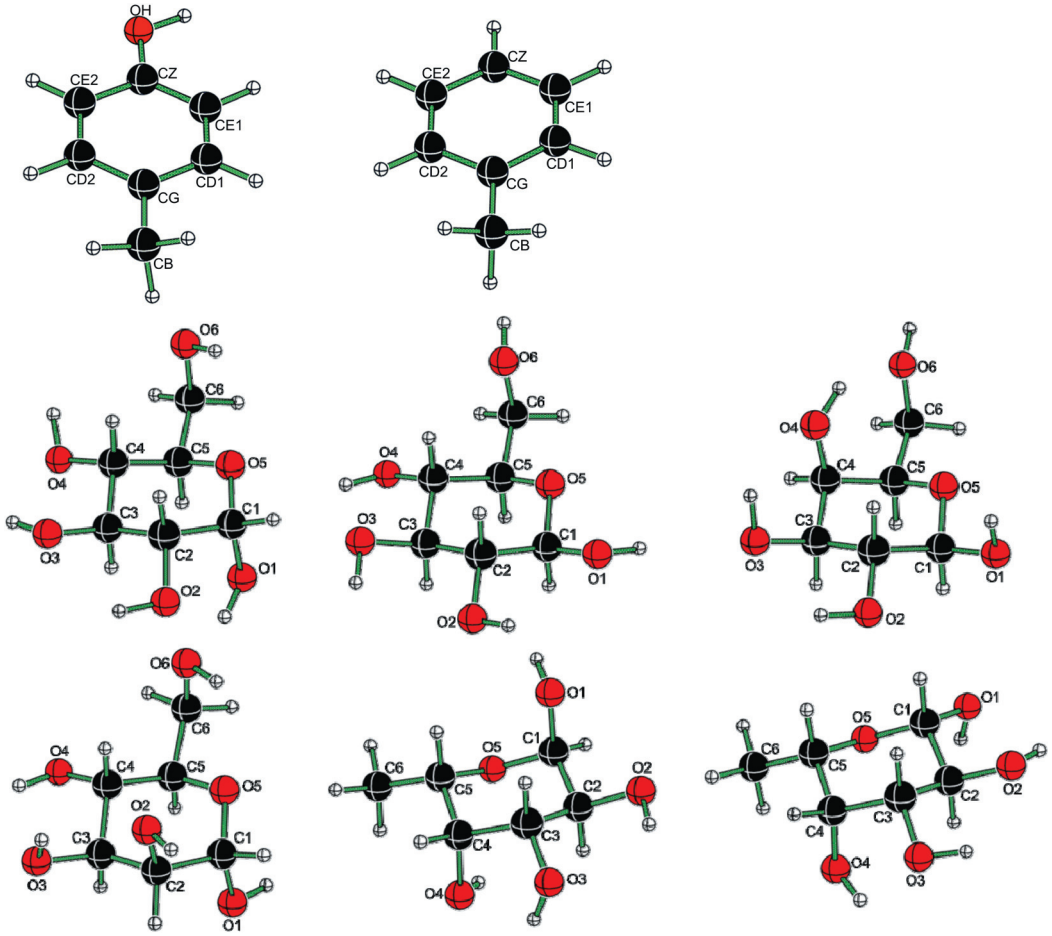

Fig. 1 Structures of, and atom nomenclatures employed for, *p*-OHTol and Tol (*top row, left to right*), α-D-glucose, β-D-glucose and β-D-galactose (*middle row, left to right*), and α-D-mannose, α-L-fucose and β-L-fucose (*bottom row, left to right*). The atom names for *p*-OHTol and Tol are the same as those used for the equivalent atoms in tyrosine and phenylalanine, respectively. The conformations of –OH, $–CH_3$ and $–CH_2OH$ groups have been set arbitrarily for the purpose of this illustration. Only non-hydrogen atoms are labeled. Color scheme: C – black, O – red, and H – ivory.

given in Tables S1 and S2,† respectively, and the renderings of these complexes, along with the distances for key inter-atomic interactions, are given in Fig. S1–S12.† Interaction energies were computed for all the 161 binary complexes (Table 2). In general, the complexes of *p*-OHTol are more stable than those of Tol for a given saccharide. This stabilization is to the extent of 3.8 to 7.8 kcal mol^{-1} at the upper bound of the interaction energy spectrum and 0.4 to 2.4 kcal mol^{-1} at the lower bound (Fig. 2). This additional stabilization is obviously due to the –OH group of *p*-OHTol. The interaction energies reported for POs found in protein–carbohydrate complexes are also in the same range, although the basis sets used in these studies are slightly different: MP2/6-311+G(d)[11] and MP2/6-311++G(d,p).[13b]

Types and strengths of interactions

There is no correlation between the type and strength of interactions. Overall, four types of interactions are seen: CH···π, CH···O, OH···π and OH···O. As expected, ρ_{bcp} decreases as the distance between the interacting atoms increases (Fig. S13†) and higher ρ_{bcp} implies stronger interactions. All the interactions that have ρ_{bcp} >0.0150 a.u. are of OH···O type; however, the converse is not true *i.e.*, ρ_{bcp} is not greater than 0.0150 for all the OH···O interactions (Tables S1 and S2†). For the CH···O and OH···O interactions, the CH and OH groups can be from either the saccharide or the aromatic residue. In the W-series, the NH···O interactions have lower ρ_{bcp} compared to OH···O in the Y-series, since the H–N atom is restricted to being in the plane of the aromatic ring. The –OH group of *p*-OHTol acts as a H-bond donor rather than as an acceptor in most of its interactions. The OH···π interaction is observed in the F-series (16 instances in 72 complexes) more often than in the Y-series (14 instances in 89 complexes); however, its frequency is highest in the W-series (29 instances in 65 complexes) because of the two ring aromatic system. The OH···π and CH···π interactions with ρ_{bcp} >0.0150 are found only in the W-series.[19]

The CH···O type of interactions are weaker and are restricted to the Y-series complexes. The CH···O interactions occur less frequently than the CH···π type. Most of the CH···O interactions are by the $–CH_3$ group of *p*-OHTol and Tol (equivalent to Cβ of Tyr and Phe); the CH groups of the aromatic ring and saccharide participate in only a few complexes. The CH···O interactions wherein the C–H is from a saccharide are observed only in the

Table 2 Interaction energies (in kcal mol^{-1}) for the complexes of saccharides with *p*-OHTol (Y-series) and Tol (F-series)[a]

α-D-Glucose		β-D-Glucose		β-D-Galactose		α-D-Mannose		α-L-Fucose		β-L-Fucose	
Y-1a	−16.40	**Y-2a**	−13.02	**Y-3a**	−14.35	**Y-4a**	−12.04	**Y-5a**	−10.32	**Y-6a**	−13.69
Y-1b	−14.49	**Y-2b**	−12.04	**Y-3b**	−13.02	**Y-4b**	−11.47	**Y-5b**	−10.16	**Y-6b**	−9.96
Y-1c	−13.63	**Y-2c**	−12.04	**Y-3c**	−13.02	**Y-4c**	−11.04	**Y-5c**	−10.14	**Y-6c**	−8.48
Y-1d	−12.31	**Y-2d**	−11.65	**Y-3d**	−12.66	**Y-4d**	−10.43	**Y-5d**	−8.91	**Y-6d**	−7.37
Y-1e	−11.50	**Y-2e**	−11.45	**Y-3e**	−12.20	**Y-4e**	−10.28	**Y-5e**	−6.86	**Y-6e**	−7.06
Y-1f	−11.50	**Y-2f**	−11.37	**Y-3f**	−11.27	**Y-4f**	−9.18	**Y-5f**	−6.86	**Y-6f**	−6.49
Y-1g	−10.16	**Y-2g**	−11.33	**Y-3g**	−11.25	**Y-4g**	−8.94	**Y-5g**	−6.68	**Y-6g**	−6.30
Y-1h	−9.95	**Y-2h**	−11.04	**Y-3h**	−11.04	**Y-4h**	−6.53	**Y-5h**	−6.51	**Y-6h**	−6.30
Y-1i	−9.66	**Y-2i**	−9.89	**Y-3i**	−10.83	**Y-4i**	−6.24	**Y-5i**	−5.87	**Y-6i**	−6.20
Y-1j	−9.23	**Y-2j**	−9.89	**Y-3j**	−9.25	**Y-4j**	−5.99	**Y-5j**	−5.70	**Y-6j**	−5.95
Y-1k	−8.73	**Y-2k**	−7.30	**Y-3k**	−7.97	**Y-4k**	−5.73	**Y-5k**	−5.62	**Y-6k**	−5.80
Y-1l	−8.49	**Y-2l**	−6.79	**Y-3l**	−6.70	**Y-4l**	−4.88	**Y-5l**	−5.15	**Y-6l**	−5.80
Y-1m	−8.41	**Y-2m**	−6.64							**Y-6m**	−5.21
Y-1n	−8.16	**Y-2n**	−6.01								
Y-1o	−7.83	**Y-2o**	−5.50								
Y-1p	−7.83	**Y-2p**	−5.43								
Y-1q	−7.66	**Y-2q**	−5.07								
Y-1r	−7.66	**Y-2r**	−4.38								
Y-1s	−6.73										
Y-1t	−6.34										
Y-1u	−6.17										
Y-1v	−6.07										
F-1a	−8.58	**F-2a**	−8.22	**F-3a**	−7.26	**F-4a**	−7.12	**F-5a**	−6.50	**F-6a**	−6.70
F-1b	−8.06	**F-2b**	−6.87	**F-3b**	−6.81	**F-4b**	−7.12	**F-5b**	−6.34	**F-6b**	−6.12
F-1c	−7.75	**F-2c**	−6.59	**F-3c**	−6.74	**F-4c**	−6.33	**F-5c**	−6.16	**F-6c**	−6.02
F-1d	−6.76	**F-2d**	−6.31	**F-3d**	−6.42	**F-4d**	−5.58	**F-5d**	−6.10	**F-6d**	−5.94
F-1e	−6.17	**F-2e**	−5.73	**F-3e**	−6.35	**F-4e**	−5.56	**F-5e**	−5.57	**F-6e**	−5.73
F-1f	−6.10	**F-2f**	−5.59	**F-3f**	−6.25	**F-4f**	−5.16	**F-5f**	−5.35	**F-6f**	−5.66
F-1g	−5.92	**F-2g**	−5.11	**F-3g**	−5.97	**F-4g**	−4.95	**F-5g**	−5.26	**F-6g**	−5.51
F-1h	−5.74	**F-2h**	−4.76	**F-3h**	−5.87	**F-4h**	−4.63	**F-5h**	−5.26	**F-6h**	−5.22
F-1i	−5.72	**F-2i**	−3.47	**F-3i**	−5.73	**F-4i**	−4.26	**F-5i**	−5.12	**F-6i**	−4.94
F-1j	−5.49			**F-3j**	−5.37	**F-4j**	−4.13	**F-5j**	−4.70	**F-6j**	−4.94
F-1k	−5.39							**F-5k**	−4.16	**F-6k**	−4.88
F-1l	−5.37									**F-6l**	−4.82
F-1m	−5.25										
F-1n	−5.12										
F-1o	−4.96										
F-1p	−4.90										
F-1q	−4.86										
F-1r	−4.32										
F-1s	−4.03										
F-1t	−3.70										

[a] The geometry optimized complexes are arranged in the increasing order of interaction energies and numbered as **Y-(1a-1v)**, **F-(1a-1t)**,…. The Y-series consists of **Y-1**, **Y-2**, **Y-3**, **Y-4**, **Y-5** and **Y-6** which are complexes of *p*-OHTol with α-D-glucose, β-D-glucose, β-D-galactose, α-D-mannose, α-L-fucose, and β-L-fucose respectively. Similarly for the F-series. In three cases of F-series, two different initial structures converged to the same structure after geometry optimization; these final structures are **F-1d**, **F-5g** and **F-6c**.

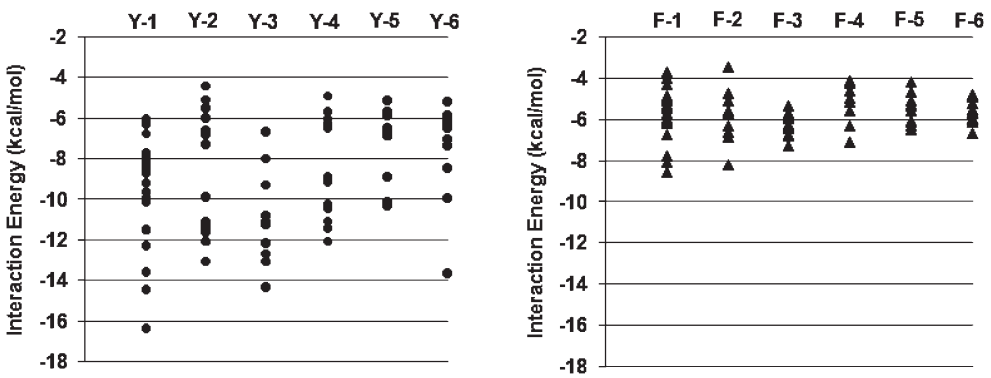

Fig. 2 Interaction energies of the various complexes of Y-series (left panel) and F-series (right panel).

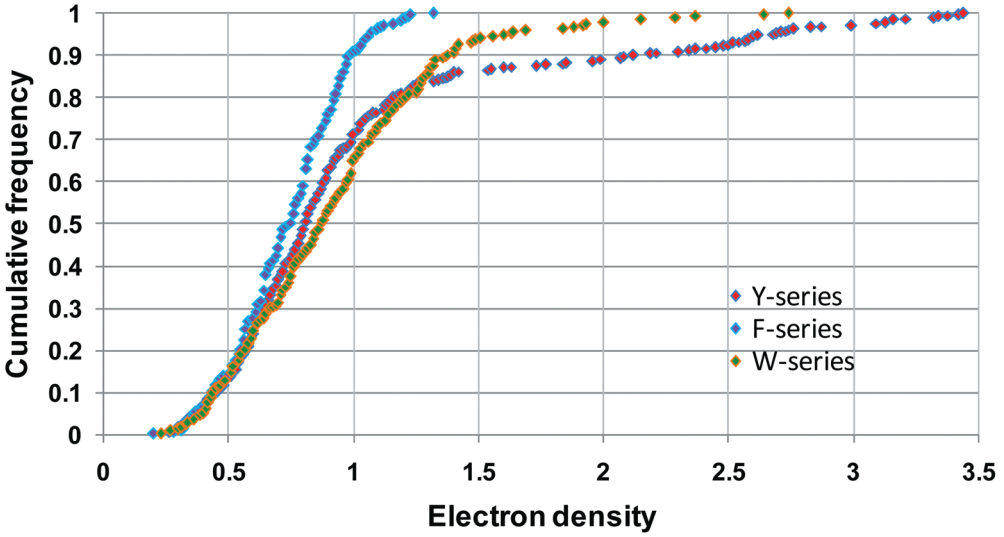

Fig. 3 Cumulative frequency graphs for the electron densities ($\rho_{bcp} \times 10^{-2}$) of 385 interactions found in the 89 complexes of the Y-series, 282 interactions found in the 75 complexes of the F-series and 276 interactions found in the 65 complexes of the W-series.

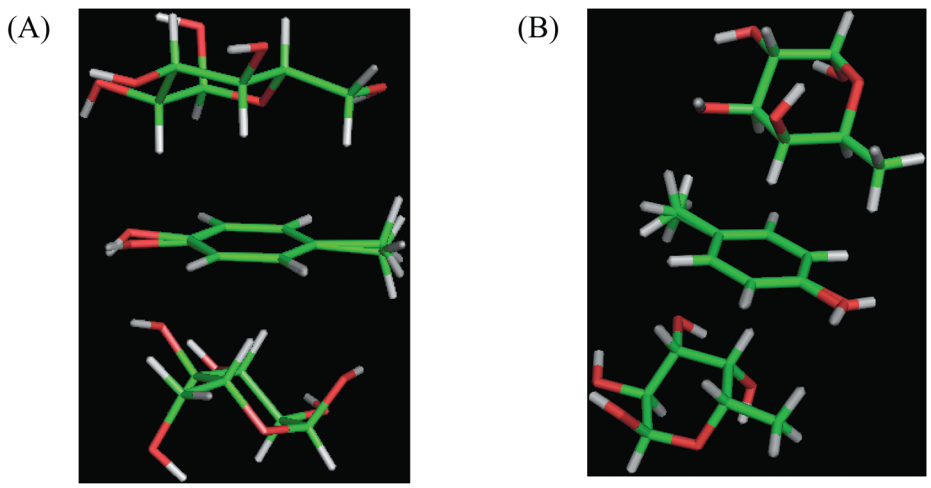

Fig. 4 Superposition of the complexes **Y-1o** and **Y-1r** (A) and **Y-5e** and **Y-5f** (B) using *p*-OHTol as the reference. It can be seen that the saccharide can interact with *p*-OHTol from either of the two sides of the planar ring due to the symmetrical disposition of the π-electron cloud. In (B), the two complexes are related to each other by C2 symmetry (CG–CZ axis) and have the same interaction energy. Other such symmetry-related pairs of binary complexes are {**Y-1e**, **Y-1f**}, {**Y-1o**, **Y-1p**}, {**Y-1q**, **Y-1r**}, {**Y-3b**, **Y-3c**}, {**F-4a**, **F-4b**}, {**F-5g**, **F-5h**} and {**F-6i**, **F-6j**}. In each of these cases, the corresponding initial geometries did not have any symmetry relationship.

Y-series and not in the F- or W-series for obvious reason. Only a few complexes have more than one CH···O interaction. The CH···O and multiple CH···π interactions are found mostly in complexes that span the lower spectrum of interaction energies.

F-series complexes tend to be weaker than those of Y- and W-series. Cumulative frequency graphs were plotted for the 385 interactions found in the 89 complexes of the Y-series, the 282 interactions found in the 75 complexes of the F-series and the 276 interactions found in the 65 complexes of the W-series (Fig. 3). It is clear from these graphs that the strongest interactions are found in the Y-series. The highest ρ_{bcp} found for an interaction is 0.0344 in the Y-series and 0.0132 a.u. in the F-series [0.0274 a.u. in the W-series]. Only 31% of all the interactions found in the Y-series have $\rho_{bcp} \geq 0.0100$ a.u. In the F- and W-series, these are 9 and 38%, respectively. This suggests that the interactions of saccharides with Tol are more likely to be weaker than those with *p*-OHTol or 3-MeIn.

Symmetry of the aromatic ring and symmetry-related POs

Symmetry-related complexes exist in the Y- and F-series, but not in the W-series. The π-electron cloud is distributed symmetrically across the plane of the aromatic ring. In addition, *p*-OHTol and Tol are symmetrical about the axis passing through the CG and CZ atoms (Fig. 1). Studies with model systems have shown that the aromatic ring can serve as an acceptor of C–H groups from both sides.[1b] Consequently, it is not surprising that (i) the saccharides can interact from both above and below the plane of the aromatic ring without compromising on the interaction energy (Fig. 4A) and (ii) a few pairs of binary complexes

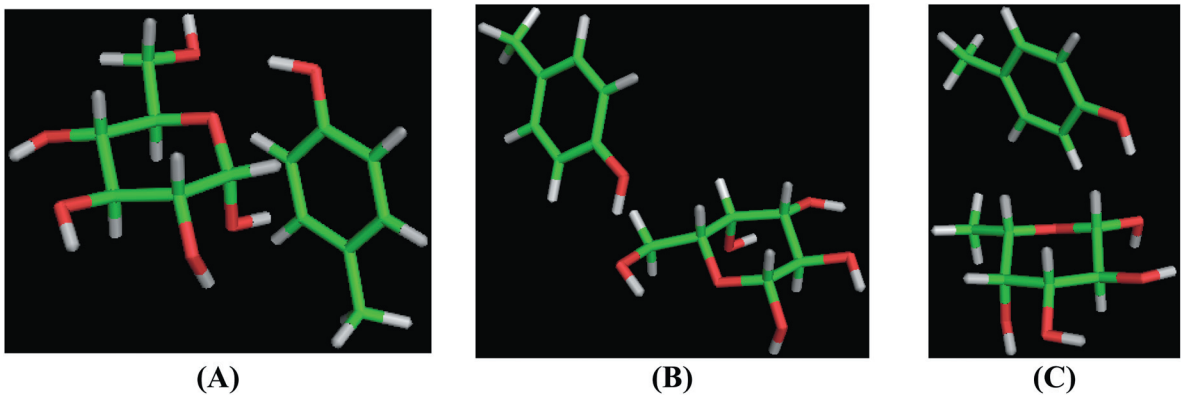

Fig. 5 Representative non-stacking complexes taken from the Y-series. The OH···O and OH···π interactions are dominant in these complexes. Shown here are the binary complexes of *p*-OHTol with (A) α-D-glucose (**Y-1a**), (B) β-D-galactose (**Y-3i**) and (C) β-L-fucose (**Y-6c**). Other complexes of this nature are **Y-(1c, 2b, 2d, 2i, 3a, 3d, 3e, 3h, 4a, 4d, 4e, and 6a)**.

are related by symmetry (Fig. 4B). Understandably, such symmetry-related complexes are not observed in the W-series.

Diversity of binding poses

A variety of stacking and non-stacking modes are available for binding. Visualization of the geometry optimized complexes (Fig. S1–S12†) clearly shows that a saccharide can interact with an aromatic residue in a large variety of poses. The saccharide stacks clearly on the aromatic ring (*vide infra*) in many of these poses whereas in few others, there is no stacking. An aromatic ring can interact with a saccharide either from above or below the pyranose ring (stacking type) or along the periphery (non-stacking type). Clearly, CH···π interactions are prevalent in the former, whereas OH···O and OH···π interactions predominate in the latter mode of binding and often give rise to non-stacking geometry.

Non-stacking complexes. These complexes are primarily stabilized by OH···O and OH···π interactions (Fig. 5, Table S1†) and the H-bond donor is often the –OH group of *p*-OHTol (Fig. S1–S6†). The hydroxyl hydrogen atoms deviate from the plane of the aromatic ring (CZ–OH dihedral angle: 10–50°) to facilitate the formation of a strong OH···O interaction. ρ_{bcp} for these interactions ranges between 0.0344 to 0.0177 a.u. (Table S1†). Although CH···π interactions are also seen in some of these complexes, they are weaker than the OH···O interactions present in the same complex. In fact, glucose and galactose prefer to interact with end-protected phenylalanine through OH···O and NH···O interactions, rather than through CH···π interactions in the gas phase.[7a,b] The complex **Y-2q** is an unusual non-stacking complex wherein the complex is essentially stabilized by CH···O interactions. In the F-series, non-stacking binding pose is observed only in **F-(1a, 1b and 3d)** (Table S2†). In these complexes, the –OH group of the saccharide forms OH···π interaction (Fig. S7, S9†). In the W-series, the non-stacking complexes are dominated by NH···O and/or OH···π interactions.

Stacking complexes. Most of the complexes obtained on optimization are in stacking geometries and the extent of stacking varies: from full or partial stacking of the pyranose ring to stacking of the $–CH_2OH$ group along with one C–H group from the pyranose ring. These complexes span a wide range of interaction energies. The interaction energies of the stacking and non-stacking type of complexes are comparable. This is so even for the fucose–benzene complexes: the interaction energies [$E_{CCSD(T)(limit)}$] for the non-stacking (OH···π mediated) type are between −5.5 and −4.0 kcal mol^{-1} and for the stacking type (CH···π mediated), between −4.9 and −4.4 kcal mol^{-1}.[14] However, unlike the fucose–benzene study where only four POs were considered for each type, in the present study, the number of complexes which are of stacking type is far more than that of non-stacking type (Tables S1 and S2†).

Non-stacking complexes are stronger than stacking complexes due to stronger OH/NH···O and OH···π interactions. All the types of interactions present in non-stacking complexes are also present in stacking complexes. "Are the strengths of interactions stabilizing the non-stacking complexes different from those in stacking complexes?" To answer this question, the cumulative frequency graphs for the strengths (ρ_{bcp}) of interactions were examined for each type of interaction by including/excluding the non-stacking complexes (Fig. 6). Clearly, the difference between stacking and non-stacking complexes lies primarily in the number of OH···O/NH···O and OH···π interactions of higher strength. It is found that (i) a larger number of OH···O interactions of relatively higher ρ_{bcp} are present in non-stacking complexes than in stacking complexes in the Y-series. Similarly, for the NH···O type interactions in the W-series. (ii) In the Y-series, the stronger OH···π interactions are present in non-stacking complexes, whereas in the F- and W-series, the strength of OH···π interaction is similar in both stacking and non-stacking complexes. (iii) A few interactions in stacking complexes can be as strong as those in non-stacking complexes.

The strengths of OH···O type of interactions show largest variation. The cumulative frequency plots show that the ρ_{bcp} values of the OH···O interaction show the largest variation and those of CH···O show the least variation. By implication, a CH···O interaction may be inferred to contribute to the same extent in different complexes. In contrast, the contribution of an

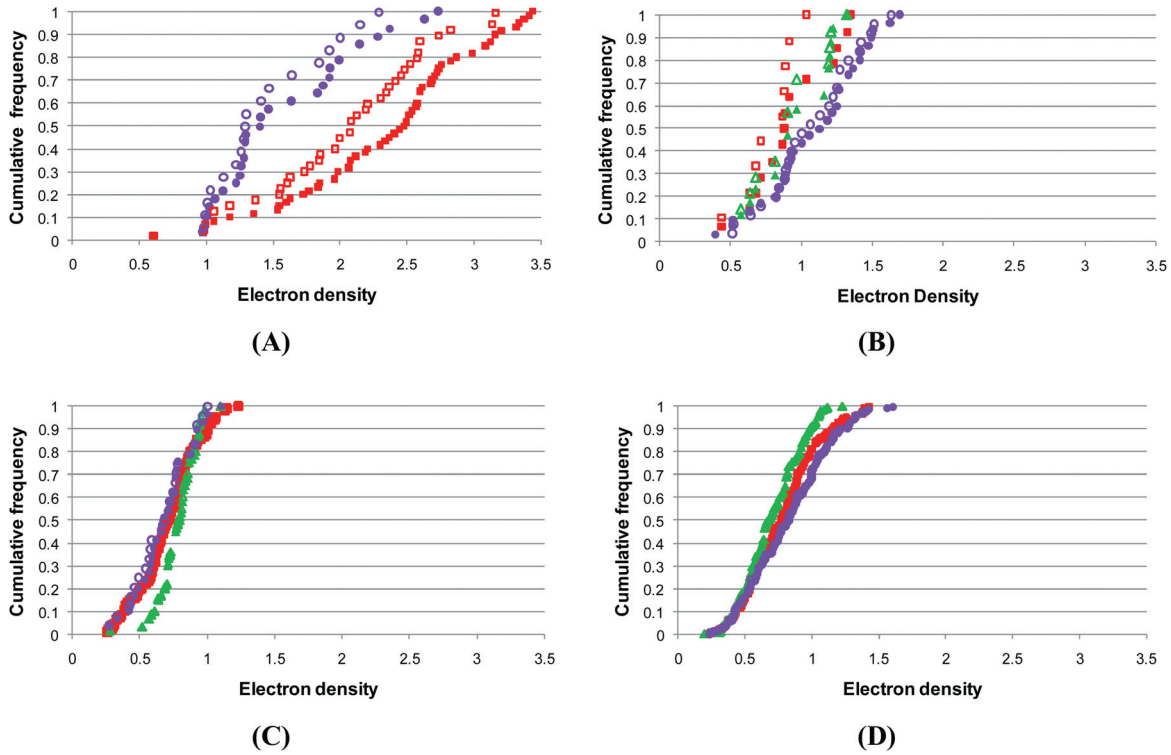

Fig. 6 Cumulative frequency graphs for the electron densities ($\rho_{bcp} \times 10^{-2}$) of the OH···O/NH···O (A), OH···π (B), CH···O (C) and CH···π (D) interactions observed in the Y-series (red square), F-series (green triangle) and W-series (violet circle). The graphs were plotted either by considering all the complexes (filled symbols) or by excluding the non-stacking complexes (unfilled symbols).

OH···O interaction can be either substantial or insignificant. Thus, as such, presence of OH···O interaction in a protein–carbohydrate complex does not indicate how important it is, especially when the data is from X-ray crystallography, since the location of hydrogen atoms cannot be precisely determined in such studies.

Asymmetry of the pyranose ring and preferred stacking modes

The size and location of the apolar patch is related to the conformation of the –CH_2OH group and configuration of the pyranose ring carbon atoms. The nature of atoms in the *a*-face and *b*-face of a pyranose ring depends on the configuration of the carbon atoms and the conformation of the pyranose rings (Fig. 1 and Fig. S14†). The size/area of the apolar patch in the *a*/*b*-face of a saccharide depends on the configuration of the carbon atoms and the conformation of the –CH_2OH group, which rotates freely in solution. The three staggered rotamers *gg*, *gt* and *tg* (Fig. S15†) are populated to various extents.[21] For example, the apolar patch formed by C3–H, C4–H and C5–H atoms in β-D-galactose can be extended if the –CH_2OH group is in *gg* or *gt* conformation.

Saccharides display very distinct stacking preferences. The stacking complexes were analyzed to gain insight into the preferences for binding through *a*-face or *b*-face and the apolar patch involved in stacking (Table 3). Overall, the interaction energies of the *a*-face and *b*-face complexes are, in general, comparable to each other, although in some cases the most stable complex is either from the *a*-face or *b*-face. The preference for *a*-face or *b*-face is absolute in the case of β-D-galactose (only through the *b*-face) and β-L-fucose (only through the *a*-face) (Fig. 7). Even α-L-fucose prefers to interact through the *a*-face since interaction through the *b*-face is seen in only one complex with *p*-OHTol (**Y-5c**) and 3-MeIn (**W-5j**). α-D-Glucose and β-D-glucose can interact through either the *a*-face or *b*-face (Fig. 7). However, α-D-glucose for *b*-face interaction and β-D-glucose for *a*-face interaction require the participation of the –CH_2OH group (Table 3). Even though α-D-mannose also can interact through either *a*- or *b*-face, the participation of the –CH_2OH group is critical for its interaction because of the axial –OH group on C2 (*a*-face) and C1 (*b*-face) atoms. Free rotation of the –CH_2OH group, as happens in solution, precludes all the modes of stacking interactions. This explains the observation that galactose, but not mannose, can stack against benzene.[4a,c]

The stacking complexes in the Y-series were superposed on each other using aromatic ring as the reference (Fig. S16†). Such a superposition was performed separately for the stacking complexes of F- and W-series also. No specific PO is preferred by any of the six saccharides, implying that any of the six saccharides can interact with the aromatic residue from a given position. In contrast, superposition of all the stacking complexes of a saccharide with any of the three aromatic residues *p*-OHTol, Tol, and 3-MeIn showed dramatic patterns and preferences:

(i) Comparison of the complexes of α-D-glucose and β-D-glucose (Fig. 8A, B; Movie S1†) and of α-L-fucose and β-L-fucose (Fig. 8E, F; Movie S2†) shows that the configuration of

Table 3 Apolar patches that interact with the aromatic residue[a]

Saccharide	*a*/*b* Face interaction	Descriptions of apolar patch interaction
α-D-Glucose	*a*	AP-124 (mainly), AP-466 (few cases)
	b	AP-356 or AP-566. At least one C6–H projects towards the centroid of the aromatic ring
β-D-Glucose	*a*	AP-246
	b	AP-135. C1–H, C3–H and C5–H are perpendicular to the aromatic ring in some complexes. In others, any one of these three C–H groups projects towards the centroid of the aromatic ring
β-D-Galactose	*a*	No stacking
	b	AP-135 or AP-345 with/without the participation of C6–H
α-D-Mannose	*a*	AP-466
	b	AP-566 in the Y- and F-series, AP-356 in the W-series
α-L-Fucose	*a*	AP-345. One of the three C–H groups projects towards the centroid of the aromatic ring in the F- and Y-series; two such interactions in the W-series
	b	AP-12 in both Y-5c and W-5a
β-L-Fucose	*a*	AP-135 or AP-345, with/without the participation of C6–H
	b	No stacking

[a] For brevity, abbreviations such as "AP-124" are used to refer to apolar patches. "AP" denotes "apolar" and "124" indicates that this apolar patch is formed by the C1–H, C2–H and C4–H groups.

the anomeric carbon determines the POs that allow favorable interactions. Thus, an aromatic residue can be used for anomeric discrimination in the case of glucose and fucose.

(ii) The distinctive differences in the preferred binding modes of β-D-glucose and β-D-galactose (Fig. 8B, D; Movie S3†) and of α-D-glucose and α-D-mannose (Fig. 8A, C; Movie S4†) shows that the aromatic residue can also be used to discriminate epimers.

(iii) Even though β-D-galactose and β-L-fucose (6-deoxy-L-galactose) use similar regions for stacking due to their topological similarities, some differences do exist because, upon superposition, C3 and C2 atoms of galactose overlap on the C5 and O5 atoms of fucose and *vice versa* (Fig. 8D, F; Movie S5†). Thus, the position of the aromatic residue can be tuned either to accommodate both galactose and fucose (using AP-135) or to discriminate between them (using the $–CH_3$ and –OH groups).

Intramolecular H-bonds and interaction energies in stacking complexes

The –OH groups, when not engaged by other interactions, can contribute to stability by additional interactions. The –OH and $–CH_2OH$ groups of a saccharide can form an arc of intramolecular hydrogen bonds: O1···O2···O3···O4···O6 (Fig. 1) depending on the conformation of the –OH/$–CH_2OH$ groups and the configuration of the carbon atoms. In the case of α-D-mannose, the –OH group at C1 cannot participate in intramolecular hydrogen bonding due to its axial orientation. Instead, it can participate in OH···π (Tol, *p*-OHTol, 3-MeIn) or OH···O (*p*-OHTol)

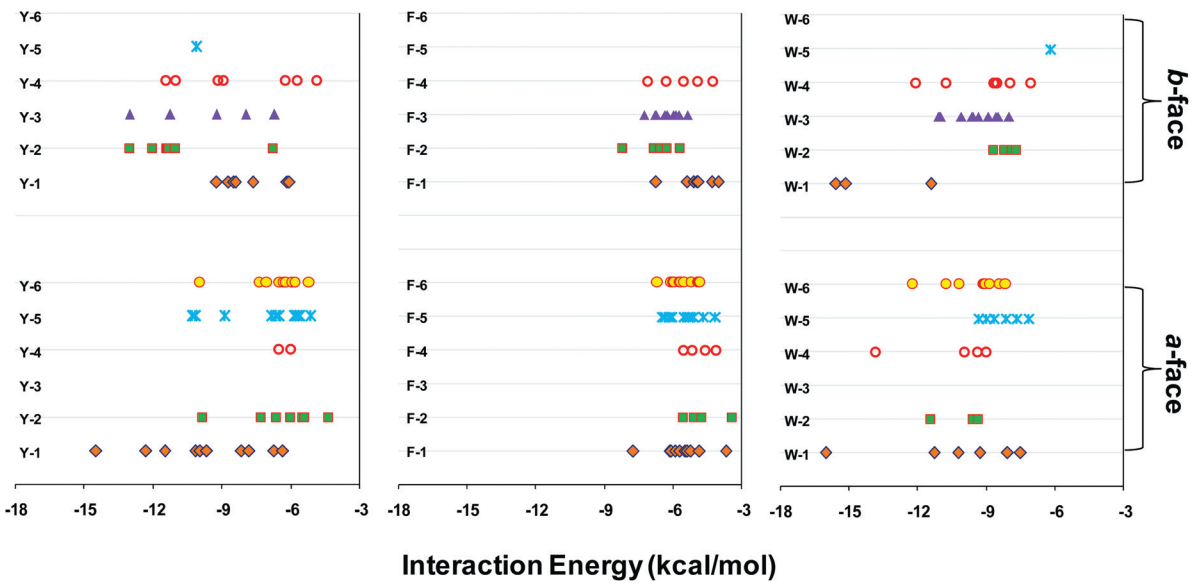

Fig. 7 Interaction energies (along the horizontal axis) for the stacking complexes of the Y-series (*left panel*), F-series (*middle panel*) and W-series (*right panel*). A different symbol has been used for each saccharide merely to assist in visual association of the data for the same group of complexes in the upper and lower halves of each panel. Nomenclature of the various complexes *viz*., **Y-1, Y-2**, *etc.* are as described in the footnote to Table 2. The interaction energies are from Table 2 for Y- and F-series.

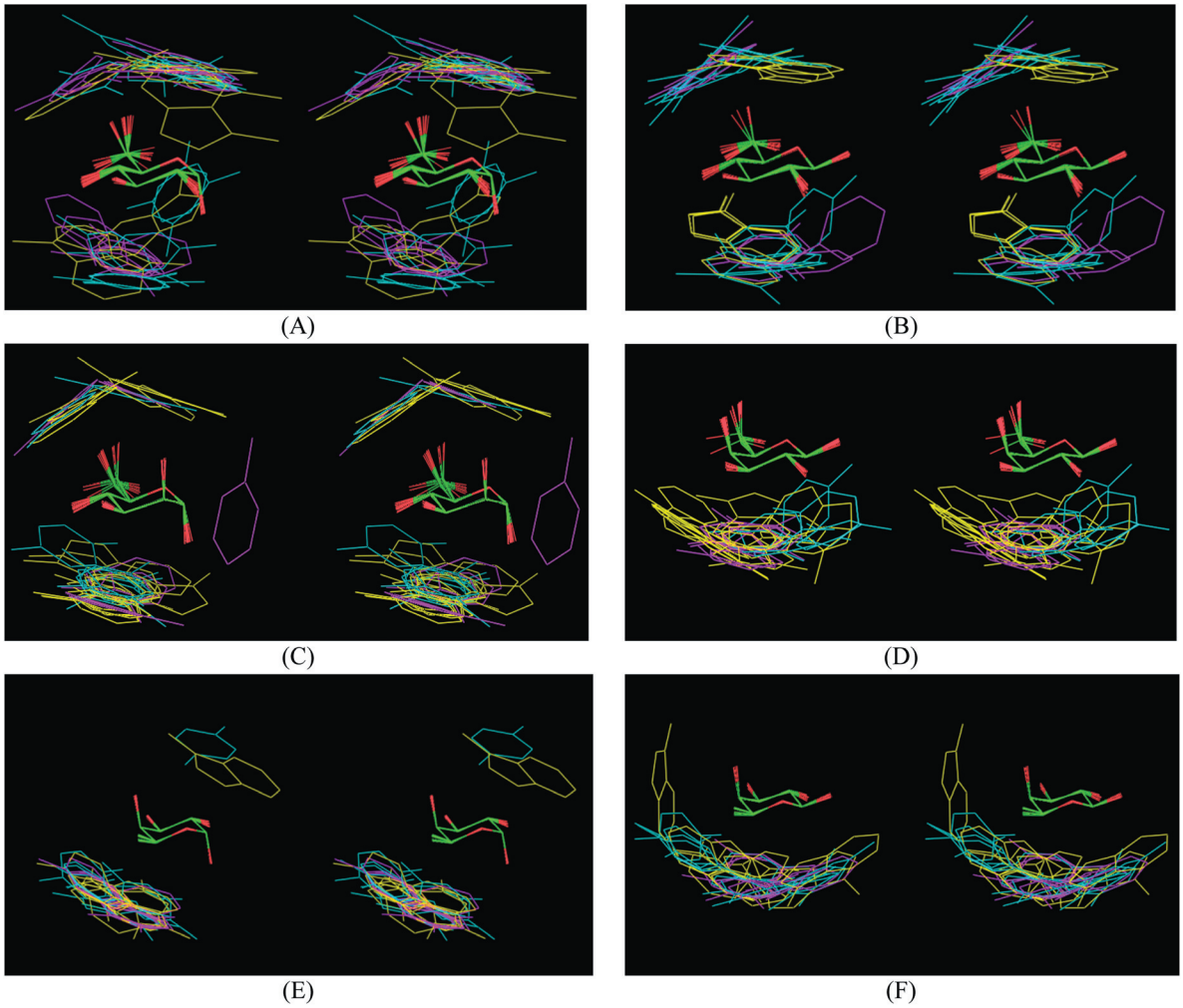

Fig. 8 Stereo views of the superposed stacking complexes of α-D-glucose (A), β-D-glucose (B), α-D-mannose (C), β-D-galactose (D), α-L-fucose (E) and β-L-fucose (F) with all the three aromatic residues. The atoms of the pyranose rings were used as reference for structural superposition. *p*-OHTol is rendered in cyan, Tol in magenta and 3-MeIn in yellow. The carbon and oxygen atoms of the saccharides are in green and red, respectively. Hydrogen atoms have been omitted for visual clarity. The saccharides are oriented in the same way in all the six panels to enable visual comparison of binding preferences.

interaction. These intermolecular H-bonds affect the interaction energy of binary complexes. However, NMR spectroscopic studies of glucose, galactose and lactose at temperatures as low as −20 °C have shown that the hydroxyl protons are more or less free rotors, except for the –OH group at glucose:C3 when it participates in inter-residue hydrogen bond (O3–H···O5′) in lactose[22] although some of the C–O torsions show biased sampling.[23] The $-CH_2OH$ group also can rotate freely in solution.

The interaction energies of the complexes **F-3a** and **F-3c** differ from each other by 0.5 kcal mol^{-1} (Table 2). The C–H groups at C3, C4 and C5 participate in CH···π interactions in both these complexes (Fig. 9A). However, in **F-3c**, the C6–H group also interacts with the aromatic residue. Despite this additional CH···π interaction, **F-3c** is less stable (by ∼0.5 kcal mol^{-1}) than **F-3a**. This is presumably because of the intramolecular H-bonds formed by the –OH groups of the saccharide in **F-3a** (Fig. S17†). The conformation of the $-CH_2OH$ group does not obviously affect the interaction energy if the PO is such that the aromatic ring is away from this group (Fig. 9B and 9C).

Potential energy surface of the complexes

Potential energy surfaces of saccharide–aromatic residue complexes are shallow. Some pairs of binary complexes have very similar POs. Even such small differences in the POs may lead to gain/loss of interactions (Fig. 10). For example, the O4–H···OH interaction is absent in **Y-2m** (−6.64 kcal mol^{-1}) but is present in **Y-2k** (−7.30 kcal mol^{-1}) (Table S1†). Comparison of **F-5c** (−6.16 kcal mol^{-1}) and **F-5e** (−5.57 kcal mol^{-1}) shows that the C6–H···CE1 interaction is absent in the latter whereas the C3–H···CD2 interaction is marginally stronger in the former (Table S2†). Such is also the case between **W-(4l, 4m)** (−7.95, −7.08 kcal mol^{-1}). Both **W-(6f, 6i)** (−9.04, −8.17 kcal mol^{-1}) have a CH···π interaction mediated by AP-345 and a CH···O

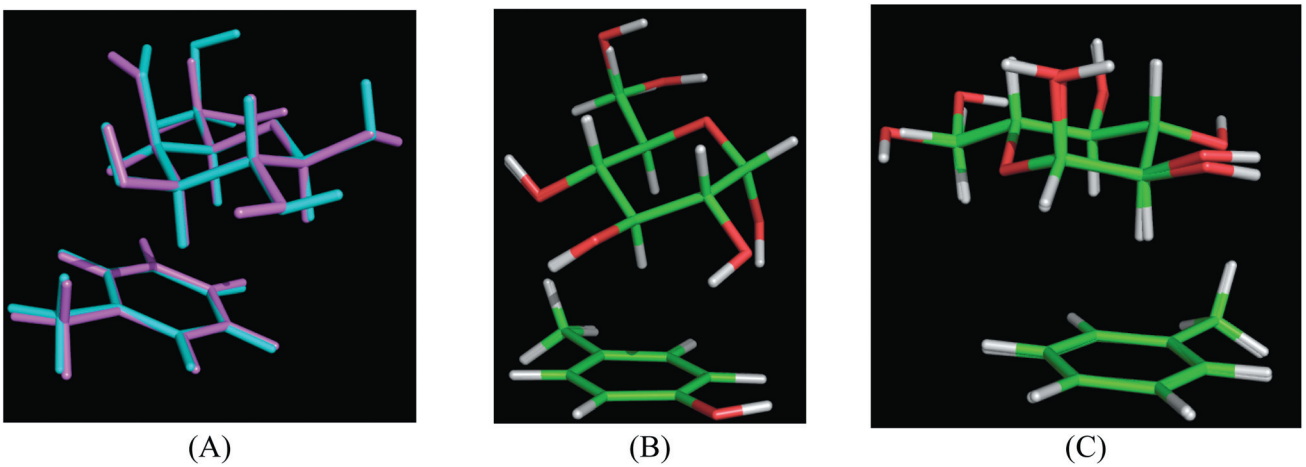

Fig. 9 Superposition of pairs of complexes which differ from each other in the conformation of the –OH and/or –CH_2OH groups of the saccharide. (A) β-D-Galactose–toluene complexes **F-3a** (cyan) and **F-3c** (magenta). The –CH_2OH group is in *gg* and *gt* conformations in **F-3a** and **F-3c**, respectively (Fig. S9†). (B) α-D-Glucose–*p*-OHTol complexes **Y-1k** and **Y-1m** wherein the –CH_2OH group is in *gt* and *gg* conformations, respectively (Fig. S1†). (C) α-D-Glucose–Tol complexes **F-1i** and **F-1j** wherein the –CH_2OH group is in *tg* and *gt* conformations, respectively (Fig. S7†).

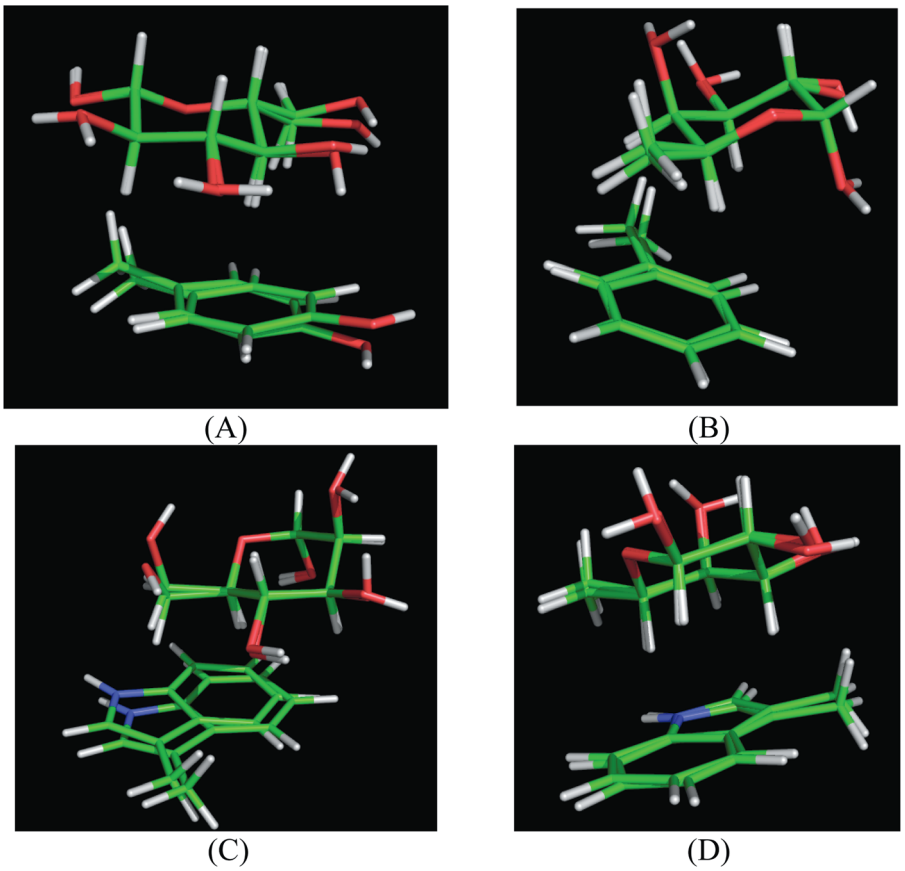

Fig. 10 Structural superposition of pairs of complexes which differ from each other marginally in their PO. (A) **Y-(2k, 2m)**, (B) **F-(5c, 5e)**, (C) **W-(4l, 4m)** and (D) **W-(6f, 6i)**. The atoms of the pyranose rings were used as reference for superposition. In all the four cases, the interaction energies of the superposed binary complexes differ from each other by <1 kcal mol^{-1}.

interaction. But the marginal change in the PO results in the gain of a C6–H···N1 interaction in **W-6f**.[24] As can be seen, the changes in interaction energies are <1.00 kcal mol^{-1} due to these marginal differences in the POs.

There are also pairs of POs that have quite different POs but have comparable interaction energies. Many of these are in the upper end of the interaction energy spectra. For example, in the Y-5 series, the interaction is through different POs in **Y-(5a, 5b,**

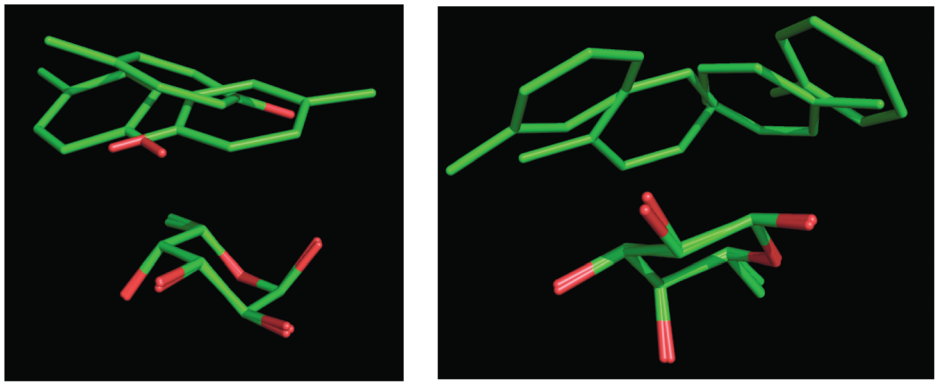

Fig. 11 Structural superposition of complexes which differ substantially from each other in their positions. (A) **Y-(5a, 5b, 5d)** and (B) **F-(6a, 6c, 6d, 6g)**. The atoms of the pyranose rings were used as reference for superposition. The difference in interaction energies of the superposed binary complexes is not more than 1.00 and 1.50 kcal mol^{-1} for (A) and (B), respectively. Hydrogen atoms are not shown for visual clarity.

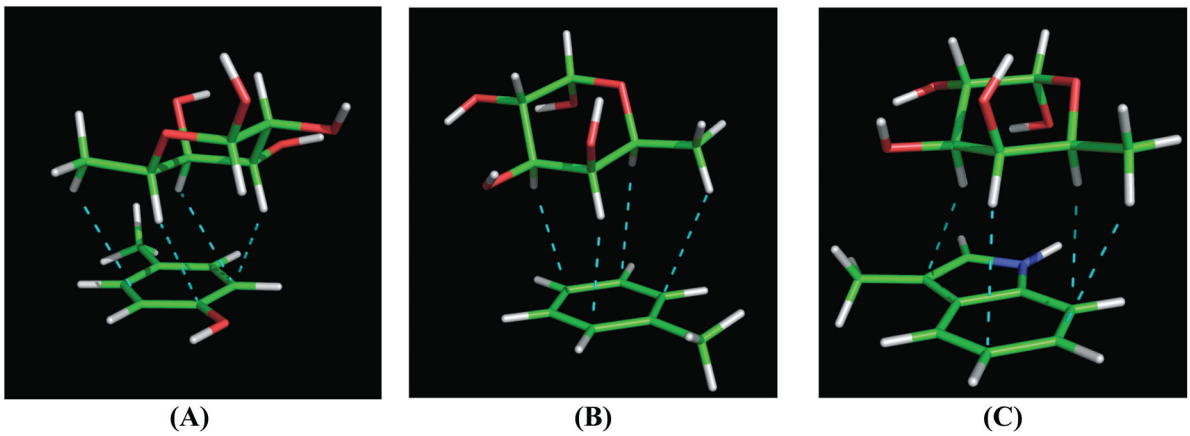

Fig. 12 Representative stacking complexes are from Y, F, and W series dominated by CH···π interactions (cyan dotted lines). (A) **Y-6m**, (B) **F-5i** and (C) **W-5e**.

5c) but the interaction energies are comparable (−10.32, −10.16 and −10.14 kcal mol^{-1}, respectively) (Fig. 11A). Another such example comes from the F-6 series (Fig. 11B): **F-(6a, 6c, 6d, 6g)** [−6.70, −6.02, −5.94, −5.51 kcal mol^{-1}]. This indicates that comparable interaction energies can still be obtained by substantial change in positions indicating that the PES is shallow; a similar inference was drawn recently from a DFT-D study on the complexes of benzene and saccharides.[17] These are the iso-energy points on the potential energy surface (PES); however, data from the present study is inadequate to draw any inferences on the path joining such iso-energy points and the energy barriers between them.

Effect of the size of the acceptor group on strength of CH···π interactions

Larger surface area of the aromatic residue enables participation of multiple C–H groups leading to stronger complexes. Some of the binary complexes in the Y- and F-series, especially with fucose, are stabilized by only CH···π interactions (up to four such interactions). The size of the aromatic ring becomes relevant in these cases. The interaction energies of such complexes range between −5.00 and −6.00 kcal mol^{-1} (Fig. 12). The occurrence of four CH···π interactions is more frequent in the W-series and the interaction energy is around −8.00 kcal mol^{-1}. These complexes are more stable than those in the Y- or F-series by about 2 to 3 kcal mol^{-1} because of the two rings. In fact, in model systems wherein the saccharide is constrained to interact with the aromatic residue in stacking mode, it has been observed that the binding enthalpy increases with the size of the aromatic ring and that the association weakens when Phe is replaced by 4-fluoro-Phe.[9]

Discussion

Stacking of saccharides on aromatic residues is observed in a number of carbohydrate-binding proteins. Site-specific mutagenesis studies of binding-site aromatic residues have shown that stacking interactions play a very important role in binding and specificity. Molecular mechanics force fields and most

functionals of density functional theory are inadequate to capture the energetics of stacking interactions due to the presence of dispersion component. Several *ab initio* quantum chemical and experimental studies have been performed to quantify the contribution of stacking interactions to binding. Experimental studies have provided the enthalpies of solvation of model saccharides by solvents such as benzene, or free energy changes associated with the addition of aromatic residues to a solution of saccharides. The computations are for gas phase but only a limited number of POs have been considered so far. The interactions of the C–H groups with the π electron cloud, which are at the heart of stacking, change depending on the mutual PO of saccharide and aromatic residue and also on the configuration of the pyranose ring atoms. Hence, in the present study, interaction energies at the MP2(full)/6-31G(d,p) level have been computed for a very large number of POs and by considering six diastereomeric hexopyranoses, which permits the delineation of the differences in energetics between some of the anomers and epimers. This, by far, is the largest number of complexes (saccharide and aromatic analogs) for which calculations of such a high level have been reported.

This study was initiated with the primary objective of exploring stacking interactions between saccharides and aromatic residues, their energetics and contributions to affinity/specificity. Hence, the initial geometries were chosen so that the interactions are primarily of CH···π type. However, the initial PO changed substantially during geometry optimization and the interaction was through a 'normal' hydrogen bond [the term 'normal' hydrogen bond is used here *vis-à-vis* a short strong or low-barrier hydrogen bond].[25] Such complexes are categorized as non-stacking complexes. This turned out to be fortuitous since it allowed comparison of the energetics of interaction through stacking with that of a normal hydrogen bond mediated interaction. Both types of interactions are observed in model systems[7a,b] and protein–carbohydrate complexes.[26,27] For example, fucose interacts simultaneously with one tryptophan residue through OH···O type interaction and through CH···π type interaction with another tryptophan in the binding sites of *Aleuria aurantia*[26] and *Ralstonia solanacearum*[27] lectins. Methylglycosides of glucose and galactose interact only through OH···O and NH···O interactions with end-protected phenylalanine in the gas phase.[7a,b] The interactions of these methylglycosides with toluene in the gas phase were found to be mediated by either CH···π or OH···π interactions depending upon the configuration at the anomeric carbon atom and the C4 ring carbon.[7c]

The –OH and $–CH_2OH$ groups are free to rotate and interact during optimization. They may form OH···O/π interactions thereby stabilizing the binary complex. These interactions may form along with (stacking type) or without (non-stacking type) CH···π interactions. In principle all the stacking complexes from lower to upper bound interaction energy spectra (Table 2) can still be favored in binding sites because the –OH groups may interact with binding site residues other than the stacking aromatic residue.

Even though multiple atom pairs are involved in interactions in a binary complex, the interaction energy is comparable to those reported for a single 'normal' hydrogen bond.[28] Thus, mutating a binding-site aromatic residue is equivalent to mutating a hydrogen bond. However, it is to be borne in mind that, be it a saccharide–aromatic residue interaction or a normal hydrogen bond, the energetics of binding of ligands to proteins either in dilute solutions or in cellular environments will be different from the *in vacuo* or gas phase condition under which the interaction energies have been computed.

The interaction energies of stacking complexes of the three aromatic residues are comparable to each other. This does not necessarily mean that replacing one aromatic residue by any of the other two will not have any effect. A Tyr may have restricted conformational freedom due to the –OH group forming H-bonds. A Phe, without such restriction, is relatively free to move allowing for a fine tuning of its interacting with the saccharide. On the other hand, the –OH group of Tyr can contribute to binding by interacting with a C–H group of saccharide. A Trp, because of its larger surface area, has the advantage of allowing the saccharide many more POs to interact with.

The utility of these data for protein engineering can be gleaned further by a comparison of some of the binary complexes. Structural superposition of **Y-1u** (−6.17 kcal mol^{-1}) and **Y-1v** (−6.07 kcal mol^{-1}) shows that the aromatic residue occupies the same position in both the complexes but the location of the Cβ and –OH groups are different (Fig. 13A). The types and strengths of interactions are same except that an additional CH···O interaction is present in **Y-1u**. Other such pairs are **Y-4c** (−11.04 kcal mol^{-1}) and **Y-4g** (−8.94 kcal mol^{-1}) (Fig. 13B), **F-3f** (−6.25 kcal mol^{-1}) and **F-3h** (−5.87 kcal mol^{-1}) (Fig. 13C) and **F-5b** (−6.34 kcal mol^{-1}) and **F-5i** (−5.12 kcal mol^{-1}) (Fig. 13D). The differences in the locations of the Cβ atoms in these pairs show that, in proteins, the aromatic residue can approach the saccharide from different directions without much compromise in the energy of interaction with the aromatic residue.

The interaction energy can also be fine tuned by a change in orientation. The saccharide can interact with the aromatic system at the same position but in a different orientation. Such a change in orientation can affect the strength of CH···π interaction marginally. For instance **F-5b** and **F-5i** are interacting with the aromatic residue with the same set of CH atoms (AP-4356) displaying four CH···π interactions (Fig. 13D). Even though the position is the same, orientations are different leading to changes in the distance between the C–H groups and acceptor; this leads to a difference in the interaction energy to the tune of ∼1.2 kcal mol^{-1}.

The results obtained in this study, and the inferences drawn from them, have implications for biological processes such as transport. Maltoporin has a layer of aromatic residues on which maltose can slide along with other polar residues. During transport, maltose continuously forms and breaks hydrogen bond with the polar residues. It has been inferred from earlier modeling studies that maltose encounters ∼4.0 kcal mol^{-1} energy barrier during transport.[18] As maltose slides, its position-orientation with respect to the aromatic residues changes and simultaneously, the hydrogen bonds are also broken/re-made. A fine balance between the energetic penalty of breaking of hydrogen bonds and the gain of interaction energy through stacking interactions ensures the smooth transport. This requires that the saccharide interacts with the aromatic residue in many different POs with comparable interaction energies. The calculations reported in the present study support this possibility.

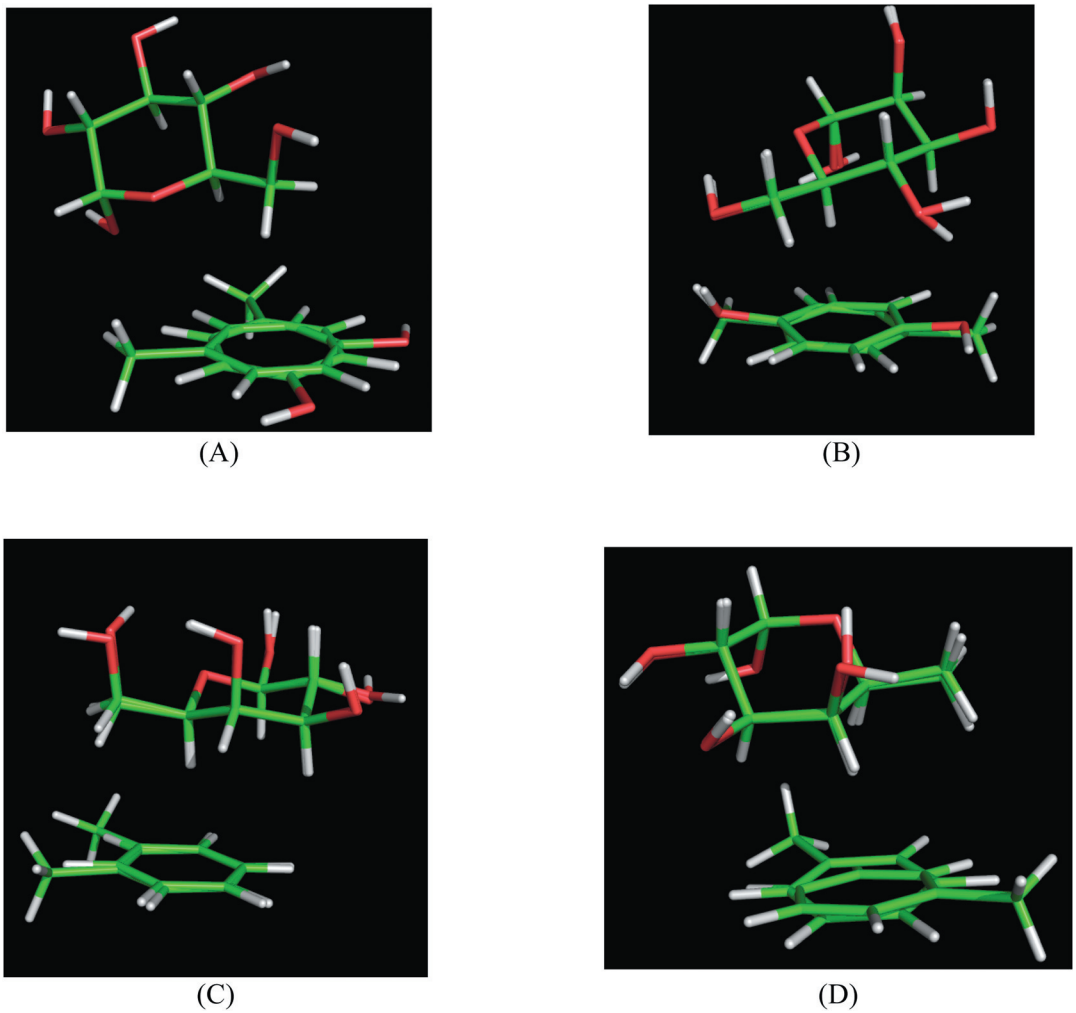

Fig. 13 Superposition of pairs of binary complex that differ from each other only in the orientation of the aromatic ring (*cf.* location of the Cβ atom). (A) **Y-1u** and **Y-1v**; (B) **Y-4c** and **Y4g**; (C) **F-3f** and **F-3h** and (D) **F-5b** and **F-5i**.

Conclusions

The present MP2 calculations on a diverse set of saccharide–aromatic residue complexes encompassing as many as 164 complexes give a reasonable estimate of gas phase interaction energies. The geometry optimized complexes display a wide range of POs, which include both stacking and non-stacking modes of interaction. All the possible types of interactions *viz*., OH···O, CH···O, OH···π and CH···π are observed in these complexes. Non-stacking complexes are dominated by OH···O and OH···π interactions. The complexes at the lower end of the interaction energy spectrum are dominated by multiple CH···π interactions. The frequency of occurrence of OH···π interaction is highest in the W-series followed by F- and Y-series. The interactions of saccharides with Tol are more likely to be weaker than those with *p*-OHTol or 3-MeIn. The saccharides can interact from both above and below the plane of the aromatic ring without compromising on the interaction energy. The aromatic residue interacts with the saccharide either through *a*-face or *b*-face or in some cases from both *a*/*b*-faces and different stacking modes are preferred by different saccharides. The energetics of the binary complexes dominated by CH···π interactions is found to be dependent on (1) the area of apolar patch, (ii) the size of the aromatic ring (3-MeIn *versus* *p*-OHTol/Tol).

Computational details

Software. Macromodel 9.5[29] was used for obtaining the initial geometry *viz*., PO of the saccharide with respect to the aromatic residue analog GAMESS[30] or Gaussian09[31a] was used for geometry optimization. Most of the calculations have been performed using GAMESS. Gaussian09 was used to take advantage of the parallel computing facility that became available during the later stages of the study. Gaussian03[31b] was used for counterpoise (CP) correction to minimize basis set superposition error (BSSE). AIM2000[32] was used for AIM analysis.

Conformational search. The saccharides were taken to be in the pyranose form and in their respective preferred chair conformations: $^4C_1$ for D-pyranoses, and $^1C_4$ for L-pyranoses.

The initial structure used for conformational sampling of α-D-glucose–*p*-OHTol complex was taken from a protein–carbohydrate complex (pdb id 1MPM). The other eleven saccharide–aromatic complexes were obtained by (i) suitably modifying the configurations of the relevant pyranose ring carbon atoms and (ii) replacing the –OH group of *p*-OHTol by –H. These initial structures were used for conformation search by the hybrid Monte Carlo minimum/low mode search method. The search consisted of 50 000 steps in vacuum.

Geometry optimization and energy calculation. Conformers in which the saccharide has stacking interaction with the aromatic residue were selected from conformational search output for geometry optimization at the MP2/6-31G(d,p) level of theory. Harmonic vibrational frequency calculations were performed to characterize the nature of the stationary points. The MP2 calculations with a moderate size basis set such as 6-31G (d,p) have been reported to give rise to higher stabilization of complexes due to large BSSE.[33] A larger basis set can be used to overcome BSSE, but this becomes computationally expensive for systems such as those considered in the present study (about 40 atoms). Hence, the counterpoise method was used to minimize BSSE.[34] The optimized structures were subjected to single point energy calculation as well as single point counterpoise correction at the MP2(full)/6-31G(d,p). In the present study, 50%-BSSE corrected[35] energy values have been used for all discussion. The optimized Cartesian coordinates for all the binary complexes are provided in the ESI.†

Characterization of interactions. AIM2000 was used for analyzing the relative strengths of various interactions in the optimized complexes. Bader's theory of atoms in molecules (AIM) is one of the widely used quantum mechanical tools to understand the H-bonding interaction.[36] The wave functions of the optimized complexes were first generated at the MP2(full)/6-31G(d,p) and topological analyses were performed. The key parameters such as the electron densities at the bond critical points (ρ_{bcp}) along the bond paths of the interacting atoms between saccharide and aromatic system have been identified. The ρ_{bcp} values are reported in atomic unit (a.u.).

Acknowledgements

PVB and RBS thank the Department of Science and Technology, Government of India for financial support through a grant-in-aid (SR/S1/PC-39/2005). MK thanks IIT Bombay for research fellowship. The authors also thank the Computer Center, IIT Bombay for generous computing facilities.

References

1 (*a*) P. V. Balaji, *Mini-Rev. Org. Chem.*, 2011, **8**, 222; (*b*) M. Nishio, *Phys. Chem. Chem. Phys.*, 2011, **13**, 13873; (*c*) P. Wildberger, C. Luley-Goedl and B. Nidetzky, *FEBS Lett.*, 2011, **585**, 499; (*d*) S. Jin, Y. Cheng, S. Reid, M. Li and B. Wang, *Med. Res. Rev.*, 2010, **30**, 171; (*e*) M. Mazik, *Chem. Soc. Rev.*, 2009, **38**, 935; (*f*) M. D. Diaz, M. C. Fernandez-Alonso, G. Cuevas, F. J. Canada and J. Jimenez-Barbero, *Pure Appl. Chem.*, 2008, **80**, 1827; (*g*) J. Fantini, *Curr. Med. Chem.*, 2007, **14**, 2911; (*h*) M. Machovik and S. Janecek, *Cell. Mol. Life Sci.*, 2006, **63**, 2710.
2 R. Dutzler, Y. F. Wang, P. Rizkallah, J. P. Rosenbusch and T. Schirmer, *Structure (Cambridge, MA, U. S.)*, 1996, **4**, 127.
3 A. Malik and S. Ahmad, *BMC Struct. Biol.*, 2007, **7**, 1.
4 (*a*) L. Bautista-Ibanez, K. Ramirez-Gualito, B. Quiroz-Garcia, A. Rojas-Aguilar and G. Cuevas, *J. Org. Chem.*, 2008, **73**, 849; (*b*) S. Vandenbussche, D. Diaz, M. C. Fernandez-Alonso, W. Pan, S. P. Vincent, G. Cuevas, F. J. Canada, J. Jimenez-Barbero and K. Bartik, *Chem.–Eur. J.*, 2008, **14**, 7570; (*c*) M. C. Fernandez-Alonso, F. J. Canada, J. Jimenez-Barbero and G. Cuevas, *J. Am. Chem. Soc.*, 2005, **127**, 7379.
5 G. Terraneo, D. Potenza, A. Canales, J. Jimenez-Barbero, K. K. Baldridge and A. Bernardi, *J. Am. Chem. Soc.*, 2007, **129**, 2890.
6 Z. R. Laughrey, S. E. Kiehna, A. J. Riemen and M. L. Waters, *J. Am. Chem. Soc.*, 2008, **130**, 14625.
7 (*a*) E. J. Cocinero, P. Çarçabal, T. D. Vaden, J. P. Simons and B. J. Davis, *Nature*, 2011, **469**, 76; (*b*) E. J. Cocinero, P. Çarçabal, T. D. Vaden, B. J. Davis and J. P. Simons, *J. Am. Chem. Soc.*, 2011, **133**, 4548; (*c*) E. C. Stanca-Kaposta, D. P. Gamblin, J. Screen, B. Liu, L. C. Snoek, B. G. Davis and J. P. Simons, *Phys. Chem. Chem. Phys.*, 2007, **9**, 4444.
8 K. Ramirez-Gualito, R. Alonso-Rios, B. Quiroz-Garcia, A. Rojas-Aguilar, D. Diaz, J. Jimenez-Barbero and G. Cuevas, *J. Am. Chem. Soc.*, 2009, **131**, 18129.
9 M. I. Chavez, C. Andreu, P. Vidal, N. Aboitiz, F. Freire, P. Groves, J. L. Asensio, G. Asensio, M. Muraki, F. J. Canada and J. Jimenez-Barbero, *Chem.–Eur. J.*, 2005, **11**, 7060.
10 V. Spiwok, P. Lipovova, T. Skalova, E. Buchtelova, J. Hasek and B. Kralova, *Carbohydr. Res.*, 2004, **339**, 2275.
11 V. Spiwok, P. Lipovova, T. Skalova, E. Vondrackova, J. Dohnalek, J. Hasek and B. Kralova, *J. Comput.-Aided Mol. Des.*, 2006, **19**, 887.
12 M. N. A. Mohamed, H. D. Watts, J. Guo, J. M. Catchmark and J. D. Kubicki, *Carbohydr. Res.*, 2010, **345**, 1741.
13 (*a*) M. S. Sujatha, Y. U. Sasidhar and P. V. Balaji, *Biochemistry*, 2005, **44**, 8554; (*b*) M. S. Sujatha, Y. U. Sasidhar and P. V. Balaji, *J. Mol. Struct. (Theochem)*, 2007, **814**, 11.
14 (*a*) S. Tsuzuki, T. Uchimaru and M. Mikami, *J. Phys. Chem. B*, 2009, **113**, 5617; (*b*) S. Tsuzuki, T. Uchimaru and M. Mikami, *J. Phys. Chem. A*, 2011, **115**, 11256.
15 R. M. Kumar, M. Elango and V. Subramanian, *J. Phys. Chem. A*, 2010, **114**, 4313.
16 R. K. Raju, A. Ramraj, M. A. Vincent, I. H. Hillier and N. A. Burton, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6500.
17 S. Kozmon, R. Matuska, V. Spiwok and J. Koca, *Chem.–Eur. J.*, 2011, **17**, 5680.
18 R. Dutzler, T. Schirmer, T. Karplus and S. Fischer, *Structure (Cambridge, MA, U. S.)*, 2002, **10**, 1273.
19 M. Kumari, P. V. Balaji and R. B. Sunoj, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6517.
20 M. S. Sujatha and P. V. Balaji, *Proteins: Struct., Funct., Bioinf.*, 2004, **55**, 44.
21 R. Stenutz, I. Carmichael, G. Widmalm and A. S. Serianni, *J. Org. Chem.*, 2002, **67**, 949.
22 L. Poppe and H. van Halbeek, *Nat. Struct. Biol.*, 1994, **1**, 215.
23 H. Zhao, Q. Pan, W. Zhang, I. Carmichael and A. S. Serianni, *J. Org. Chem.*, 2007, **72**, 7071.
24 Refer to Supporting Information of ref. 19.
25 (*a*) S. Rajagopal and S. Vishveshwara, *FEBS Lett.*, 2005, **272**, 1819; (*b*) I. Rozas, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2782.
26 (*a*) M. Fujihashi, D. H. Peapus, N. Kamiya, Y. Nagata and K. Miki, *Biochemistry*, 2003, **42**, 11093; (*b*) M. Wimmerova, E. Mitchell, J. F. Sanchez, C. Gautier and A. Imberty, *J. Biol. Chem.*, 2003, **278**, 27059.
27 N. Kostlanova, E. P. Mitchell, H. Lortat-Jacob, S. Oscarson, M. Lahmann, N. Gilboa-Garber, G. Chambat, M. Wimmerova and A. Imberty, *J. Biol. Chem.*, 2005, **280**, 27839.
28 (*a*) K. Wendler, J. Thar, S. Zahn and B. Kirchner, *J. Phys. Chem. A*, 2010, **114**, 9529; (*b*) T. E. Creighton, *Proteins: Structures and Molecular Properties*, W. H. Freeman, New York, 2nd edn, 1993, p. 148; (*c*) G. R. Desiraju and T. Steiner, *The Weak Hydrogen Bond in Structural Chemistry and Biology*, Oxford University Press, New York, 1999, p. 12; (*d*) A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman, New York, 3rd edn, 1999, p. 330; (*e*) G. Buemi and F. Zuccarello, *Chem. Phys.*, 2004, **306**, 115; (*f*) A. Muller and S. Leutwyler, *J. Phys. Chem. A*, 2004, **108**, 6156; (*g*) S. Y. Shen, D. Y. Yang, H. L. Selzle and E. W. Schlag, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 12683.
29 *MacroModel, version 9.5, Schrödinger*, LLC, New York, 2007.

30 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. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347.
31 (*a*) M. J. Frisch, *et al.*, *GAUSSIAN 09 (Revisions A.02 and E.01)*, Gaussian, Inc., Wallingford CT, 2009; (*b*) M. J. Frisch, *et al.*, *GAUSSIAN 03 (Revisions C.02 and E.01)*, Gaussian, Inc., Wallingford CT, 2004. (See ESI† for the full list).
32 F. Biegler-König and J. Schoenbohm, *J. Comput. Chem.*, 2002, **23**, 1489.
33 L. Sobczyk, S. J. Grabowski and T. M. Krygowski, *Chem. Rev.*, 2005, **105**, 3513.
34 S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553.
35 K. S. Kim, Y. J. Lee, H. S. Choi, J. Kim and J. H. Jang, *Chem. Phys. Lett.*, 1997, **265**, 497.
36 R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Clarendon Press, Oxford, 1990.

