

## Modelling of carbohydrate–aromatic interactions: *ab initio* energetics and force field performance

Vojtěch Spiwok<sup>a,\*</sup>, Petra Lipovová<sup>a</sup>, Tereza Skálová<sup>b</sup>, Eva Vondráčková<sup>b</sup>,  
Jan Dohnálek<sup>b</sup>, Jindřich Hašek<sup>b</sup> & Blanka Králová<sup>a</sup>

<sup>a</sup>Department of Biochemistry, Institute of Chemical Technology in Prague, Technická 5, 166 28, Prague 6, Czech Republic; <sup>b</sup>Institute of Macromolecular Chemistry, Academy of Sciences of the Czech Republic, Heyrovského náměstí 2, 162 06, Praha 6, Czech Republic

Received 18 August 2005; accepted 12 December 2005  
© Springer 2006

**Key words:** *ab initio*, carbohydrate recognition, C–H/ $\pi$  interactions, force field, glycobiology, glycosidases, lectins

### Summary

Aromatic amino acid residues are often present in carbohydrate-binding sites of proteins. These binding sites are characterized by a placement of a carbohydrate moiety in a stacking orientation to an aromatic ring. This arrangement is an example of CH/ $\pi$  interactions. *Ab initio* interaction energies for 20 carbohydrate–aromatic complexes taken from 6 selected ultra-high resolution X-ray structures of glycosidases and carbohydrate-binding proteins were calculated. All interaction energies of a pyranose moiety with a side chain of an aromatic residue were calculated as attractive with interaction energy ranging from –2.8 to –12.3 kcal/mol as calculated at the MP2/6-311+G(d) level. Strong attractive interactions were observed for a wide range of orientations of carbohydrate and aromatic ring as present in selected X-ray structures. The most attractive interaction was associated with apparent combination of CH/ $\pi$  interactions and classical H-bonds. The failure of Hartree–Fock method (interaction energies from +1.0 to –6.9 kcal/mol) can be explained by a dispersion nature of a majority of the studied complexes. We also present a comparison of interaction energies calculated at the MP2 level with those calculated using molecular mechanics force fields (OPLS, GROMOS, CSFF/CHARMM, CHEAT/CHARMM, Glycam/AMBER, MM2 and MM3). For a majority of force fields there was a strong correlation with MP2 values. RMSD between MP2 and force field values were 1.0 for CSFF/CHARMM, 1.2 for Glycam/AMBER, 1.2 for GROMOS, 1.3 for MM3, 1.4 for MM2, 1.5 for OPLS and to 2.3 for CHEAT/CHARMM (in kcal/mol). These results show that molecular mechanics approximates interaction energies very well and support an application of molecular mechanics methods in the area of glycochemistry and glycobiology.

**Abbreviations:** AMBER – assisted model building with energy refinement; B3LYP – Becke–Slater-HF 3-term exchange and Lee–Yang–Parr correlation hybrid functional; BSSE – basis set superposition error; CBM – carbohydrate-binding module; CBS – complete basis set; CCSD(T) – coupled cluster with single, double and perturbative triple excitation; CHARMM – chemistry at Harvard molecular mechanics; CHEAT – carbohydrate hydroxyl groups represented by extended atoms; CSFF – carbohydrate solution force field; DFT – density functional theory; GROMOS – Groningen molecular simulation; HF – Hartree–Fock method; MM2 – molecular mechanics version 2; MM3 – molecular mechanics version 3;

\*To whom correspondence should be addressed. Fax: +420-220-445-167; E-mail: spiwokv@vscht.cz

MP2 – Møller–Plesset perturbation theory; second order; OPLS – optimized potentials for liquid simulations; PDB – protein data bank; RMSD – root mean square deviation.

## Introduction

Understanding quantum-mechanical nature and thermodynamic consequences of non-covalent bonds is necessary for a detailed insight into issues of molecular recognition, structure–activity relationship, supramolecular chemistry, crystal and solid phase chemistry, biomolecular structure and so forth [1]. Electrostatic and van der Waals interactions and classical hydrogen bonds are recognized as the main non-covalent interactions in biology. Terms describing these interactions are included in most molecular mechanics force fields or scoring functions (empirical free energy models). Beside these “canonical” interactions, there are also interactions as  $\pi$ – $\pi$  stacking,  $\pi$ -cation, CH/ $\pi$ , YH/ $\pi$  and CH/Y interactions (where Y is N, O, S or halogen) that are playing an important role in biology and supramolecular chemistry. The role of these “non-canonical” interactions is often omitted in biological studies. In molecular mechanics force fields these interactions are generally not explicitly described. Research on the role of non-covalent interactions in specific biological systems can therefore bridge a gap between theoretical chemistry and biological sciences. *Ab initio* quantum chemistry methods provide a possibility of calculation intra- and intermolecular forces. Coordinates of a complex and specification of a method/basis form the only input required for these methods. CH/ $\pi$  interactions are an example of weak hydrogen bonds. They have been studied for decades using a variety of experimental and theoretical techniques [2]. The study of Tamres showed that mixing of chloroform with benzene (or with some other aromatic compounds) is accompanied by release of heat [3]. Tamres provided evidence that an aromatic ring can act as an H-bond acceptor [3]. CH/ $\pi$  interactions are (together with CH/N, CH/O and CH/X interactions) also known as improper blue-shifting hydrogen bonds or anti-hydrogen bonds because they decrease C–H donor bond length contrary to “usual” H-bonds [4]. In a classical H-bond (X–H ... Y), the strength of X–H bond is reduced due to a charge density transfer from the lone pair

of Y (or  $\pi$  electrons) to the antibonding  $\sigma^*$  orbital of X–H. This leads to a weakening of X–H bond and an increase of its distance. In blue-shifting H-bonds, on the other hand, charge is transferred from the lone pair of Y (or  $\pi$  electrons) to a more remote part of the molecule containing X–H bond. As a secondary effect, position of the donor proton of X–H is changed and X–H distance is reduced [4]. CH/ $\pi$  interactions were also studied by database mining of crystallographic databases of proteins [5] and small molecules [6, 7]. The fact that CH/ $\pi$  interactions are intermolecular forces and not purely a result of the hydrophobic effect is illustrated by *ab initio* calculations of interaction energies. Interaction energies of CH/ $\pi$  interactions have been calculated by *ab initio* methods for a number of model inter-molecular complexes including methane–benzene [8–10], ethane–benzene [8], ethylene–benzene [8], acetylene–benzene [8], T-shaped benzene dimer [11–14] and chloroform–fluorobenzene [15] at the different levels of theory including second order Møller and Plesset theory [6, 8, 12, 15], configurational interaction [11], coupled cluster [8, 9, 13, 14] and symmetry adapted perturbation theory [9] methods. The interaction energy of methane–benzene complex was calculated as –1.45 kcal/mol as an estimated basis set limit at the CCSD(T) level [8]. Application of non-correlated methods (e.g. Hartree–Fock) generally fails to describe these interactions. It was shown that dispersion energy is dominant in the interaction of benzene with methane but the electrostatic term increases in a series of complexes with ethane, ethylene and acetylene [8].

Recent advances in glycomics – a research on a saccharide complement of genome and proteome illustrate an enormous importance of carbohydrate recognition in signal transduction, cell–cell interactions, pathogen entry to host cell, cancer metastasis, inflammation, fertility and development. CH/ $\pi$  interactions are assumed to play an important role in carbohydrate–protein interactions because most carbohydrate-processing enzymes and carbohydrate-binding proteins contain one or multiple aromatic amino acid residues in their binding sites [16, 17]. These complexes are

characterized by parallel orientation of aromatic ring and pyranose or furanose moiety with C–H bonds of a carbohydrate pointing onto the plane of an aromatic ring. The role of aromatic amino acid residues in binding of carbohydrate ligands was studied in a variety of model proteins by site-directed mutagenesis [18–20], covalent modification [18], and by calorimetric methods [20]. Mutation of such aromatic amino acid residue (to alanine or other non-aromatic residues) generally leads to a significant drop in affinity or activity towards a carbohydrate ligand/substrate [18–20]. A recent calorimetric study showed decreased enthalpy of binding of carbohydrate to protein mutants lacking the aromatic residue [20]. Interaction of a carbohydrate with an aromatic ring is also proposed to be important in design of artificial carbohydrate receptors as many successful carbohydrate receptors employ aromatic systems [21]. Aromatic–carbohydrate interactions may also play an important role in solubilization of carbon-based nanomaterials by saccharides, drug solubilization by cyclodextrines, cellulose–lignin assembly, adhesion of pathogens to medical devices, in chromatography of saccharides on porous graphite columns and in other interactions.

Since 2004, carbohydrate–aromatic interactions are a subject of *ab initio* quantum chemistry studies [22–25]. In our study [22] we focused on carbohydrate–aromatic interactions in  $\beta$ -galactosidase from *E. coli*. Interaction energy for the most favourable complex (glucose moiety of allolactose in a shallow binding mode) was calculated as  $-5.2$  kcal/mol at the MP2/6-31+G(d) level. Interaction energies calculated on HF/6-31+G(d) and B3LYP/6-31+G(d) were small or repulsive. Also the profile of interaction energy as a function of intermolecular distance was presented in this study [22]. Two studies of Sujatha et al. use the density functional theory method with B3LYP functional [23] and MP2 [24] to study carbohydrate–aromatic pairs taken from X-ray structures of carbohydrate–protein complexes. Interaction energies at the MP2/6-311++G(d,p) level for model carbohydrate–aromatic complexes were in the range of  $-3.2$  to  $-8.2$  kcal/mol [24]. Authors also addressed the issue of specificity of studied proteins towards galactose vs. glucose. The study of Fernández-Alonso et al. [25] performed *ab initio* geometry optimization with BSSE correction at the MP2/6-31G(d,p) level for fucose–benzene

complex derived from the structure of galactose-binding lectin. Interaction energy of this complex was calculated as  $-3.0$  kcal/mol. Potential surface of this complex was extensively studied [25].

In recent years force field developers focused on carbohydrates, owing to success of several carbohydrate-based drugs and development of a novel discipline – glycomics. Carbohydrate-tuned versions were developed [26–29] and evaluated [30, 31] for several well established biomolecular force fields in recent years. This attempt focused mainly on dihedral angle parameters as these are recognized as a main reason of failure of general purpose (i.e. not carbohydrate tuned) force fields in carbohydrate modelling. However, carbohydrate–aromatic interactions represent another possible weak point and therefore we decided to evaluate performance of current force fields in modelling of carbohydrate–aromatic interactions. Molecular modelling of carbohydrate–protein interactions using molecular dynamics simulation or protein–ligand docking seems to be more difficult in comparison with modelling of interactions of proteins with other types of ligands, probably due to a high importance of hydrogen bonds and competition between ligand and water [32], presence of water–mediated interactions [33], pseudo-symmetrical distribution of H-bond donors/acceptors in the ligand [34] and possibly also due to inadequate modelling of carbohydrate–aromatic interactions. Addressing the issue of carbohydrate–aromatic interactions is necessary for accurate modelling of protein–ligand interaction and is critical for a performance of design of carbohydrate-like pharmaceuticals. Based on the results of our study [22] we decided to focus on the role of carbohydrate–aromatic interactions in a wider range of carbohydrate–protein complexes using the second order Møller and Plesset perturbation theory [35]. Another goal was to test the predictive power of the force field approach on carbohydrate–protein interactions.

## Methods

### *Model structures*

Experimentally determined X-ray structures of carbohydrate–protein complexes were selected

using the database-mining tool GlyVicinity [36]. Six structures with the highest resolution (1.3 Å or better) were selected. The selected structures are listed in Table 1 and further described in the Results and discussion section. Figures illustrating individual complexes can be obtained as a supplementary material. To make *ab initio* calculations feasible it was necessary to reduce the size of the system. Aromatic amino acid residues were represented as *p*-cresol (1-hydroxy-4-methylbenzene) and 3-methylindole for tyrosine and tryptophan, respectively. Carbohydrates were represented as a single pyranose moiety. For thioglycoside ligands, atom types of sulphur were changed to oxygen. Hydrogen atoms were added to each complex using

OpenBabel 1.100.0 [openbabel.sf.net] with standard distances. Then the structures were *ab initio* minimized at the Hartree–Fock level (HF/MINI) with rational function optimizer as implemented in Gamess US package (version 14 Jan. 2003 R2) [43]. The minimization was terminated when the largest component of the energy gradient was less than 1/1000 Hartree/Bohr, and the root-mean-square gradient less than 1/3000 Bohr within the same step. A more thorough minimization would be required to obtain optimal conformations of complexes. On the other hand, a thorough minimization at the HF or DFT level is likely to lead to a significant separation of interacting molecules because non-correlated and DFT methods usually

Table 1. List of the studied complexes with results of calculation of *ab initio* interaction energies.

Carbohydrate–protein complex	PDB ID	Resolution (Å)	Ref.	Carbohydrate residue	Aromatic residue	Donor bond	Geometric parameters			Interaction energy (kcal/mol)	
							<i>d</i> (Å)	<i>r</i> (Å)	$\alpha$ (°)	HF	MP2
<i>Clostridium thermocellum</i> endoglucanase A (mutant E59Q)	1KWF	0.94	[37]	Glc401(B)	W205(A)	C5–H	3.09	3.01	77	–1.2	–3.8
				Glc402(B)	W132(A)	C3–H	3.23	3.22	90	–1.3	–4.4
						(C5–H)	(3.35)	(3.33)	(84)		
				Glc404(B)	Y372(A)	C1–H	2.96	2.96	90	–0.9	–3.9
				Glc404(C)	Y372(A)	C1–H	3.03	2.98	80	–0.6	–3.4
				Glc405(B)	Y277(A)	C2–H	2.85	2.84	87	0.7	–3.6
				Glc405(C)	Y277(A)	C2–H	3.03	2.97	78	0.4	–3.7
Concanavalin A	1I3H	1.20	[38]	Glc406(C)	Y369A(A)	C2–H	3.33	3.01	65	–2.7	–6.0
				Glc406(C)	Y369B(A)	C2–H	3.21	2.71	58	–6.9	–12.3
				Man101(B)	Y12(A)	both	6.33	3.32	32	–0.6	–2.8
						C6–H	(6.41)	(3.32)	(31)		
						(C3–H)	(3.45)	(3.11)	(64)		
Cholera toxin B	3CHB	1.25	[39]	Gal104(D)	W88(D)	C5–H	3.47	3.14	65	–2.7	–6.4
<i>Streptomyces lividans</i> xylan binding domain Cbm13	1KNM	1.20	[40]	Lat132	W34(A)	C4–H	3.20	2.93	66	–1.6	–5.6
				Lat133	Y117(A)	C4–H	3.20	2.98	69	–0.6	–3.4
				Lat133 <sup>1</sup>	Y74(A) <sup>1</sup>	C4–H	3.21	3.21	90	–0.7	–3.5
<i>Humicola insolens</i> cellobiohydrolase Cel6A (mutant D405N)	1OC7	1.11	[41]	Sgc602(A)	W371(A)	C5–H	2.76	2.61	71	0.2	–4.9
				Sgc603(A)	W274(A)	C3–H	2.89	2.90	90	–0.8	–5.2
				Ma3605(A)	W277(A)	C5–H	3.46	3.28	71	1.0	–3.3
<i>Piromyces equi</i> carbohydrate binding module (CBM29-2)	1GWM	1.15	[42]	Bgc1156(A) <sup>2</sup>	Y46(A)	C5–H	2.77	2.77	87	–0.2	–4.0
				Bgc1156(A) <sup>3</sup>	Y46	C5–H	3.23	3.02	69	–0.5	–3.2
				Bgc1158(A)	W26(A)	C5–H	3.17	3.12	80	–1.1	–3.7
				Bgc1160(A)	W24(A)	C5–H	3.24	3.07	71	–1.1	–4.3

Chain notation is given for each residue as a character in parentheses. Donor C–H bonds were assigned based on visual inspection. The second C–H bond and corresponding geometrical parameters are given in parentheses if two C–H bonds are involved in interaction. Parameter *d* is distance between the donor hydrogen atom and the plane of the aromatic ring measured in the direction of the donor C–H bond. Parameter *r* is a perpendicular distance between the donor hydrogen atom and the plane of the aromatic ring. Parameter  $\alpha$  is an angle between the donor C–H bond and the plane of the aromatic ring. Interaction energies were calculated in 6-311 + G(d) basis set. 1 – interactions with a crystallographically related molecule of protein, 2 – the nearer position of Bgc1156, 3 – the farther position of Bgc1156.

fail to model CH/ $\pi$  interactions and so these are calculated as repulsive. Therefore we decided to use this minimization procedure with a relatively high convergence criteria (i.e. high value of energy root-mean-square gradient when the optimization is terminated), which was shown to be suitable for fixing geometries of covalent bonds without disturbing the overall geometry of a complex [22].

#### Calculation of *ab initio* interaction energies

Interaction energies were calculated using supra-molecular method (as a difference of *ab initio* energies of the whole system and both subsystems) for minimized complexes. Calculations were performed at the HF/6-311+G(d) and MP2/6-311+G(d) level using Gamess US package [43]. Basis sets implemented in the program were used. BSSE was corrected using the standard counterpoise correction [44]. Calculation of geometrical parameters and structure alignments were performed in a standard spreadsheet editor. Profile of interaction energy as a function of distance was taken from the reference 22. Briefly, complex of Trp999 from *E. coli*  $\beta$ -galactosidase with glucose (a moiety of allolactose) was minimized at the HF/MINI level as described above. Then the indole ring of tyrosine was shifted towards and outwards the glucose moiety in the direction of the putative donor C–H bond (C-6 of glucose) as illustrated in Figure 1. Interaction energy on HF/6-31+G(d) and MP2/6-31+G(d) was calculated for five distances (2.31, 2.71, 3.11, 3.51 and 3.91 Å measured from the donor hydrogen atoms to the aromatic ring in the direction of the donor C–H bond).

#### Calculation of force field interaction energies

Force field calculations of interaction energies were performed in a standard spreadsheet editor for *ab initio* minimized geometries using published force field parameters (except MM2 and MM3 force fields). Original publications cited in Table 2 and parameter files of Tinker package [51] were used as a source of force field parameters. MM2 and MM3 interaction energies were calculated using Tinker package. Lone pair dummy atoms in MM2 force field were added by geometry optimization with position constraints applied on all real atoms while lone pair dummy atoms were free to move.

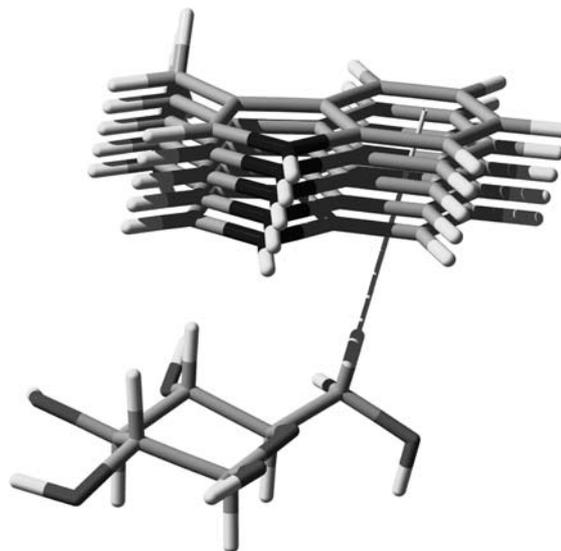


Figure 1. Complex of Trp999 from *E. coli*  $\beta$ -galactosidase with glucose (a moiety of allolactose) used for modelling of profile of interaction energy as a function of distance. Indole ring of tryptophan was shifted in a direction of C6–H bond (shown as a thin stick).

Relative dielectric constant was set to 1.5 for MM2 and MM3 force fields (recommended value) and 1.0 for all other force fields. All images were generated using SwissPDB viewer [52] together with PoVRay [www.povray.com].

## Results and discussion

#### Selection of model carbohydrate–aromatic complexes

Six X-ray structures of carbohydrate-binding proteins with the highest available structure resolution were selected using GlyVicinity tool [36]. These structures include examples of various biological functions related to carbohydrate recognition. There are two examples of glycosidases (*Clostridium thermocellum* endoglucanase A and *Humicola insolens* cellobiohydrolase Cel6A), two examples of carbohydrate binding modules (*Streptomyces lividans* xylan binding domain CBM13 and *Piromyces equi* carbohydrate binding module CBM29-2), one lectin (Concanavalin A) and one carbohydrate-binding bacterial toxin (cholera toxin). Recognized glycopyranose moieties include  $\beta$ -D-glucopyranoses,  $\beta$ -D-galactopyranoses and

Table 2. Performance of tested force fields.

Carbohydrate force field	Protein force field	With the circled complex				Without the circled complex			
		<i>A</i>	<i>B</i>	<i>R</i> <sup>2</sup>	RMS	<i>A</i>	<i>B</i>	<i>R</i> <sup>2</sup>	RMS
OPLS-AA [26]	OPLS-AA [45]	1.02	-1.25	0.94	1.5	1.04	-1.18	0.78	1.5
GROMOS 43A1 [46]	GROMOS 43A1 [46]	0.71	-1.80	0.68	1.2	0.60	-2.25	0.25	1.2
CSFF [27]	CHARMM c31b1 [47]	0.78	-1.76	0.89	1.0	0.85	-1.49	0.69	1.0
CHEAT [28]	CHARMM c31b1 [47]	0.01	-3.92	0.00	2.3	0.24	-2.98	0.06	1.2
Glycam04 [29]	AMBER [48]	0.82	-1.81	0.87	1.2	0.92	-1.42	0.67	1.3
MM2 [49]	MM2 [49]	0.76	-1.60	0.62	1.4	0.73	-1.73	0.25	1.4
MM3 [50]	MM3 [50]	0.47	-1.71	0.84	1.3	0.64	-1.02	0.73	0.7

Force field energy as a function of *ab initio* energy was fitted with equation  $E(\text{force field}) = A \cdot E(\text{MP2}) + B$ . Values of *A* and *B* as well as correlation coefficient and average absolute value of a difference between force field and MP2 energy are listed. Values of *B* and absolute value of a difference between force field and MP2 energy are in kcal/mol. Analysis with and without the complex Tyr369B-Glc406 from *Clostridium thermocellum* endoglucanase A (circled in Figure 4) was performed.

$\alpha$ -D-mannopyranose. These complexes represent a wide range of carbohydrate-protein affinities. While CBMs are rather weak and non-specific carbohydrate binders, interaction of cholera toxin with its ligand is the interaction with one of highest known affinity among carbohydrate-protein interactions ( $\Delta G = -10$  kcal/mol for the pentasaccharide fragment of GM1 ganglioside from calorimetric measurements) [53]. These structures contain 20 carbohydrate-aromatic complexes including those where alternative conformation/position of a ligand or a side chain was resolved in corresponding crystal structures. Proteins as well as complexes are listed in Table 1.

#### Optimized geometries of carbohydrate-aromatic complexes

The first stage of calculation of interaction energies was a geometrical optimization of studied complexes using an *ab initio* energy minimization. The fact that CH/ $\pi$  interactions are generally underestimated using HF and most DFT techniques or even interpreted as repulsive would likely lead to a significant separation of both interacting molecules in thorough geometry optimization at the HF or DFT level. A profile of interaction energy as a function of distance of sugar and aromatic residue published in reference [22] showed that the optimal distance calculated at the HF level was approximately 0.4 Å longer than that calculated at the MP2 level. Therefore we decided to use a rapid minimization with a relatively high convergence criteria. The aim of

this procedure is to fix geometries of covalent bonds without disturbing an overall geometry of a complex. Differences between initial and geometry optimized structures of studied complexes (described as RMSD values for non-hydrogen atoms) were within the range of 0.15–0.55 Å. This shows that the minimized complexes represent a good approximation of real conformations and that the applied *ab initio* minimization procedure provides a good accuracy with an efficient use of computer resources.

Geometry of studied carbohydrate-aromatic complexes is illustrated in Figure 2 generated by alignment of carbon atoms of hexopyranose moieties of the *ab initio* minimized structures of the complexes. In order to illustrate geometries of studied complexes, C–H bonds which are expected to contribute most to interactions were assigned for each complex by visual inspection and corresponding geometric parameters were calculated (listed in Table 1.). Beside these putative CH/ $\pi$  donors also other C–H bonds can contribute to interactions. For each carbon atom of hexopyranose moiety (C1 to C6) at least one example of interaction of the corresponding C–H bond with an aromatic residue was observed among studied complexes. Among the studied complexes, there is a wide range of angles between a donor C–H bond and the plane of an aromatic ring ranging from 31° to nearly perpendicular. Examples of C–H bonds pointing nearly into the centre of an aromatic ring as well as those pointing onto the aromatic C–C bond or outside the ring were observed. Perpendicular (i.e. shortest) distances between the hydrogen

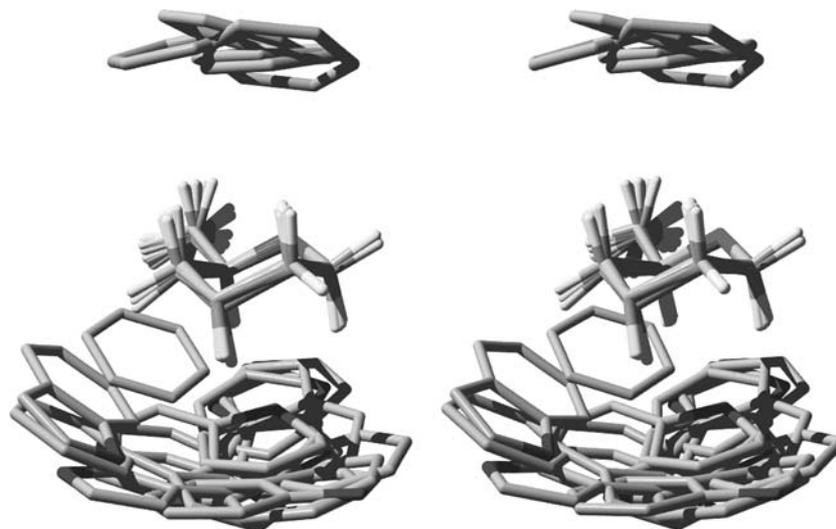


Figure 2. Stereo-view illustrating geometry of minimized complexes. The image was generated by superposition of carbon atoms of the carbohydrate moieties. Hydroxyl groups of carbohydrates as well as  $C\beta$ , oxygen and hydrogen atoms of aromatic residues are not shown for simplicity.

atom of a donor C–H bond and the plane of an aromatic ring were in the range of 2.61–3.33 Å.

#### Interaction energies

Figure 2 illustrates that selected complexes represent a wide range of orientations of an aromatic system towards a carbohydrate moiety. All interactions were calculated as attractive at the MP2/6-311+G(d) level with interaction energies in the range of –2.8 to –12.3 kcal/mol. *Clostridium thermocelum* endoglucanase A is an inverting glycosidase, which forms a part of cellulosome – an extracellular multi-enzymatic cellulose-degrading complex of this thermophilic bacterium. A high-resolution crystal structure of a complex of E95Q inactive mutant with cellopentaose is available (PDB ID 1KWF, resolution 0.94 Å) [37]. The active site is a long cavity composed of six sub-sites each binding a single glucose moiety denoted –3 to +3 according to the location of the cleaved glycosyl bond (carbohydrate residues Glc401 to Glc406, respectively as denoted in the PDB record). In sites +1 and +2 the electron density map reveals two distinct positions of glucose moieties (chain B and C of PDB record) and *ab initio* minimizations and interaction energy calculations were performed for both alternative positions. There are two tryptophan residues (Trp205 and Trp132) interacting with glucose moieties –3

and –2, respectively. Glucose moieties in sub-sites +1, +2 and +3 interact with Tyr372, Tyr277 and Tyr369, respectively. Interaction energies of complexes associated with Tyr372 and Tyr277 calculated at the MP2/6-311+G(d) level were in the range of –3.4 to –3.9 kcal/mol. The most attractive interactions were calculated for complexes of Tyr369 and Glc406. There are two distinct positions of the side-chain of the residue Tyr369 as revealed from electron density maps (denoted Tyr369A and Tyr369B) differing in the  $N-C_\alpha-C_\beta-C_\gamma$  dihedral angles [37]. The *ab initio* energy minimization and the calculation of interaction energy were performed for both positions. Interaction energies at the MP2/6-311+G(d) level were –6.0 and –12.3 kcal/mol for the complexes Tyr369A–Glc406 and Tyr369B–Glc406, respectively. These high values can be explained by an important role of classical H-bonds. There is a classical H-bond between O6–H bond of the glucose and the phenol group of the tyrosine in the complex Tyr369A–Glc406 as apparent by visual inspection. In the complex Tyr369B–Glc406 there are two apparent H-bonds (between O1–H bond of the glucose and the phenol group of tyrosine and between the phenol group of tyrosine and O5 atom of glucose). Also high interaction energies calculated by non-correlated method at the HF/6-311+G(d) level (–2.7 and –6.9 kcal/mol for Tyr369A–Glc406 and Tyr369B–Glc406,

respectively) indicate a significant contribution of classical hydrogen bonds. Interaction energies of other complexes calculated on HF/6-311+G(d) were in the range of  $-1.3$  to  $0.7$  kcal/mol. The Tyr369–Glc406 complex shows that CH/ $\pi$  interactions can be present simultaneously with classical H-bonds.

Concanavalin A is one of the most widely used and well characterized lectins. It possesses a high affinity towards  $\alpha$ -linked mannose. A structure of the complex of Concanavalin A with  $\alpha$ 1-2 manno-  
biose is available at 1.2 Å resolution (PDB ID 113H) [38]. The binding site of Concanavalin A is relatively shallow. The aromatic residue Tyr12 interacts with both hydrogen atoms on the C6 carbon of the non-reducing moiety of manno-  
biose. Contribution of these two hydrogen atoms seems to be approximately the same (perpendicular distances between each hydrogen atom and the plane of the aromatic ring is nearly the same, see Figure 3c.). The interaction energy was calculated at the MP2/6-311+G(d) level as  $-2.8$  kcal/mol. This was the least attractive interaction among the studied complexes. This indicates that CH/ $\pi$  interactions in this “fork” orientation with equivalent contributions of two C–H bonds on a single carbon atom are likely to be less favourable compared to a single C–H bond pointing onto an aromatic ring.

Cholera toxin is an extra-cellular assembly of A and B subunits ( $AB_5$  quaternary structure) produced by *Vibrio cholera*. The subunit A is responsible for activation of G-protein inside the target cell of a host while the subunit B is responsible for toxin targeting. Subunit B recognizes a non-reducing terminal  $\beta$ -galactopyranosyl moiety of  $G_{M1}$ -ganglioside of lipid rafts. Recognition of  $G_{M1}$ -ganglioside promotes endocytosis. The structure of cholera toxin B pentamer in complex with oligosaccharide fragment of  $G_{M1}$ -ganglioside is available (PDB ID 3CHB, resolution 1.25 Å) [39]. As cholera toxin B is a pentamer in the crystal structure only the subunit D was selected for this study. Its binding site is relatively large in order to accommodate the major part of the  $G_{M1}$  penta-  
saccharide. The galactopyranosyl moiety of the ligand interacts with Trp88. The axis of the indole ring of tryptophan is aligned with the C3–C4–C5–C6 chain of galactose. Therefore corresponding hydrogen atoms point towards the indole ring of tryptophan. C5–H and C3–H bonds probably contribute most to CH/ $\pi$  interaction. The interaction energy of this complex calculated on MP2/6-311+G(d) was the second most attractive of all studied complexes ( $-6.4$  kcal/mol). This strong interaction could be explained by involvement of multiple C–H bonds in the CH/ $\pi$  interactions. The fact that complex of W88(D)–Gal104(D) from

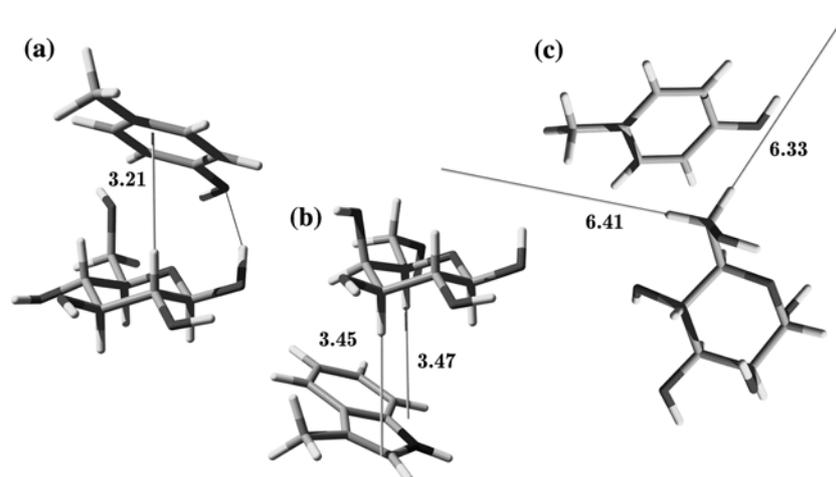


Figure 3. Three illustrative carbohydrate-aromatic complexes: (a) The complex with the most stabilizing interaction energy (Y369B(A)–Glc406(C) of *Clostridium thermocellum* endoglucanase A), (b) the complex with the second most stabilizing interaction energy (complex W88(D)–Gal104(D) of Cholera toxin B) and (c) the complex with the lowest stabilizing interaction energy (complex Y12(A)–Man101(B) of Concanavalin A) among studied complexes. Putative CH/ $\pi$  donor bonds were prolonged to their intersections with a plane of an aromatic ring (thin lines, distances are illustrated). Note that complex Y369B(A)–Glc406(C) of *Clostridium thermocellum* endoglucanase A characterized by a presence of classical H-bond (thin line in a). Structures of other complexes can be obtained as a supplementary material.

Cholera toxin B possesses the second strongest interaction without any apparent classical H-bonds coincides with very high affinity of cholera toxin with its ligands [53].

Carbohydrate binding modules are often present as domains of polysaccharide processing enzymes and enzymatic complexes. These non-catalytic domains help targeting of catalytic enzymes to substrate. Family 13 CBM is a part of *Streptomyces lividans* family 10A xylanase (for nomenclature of CBMs and glycosidases see references [54, 55]). A crystal structure of this CBM is known for an unliganded state (PDB IDs 1KNL) and for complexes with lactose and xylopentaose (PDB IDs 1KNM and 1KNN, respectively) [40]. The structure of the complex with lactose (1KNM, resolution 1.2 Å) was used in this study. This CBM has  $\beta$ -trefoil fold with three independent binding sites ( $\alpha$ ,  $\beta$  and  $\gamma$ ) on each "leaf". Two binding sites ( $\alpha$  and  $\beta$ ) are occupied by lactose molecules and one binding site is empty ( $\gamma$ ) in the PDB file. One lactose molecule Lat133 binds simultaneously to the site  $\beta$  as found in the PDB file and to the site  $\gamma$  of the neighbour crystallographically related protein molecule. The complex with Lat133 in site  $\gamma$  was therefore generated by application of a symmetry operator as described in the PDB file. In the binding site  $\alpha$  the non-reducing end of lactose interacts with Trp34. The geometry of this interaction is similar to that of cholera toxin B – galactose complex and this interaction is also one of the most attractive (–5.6 kcal/mol at the MP2/6-311 + G(d) level). In the binding site  $\beta$  the non-reducing end (galactose moiety) of lactose interacts with Tyr117. In this interaction the C4–H bond is apparently the main CH/ $\pi$  donor. Interaction energy at the MP2/6-311 + G(d) level is –3.4 kcal/mol. The glucose moiety in the  $\gamma$  site interacts with Tyr74 and energy of this interaction at the MP2/6-311 + G(d) level is –3.5 kcal/mol.

Cellobiohydrolase in another important part of cellulose-degrading systems. A crystal structure of *Humicola insolens* cellobiohydrolase Cel6A is available for unliganded state as well as for complexes with thioglycosidic-bond-containing oligosaccharide mimetics [41]. The structure of the D405N mutant complexed with methyl–4, 4<sup>II</sup>, 4<sup>III</sup>, 4<sup>IV</sup>-tetrathio- $\alpha$ -cellopentaoside was selected (PDB ID 1OC7, resolution 1.10 Å) for this study. The binding site in the form of tunnel-like cavity hosts six monosaccharide moieties. These sub-sites

are denoted as –2 to +4 according to the site of cleavage of glycosidic bond with four aromatic residues involved in interaction. Interaction energies were calculated for complexes of Trp371–Sgc602, Trp274–Sgc603 and Trp277–Ma3605. The structures of complexes were modified for calculations by replacement of sulphur atoms of thioglycosides by oxygen followed by an *ab initio* minimization. As C–S bonds in thioglycosides are longer than C–O bonds, resulting structures of the *ab initio* minimization were carefully evaluated. The used minimization procedure turned out to be suitable for bond lengths fixing and no significant disturbance of overall geometry was observed. The RMSD values between initial and minimized positions of non-hydrogen atoms of carbohydrate moieties were in the range of 0.21–0.36 Å. Interaction energies calculated at the MP2/6-311 + G(d) level were –4.9, –5.2 and –3.3 kcal/mol for complexes Trp371–Sgc602, Trp274–Sgc603 and Trp277–Ma3605, respectively.

The carbohydrate binding module CBM29-2 from *Piromyces equi* is another non-catalytic carbohydrate binding module which targets a microbial cellulose-degrading enzyme to its substrate (PDB ID 1GWM, resolution 1.15 Å) [42]. This protein binds a variety of manno- and glucooligosaccharides with low substrate specificity. The binding site of this protein is a shallow cavity with three aromatic amino acid residues (Trp24, Trp26 and Tyr46). Electron density map of residue Bgc1156 (as denoted in the PDB file) found in vicinity of residue Tyr46 revealed two distinct binding sites of glucose (denoted as the nearer and the farther) and *ab initio* energy minimization and interaction energy calculation were performed for both of these positions. Energies of interaction of Trp24 and Trp26 with corresponding glucose moieties were –4.3 and –3.7 kcal/mol, respectively, as calculated at the MP2/6-311 + G(d) level. Interaction energies of complexes Tyr46–Bgc1156 were –4.0 and –3.2 kcal/mol in the nearer and the farther mode, respectively, as calculated on MP2/6-311 + G(d).

Relatively high interaction energies of studied complexes (–2.8 to –12.3 kcal/mol) indicate that CH/ $\pi$  interactions play an important role in carbohydrate recognition. In order to estimate a contribution of CH/ $\pi$  interactions in stabilization of carbohydrate–aromatic complexes, interaction energy was calculated for the complex Tyr277–

Glc405(chainB) from *Clostridium thermocelum* endoglucanase A in which aromatic side chain was replaced by a non-aromatic moiety (cyclopentane). Cyclopentane was placed in parallel to the benzene ring of tyrosine so that their centres of mass were aligned. Geometry optimization and interaction energy calculation was performed as described in Methods. Resulting interaction energy at the MP2/6-311+G(d) level (−1.9 kcal/mol) was approx. half of that of the original complex (−3.6 kcal/mol).

The fact that interaction energies (−2.8 to −12.3 kcal/mol) were more attractive than those of methane–benzene complex [8] could be explained by involvement of multiple C–H bonds of a carbohydrate in the interaction. A visual inspection of the weakest complex from Concanavalin A shows that this complex possesses the weakest CH/ $\pi$  contact. This also explains the fact that interaction energies of tryptophan complexes tend to be more attractive than those of tyrosine complexes because the  $\pi$ -rich surface of tryptophan is larger. More attractive HF and MP2 energies in the complexes of Tyr369–Glc406 of 1KWF correspond to the fact that classical H-bonds are formed in these complexes. Directionality of CH/ $\pi$  interactions on the other hand does not seem to be as important as the number of involved C–H bonds. Relatively high interaction energies were observed for complexes with nearly perpendicular (e.g. Trp274–Sgc603 of 1OC7) as well as non-perpendicular (e.g. Trp34–Lat132 of 1KNM) orientation of a C–H bond towards an aromatic ring. Interactions of non-perpendicular geometry are characteristic for a galactose moiety (C4–H bond is equatorial). Also attractive interaction energies were observed for complexes where a C–H bond was pointing at the centre of an aromatic ring (e.g. Tyr274–Sgc603 of 1OC7) as well as in the case where it pointed nearly at the aromatic C–C bond (e.g. Tyr277–Glc405 of 1KWF). This finding corresponds with the previous database-mining and *ab initio* study of general CH/ $\pi$  interactions [6] as well as with results of studies of aromatic–carbohydrate interactions [22–25] which showed little directionality of CH/ $\pi$  interactions. Our view of carbohydrate–aromatic complexes can be concluded as unexpectedly attractive while the carbohydrate can “slide” along the aromatic ring. Another explanation of high interaction energies can be the fact that electronegative oxygen atom is

bound to a carbon of each C–H bond involved in the interaction. Interaction of chloroform with fluorobenzene was stronger than that of methane–benzene complex [15] due to the effect of chlorine atoms of chloroform.

We have to comment on application of density functional theory methods in calculation of interaction energies of carbohydrate–protein complexes. In quantum organic chemistry these methods represent an alternative to highly correlated calculations for their satisfactory accuracy at high speed and with favourable scaling. On the other hand, it seems that application of DFT within the standard Kohn–Sham theory is not capable to calculate interaction energies of complexes with high contribution of dispersion forces. Most of 25 density functionals that were compared in the study of Johnson et al. evaluated the interaction in the parallel benzene dimer as repulsive [56]. The interaction in the T-shaped benzene dimer (with CH/ $\pi$  interaction) was calculated as attractive using most functionals but majority of them significantly underestimated the interaction energy [56]. For example with the B3LYP/6-31G(d,p) functional the interaction energy was calculated as −0.63 kcal/mol [56] whereas the value achieved with CCSD(T)/CBS is −1.45 kcal/mol [8]. The Perdew–Wang 91 functional is probably the only commonly used functional which approximates relatively well a profile of energy as a function of distance for complexes of dispersion nature. Also other studies on carbohydrate–aromatic complexes [22–25] revealed that B3LYP interaction energies were low or repulsive in contrast to MP2 energies. Application of functionals with empirical correction terms [57] seems to be a viable alternative.

On the other hand, MP2 is known to slightly overestimate attractive interactions in complexes of dispersion nature. Comparison of interaction energies of different methane–benzene complexes at the MP2/CBS and CCSD(T)/CBS levels (CBS stands for complete basis set) showed that MP2 energies were calculated as more attractive by 0.18–0.33 kcal/mol (19–25%) [8]. Therefore we can expect that MP2 interaction energies presented in this study are also overestimated by approximately 20%. Detailed elucidation of this would

require an application of a higher level of theory (e.g. coupled cluster).

*Role of carbohydrate–aromatic interactions in carbohydrate recognition*

The results of this study show that carbohydrate–aromatic interactions are relatively strong attractive interactions of quantum-mechanical origin (i.e. not purely hydrophobic). Additionally, there could be also a contribution of hydrophobic interactions of statistical-thermodynamical origin. As most protein–ligand interactions are experimentally characterized in terms of thermodynamic free energy rather than interaction energy the question arises how important the entropic contribution is in carbohydrate–aromatic interactions. Carbohydrates interact with aromatic amino acid residues in proteins generally by axial C–H bonds while equatorial O–H groups can freely form classical H-bonds with other amino acids or with protein-bound water molecules. A hydrophobic interaction between the hydrophobic surface of an aromatic ring and the hydrophobic axial face of a carbohydrate is likely to stabilize carbohydrate–protein interaction. While classical H-bonds and CH/ $\pi$  interactions are both relatively strong the main difference is that classical H-bond donors/acceptors of protein/ligand compete with water molecules upon complex formation. Formation of a classical H-bond does not necessarily have to stabilize the protein–ligand complex. Such competition with water is likely to be weaker in case of carbohydrate–aromatic interactions. Elucidation of free energy nature of carbohydrate–aromatic interaction would also require a detailed characterization of aromatic–water (eg. OH/ $\pi$ ) and carbohydrate–water (eg. OH/O and CH/O) interactions as well as a characterization of hydrophobic effect. Hydrophobic contribution to carbohydrate–arene interaction was addressed by Morales and Penadés [58]. Conformation of cyclic glycopeptide (a cyclic molecule formed by two maltose molecules covalently connected by two aromatic linkers) was studied in water and in non-polar environment by NMR. Carbohydrate–aromatic interactions were dominant in water but were disturbed in a non-polar environment. Moreover, aromatic amino acid residues can play an important role in formation of the shape of a binding site with important thermodynamic contributions. We

can therefore expect that formation of aromatic–carbohydrate complexes can be enhanced by a hydrophobic effect.

*Comparison of ab initio and force field energies*

As a majority of applications of carbohydrate modelling is based on molecular mechanics force fields, it became commonplace to test a predictive power of currently used force fields. Interaction energies were calculated using the tested force fields and compared with *ab initio* values. Similarly to *ab initio* calculations, interaction energies were calculated so that a sum of energies of both subsystems is subtracted from energy of the whole complex. Interaction energies are calculated for a fixed conformation (single point) and thus covalent interaction terms (bonds, angles and torsions) do not play any role. Non-covalent interactions are treated by Coulombic and Lennard–Jones terms in OPLS-AA, CSFF/CHARMM, CHEAT/CHARMM, GROMOS and Glycam/AMBER force fields. These force fields contain explicit hydrogen atoms except for Gromos and CHEAT in which C–H bonds and O–H bonds, respectively, are modelled as united atoms. Partial charges on hydrogen atoms of C–H bonds of carbohydrates were ranging from zero (Glycam) to 0.100 (C-1 in OPLS). In MM2 force field, electrostatic interactions are modelled as dipole–dipole interactions (with lone-pair dummy atoms). Van der Waals interactions are modelled using a Buckingham potential. In MM3 force field, electrostatic interactions are treated either as charge–charge or charge–dipole interactions. Moreover, a special H-bond potential is added in MM3. For references of tested force fields see Table 2.

As the complex of Tyr369B–Glc406 from *Clostridium thermocellum* endoglucanase A is apparently dominated by classical H-bonds, correlation between force field and *ab initio* values was tested with and without inclusion of this complex (circled in Figure 4). Regression without this point gives an insight into a performance of tested force fields to model carbohydrate–aromatic interactions, while regression with inclusion of this point illustrates whether these interactions are over/underestimated if compared with classical H-bonds. Results of regression of force field energies vs. MP2 energies are illustrated in Figure 4. Corresponding parameters of linear regression

of force field energies as a function of MP2 energies are listed in Table 2.

Carbohydrate–aromatic interactions were most accurately modelled in all-atom force fields containing Coulombic and Lennard–Jones noncovalent terms (OPLS-AA, CSFF/CHARMM and Glycam/AMBER). The highest correlation was found for OPLS-AA force field ( $R^2=0.94$  and

$R^2=0.78$ , with and without the complex circled in Figure 4, respectively). Energy decomposition of interaction energies calculated using these three force fields revealed that they possessed approximately the same values of Lennard–Jones energies (generally  $-5$  to  $-3$  kcal/mol for OPLS-AA) while differing mainly in electrostatic energies. In the other words, there was a strong correlation

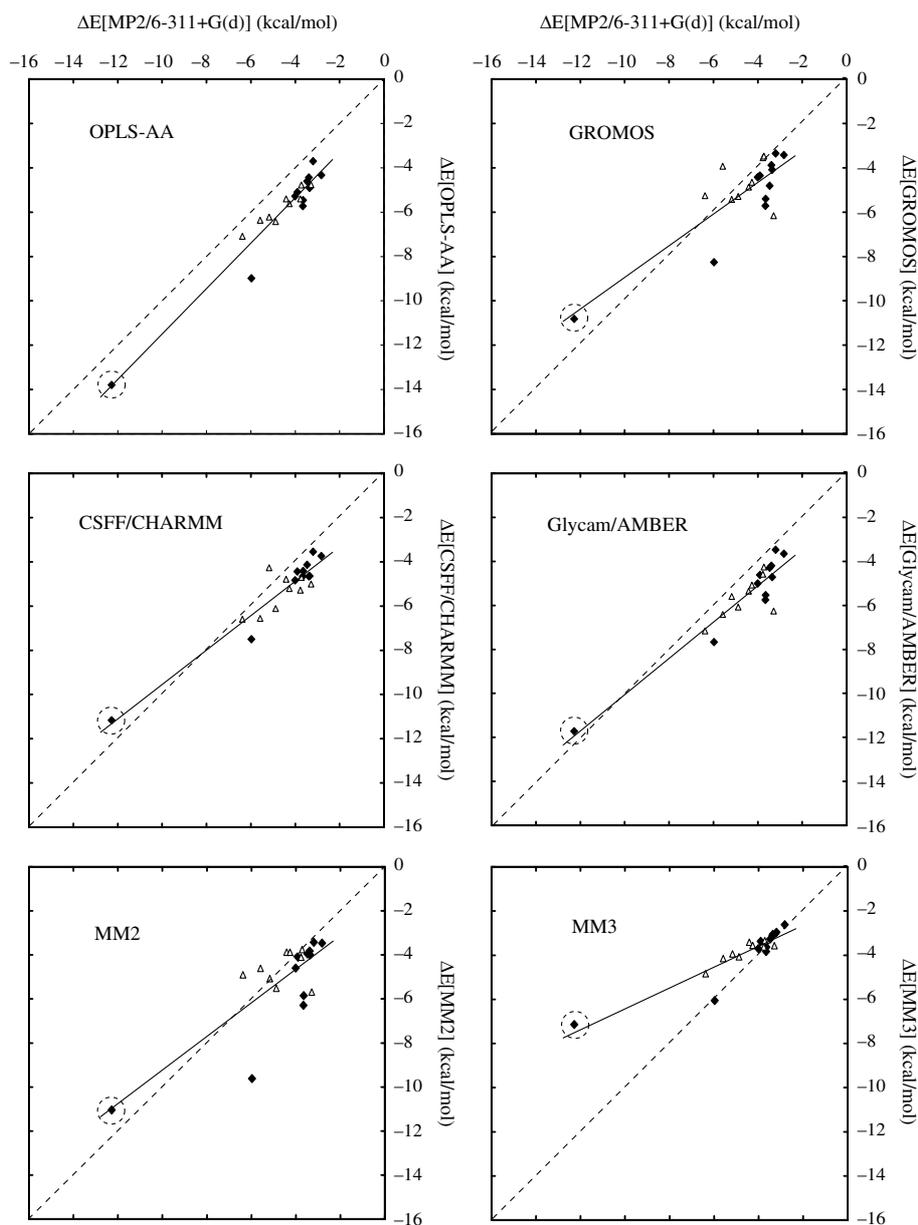


Figure 4. Interaction energies of studied complexes calculated with tested force fields (vertical) vs. *ab initio* at the MP2/6-311+G(d) level (horizontal). Interactions with tyrosines are shown as filled squares and with tryptophans as opened triangles. The circled point is the complex of Tyr369B–Glc406 from *Clostridium thermocellum* endoglucanase A (mutant E95Q) which is dominated by the classical H-bond. See Table 2. and text for details.

between electrostatic force field energies and MP2 energies. This indicates that dispersion energies (accounted mainly by Lennard–Jones terms of force fields) represent the major stabilizing factor of carbohydrate–aromatic interactions while electrostatics determines an orientation of interacting parts of complexes. This is in agreement with decomposition of *ab initio* interaction energies of benzene–alkane complexes [8]. These three force fields interpreted interaction energies as more attractive (approximately 1 kcal/mol). This is illustrated especially by a profile of interaction energy as a function of intermolecular distance. Profiles of these three force fields differed mainly at low intermolecular distances with best performance of Glycam/AMBER. Decomposition of force fields interaction energies can be obtained as a supplementary material.

Relatively good performance was observed with GROMOS force field, taking into account the fact that aliphatic atoms are modelled implicitly (i.e. C–H is modelled as a united atom). A correlation between GROMOS and MP2 energies was poor; however, all interactions were correctly modelled as attractive. The united atom model is less accurate than all-atom model probably due to the fact that C–H bonds play a key role in carbohydrate–aromatic interactions. The trend observed in OPLS-AA, CSFF/CHARMM and Glycam/AMBER (i.e. nearly constant Lennard–Jones terms and difference mainly in electrostatic terms) was not observed in GROMOS. Similarly to previously described force fields, GROMOS modelled interaction energy as more attractive in the profile of interaction energy as a function of distance. Most apparent difference between GROMOS and MP2 energy was at low distance which indicates a possible limit of united atom approach.

Results of CHEAT/CHARMM (correlation not shown) were heavily affected by the fact that hydroxyl hydrogens are modelled implicitly in this force field (O–H group as a united atom). This leads to incorrect modelling of classical H-bonds and interaction energies of H-bond-rich complexes are incorrect. This force field was developed for molecular mechanics studies of complex carbohydrates in vacuum or in solvent and is not suitable for modelling of carbohydrate–protein interactions.

MM2 force field (in which electrostatic interactions are modelled as dipole–dipole interactions with lone pairs dummy atoms) showed surprisingly

poor correlation. The trend observed in OPLS-AA, CSFF/CHARMM and Glycam/AMBER (i.e. nearly constant Lennard–Jones terms and difference mainly in electrostatic terms) was not observed in MM2 force field. On the other hand, the profile of interaction energy as a function of intermolecular distance was most accurately modelled in MM2 and MM3 force fields. This could be explained by the fact that this force field does not significantly overbind carbohydrate–aromatic interactions.

MM3 force field contains a special term for classical H-bonds. This force field showed relatively good correlation with MP2 energies ( $R^2=0.84$ ), however, interaction energy of the complex dominated with a classical H-bond (circled in the Figure 4) was modelled as less attractive. Moreover, electrostatic term of force field interaction energy correlated well with MP2 energy, similarly to OPLS-AA, CSFF/CHARMM and Glycam/AMBER force fields. The fact that MM3 force field underestimates classical H-bonds in carbohydrate modelling has been previously observed [31]. This fact can be partially explained by higher value of relative dielectric constant (set to 1.5, recommended value). Changing the value to 1.0 lead to a higher slope of the correlation and thus better performance (0.79, data not shown). However, without testing this for interactions other than carbohydrate–aromatic, we cannot recommend readers to use a relative dielectric constant set to one. Similarly to MM2 force field, the profile of interaction energy as a function of intermolecular distance was accurately modelled in this force field. This could be explained by the fact that MM2 and MM3 force fields do not significantly overestimate attractive energies of carbohydrate–aromatic interactions (Figure 5).

The tested force fields generally interpreted interactions as slightly more attractive (by approx. 1 kcal/mol) if compared to MP2 energies. Taking into account the fact that MP2 interaction energies in complexes of dispersion nature are usually calculated as more attractive (e.g. if compared with coupled cluster results), we can expect that force field energies are overestimated by 1–2 kcal/mol. However, it seems that a strong correlation between force field and *ab initio* data is more important for a performance of tested force fields (realistic molecular dynamics simulation, structure refinement etc.) than absolute values of interaction energies.

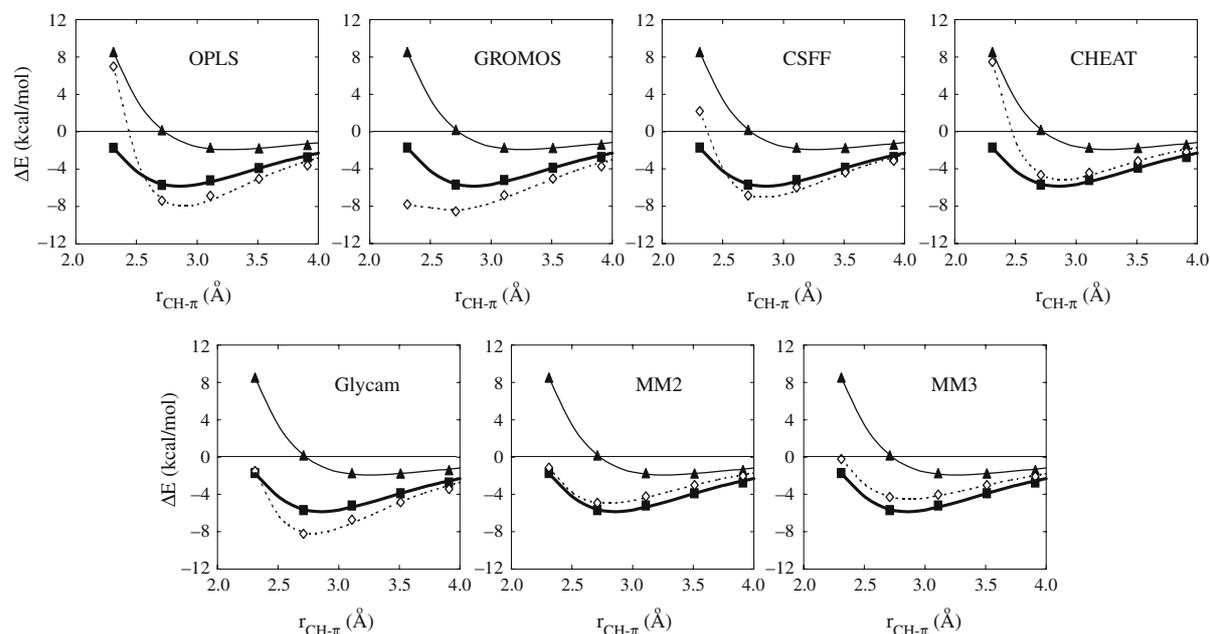


Figure 5. Energy of complex Trp999 from *E. coli*  $\beta$ -galactosidase with glucose (a moiety of allolactose) as a function of distance calculated at the MP2/6-31+G(d) level, (solid squares) HF/6-31+G(d) level (solid triangles) and using each of tested force fields (open diamonds). Points were fitted using the Buckingham potential.

We can conclude that force fields are successfully applicable on carbohydrate–aromatic complexes and no additional CH/ $\pi$  term is necessary. The fact that a significant correlation was found between MP2 and force field energies supports an application of the force field approach in a modelling of carbohydrate recognition. These results show that carbohydrate–aromatic interactions can be modelled using force field approach as a combination of Coulombic and van der Waals term (e.g. using Lennard–Jones potential) and there is no need for implication of additional terms for CH/ $\pi$  interaction into current force fields. On the other hand, addition of special CH/ $\pi$  terms into empirical scoring functions for carbohydrate–protein scoring could be useful. Contrary to force fields, these scoring functions aim to predict free energy rather than interaction energy and such CH/ $\pi$  term in a scoring function could address effects of solvation/desolvation and entropy.

CH/ $\pi$  interactions play an important role in carbohydrate recognition by glycosidases and carbohydrate-binding proteins. Interaction energies of studied model complexes (in the range of  $-12.3$  and  $-2.8$  kcal/mol) were attractive for a

wide range of orientations of the carbohydrate with respect to the aromatic system. The presented study shows that post-Hartree–Fock methods must be applied to produce a quantitative insight into these interactions; nevertheless, performance of the force field approach is satisfactory for most molecular-mechanics-based modelling of carbohydrate recognition.

#### Acknowledgements

Authors would like to gratefully acknowledge the Czech Science Foundation (GACR 204/02/0843) and the Academy of Sciences of the Czech Republic (projects B500500512 and AVOZ 40500505) for financial support.

#### References

- Hobza, P. and Zahradník, R., Weak Intermolecular Interactions in Chemistry and Biology. Elsevier Scientific Pub. Co, Amsterdam, New York, 1980.
- Nishio, M., Hirota, M. and Umezawa, Y., The CH/ $\pi$  Interaction: Evidence, Nature, and Consequences. John Wiley & Sons, New York, 1998.
- Tamres, M., J. Am. Chem. Soc., 74 (1952) 3375.

4. Hobza, P. and Havlas, Z., *Chem. Rev.*, 100 (2000) 4253.
5. Brandl, M., Weiss, M.S., Jabs, A., Sühnel, J. and Hilgenfeld, R., *J. Mol. Biol.*, 307 (2001) 357.
6. Takahashi, O., Kohno, Y., Iwasaki, S., Saito, K., Iwaoka, M., Tomoda, S., Umezawa, Y., Tsuboyama, S. and Nishio, M., *Bull. Chem. Soc. Jpn.*, 74 (2001) 2421.
7. Nishio, M., *Cryst. Eng. Comm.*, 6 (2004) 130.
8. Tsuzuki, S., Honda, K., Uchimar, T., Mikami, M. and Tanabe, K., *J. Am. Chem. Soc.*, 122 (2000) 3746.
9. Tarakeshwar, P., Choi, H.S. and Kim, K.S., *J. Am. Chem. Soc.*, 123 (2001) 3323.
10. Bagnò, A., Saielli, G. and Scorrano, G., *Chem. Eur. J.*, 8 (2002) 2047.
11. Karlstrom, G., Linse, P., Wallqvist, A. and Jonsson, B., *J. Am. Chem. Soc.*, 105 (1983) 3717.
12. Hobza, P., Selzle, H.L. and Schlag, E.W., *J. Am. Chem. Soc.*, 116 (1994) 3500.
13. Hobza, P., Selzle, H.L. and Schlag, E.W., *J. Phys. Chem.*, 100 (1996) 18790.
14. Tsuzuki, S., Uchimar, T., Matsumura, K., Mikami, M. and Tanabe, K., *Chem. Phys. Lett.*, 319 (2000) 547.
15. Hobza, P., Spirko, V., Havlas, Z., Buchhold, K., Reimann, B., Barth, H.-D. and Brutschy, B., *Chem. Phys. Lett.*, 299 (1999) 180.
16. Muraki, M., *Protein Pept. Lett.*, 9 (2002) 195.
17. Sujatha, M.S. and Balaji, B.V., *Proteins*, 55 (2004) 44.
18. Muraki, M. and Harata, K., *Biochemistry*, 39 (2000) 292.
19. Huber, R.E., Hakda, S., Cheng, C., Cupples, C.G. and Edwards, R.A., *Biochemistry*, 42 (2003) 1796.
20. Zolotnitsky, G., Cogan, U., Adir, N., Solomon, V., Shoham, G. and Shoham, Y., *Proc. Natl. Acad. Sci. USA*, 101 (2004) 11275.
21. Davis, A.P. and Wareham, R.S., *Angew. Chem. Int. Ed. Engl.*, 38 (1999) 2978.
22. Spiwok, V., Lipovová, P., Skálová, T., Buchtelová, E., Hašek, J. and Králová, B., *Carbohydr. Res.*, 339 (2004) 2275.
23. Sujatha, M.S., Sasidhar, Y.U. and Balaji, P.V., *Protein Sci.*, 13 (2004) 2502.
24. Sujatha, M.S., Sasidhar, Y.U. and Balaji, P.V., *Biochemistry*, 44 (2005) 8554.
25. del Carmen Fernandez-Alonso, M., Cañada, F.J., Jiménez-Barbero, J. and Cuevas, G., *J. Am. Chem. Soc.*, 127 (2005) 7379.
26. Damm, W., Frontera, A., Tirado-Rives, J. and Jorgensen, W.L., *J. Comp. Chem.*, 18 (1997) 1955.
27. Kuttel, M., Brady, J.W. and Naidoo, K.J., *J. Comput. Chem.*, 23 (2002) 1236.
28. Kouwijzer, M.L.C.E. and Grootenhuys, P.D.J., *J. Phys. Chem.*, 99 (1995) 13426.
29. Woods, R.J., Dwek, R.A., Edge, C.J. and Fraser-Reid, B., *J. Phys. Chem.*, 99 (1995) 3832.
30. Pérez, S., Imbert, A., Engelsen, S.B., Gruza, J., Mazeau, K., Jiménez-Barbero, J., Poveda, A., Espinosa, J.-F., van Eyck, B.P., Johnson, G.J. et al., *Carbohydr. Res.*, 314 (1998) 141.
31. Hemmingsen, L., Madsen, D.E., Esbensen, A.L., Olsen, L. and Engelsen, S.B., *Carbohydr. Res.*, 339 (2004) 937.
32. Kubinyi, H., In Testa, B., van de Waterbeemd, H., Folkers, G. and Guy, R., (Eds.), *Pharmacokinetic Optimization in Drug Research. Biological, Physicochemical, and Computational Strategies*, Helvetica Chimica Acta and Wiley-VCH, Zürich, 2001, pp. 513–524.
33. Ladbury, J.E., *Chem. Biol.*, 3 (1996) 973.
34. Kellenberger, E., Rodrigo, J., Muller, P. and Rognan, D., *Proteins*, 57 (2004) 225.
35. Møller, C. and Plesset, M.S., *Phys. Rev.*, 46 (1934) 618.
36. Lütteke, T., Frank, M. and von der Lieth, C.W., *Carbohydr. Res.*, 339 (2004) 1015.
37. Guérin, D.M.A., Lascombe, M.-B., Costabel, M.B., Souchon, H., Lamzin, V., Béguin, P. and Alzari, P.M., *J. Mol. Biol.*, 316 (2002) 1061.
38. Sanders, D.A.R., Moothoo, D.N., Raftery, J., Howard, A.J., Helliwell, J.R. and Naismith, J.H., *J. Mol. Biol.*, 310 (2001) 875.
39. Merritt, E.A., Kuhn, P., Sarfaty, S., Erbe, J.L., Holmes, R.K. and Hol, W.G.J., *J. Mol. Biol.*, 282 (1998) 1043.
40. Notenboom, V., Boraston, A.B., Williams, J.S., Kilburn, D.G. and Rose, D.R., *Biochemistry*, 41 (2002) 4246.
41. Varrot, A., Frandsen, T.B., von Ossowski, I., Boyer, V., Cottaz, S., Driguez, H., Schüle, M. and Davies, G.J., *Structure*, 11 (2003) 855.
42. Charnock, S.J., Bolam, D.N., Nurizzo, D., Szabó, L., McKie, V.A., Gilbert, H.J. and Davies, G.J., *Proc. Natl. Acad. Sci. USA*, 99 (2002) 14077.
43. Schmidt, M.W., Baldrige, 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., *J. Comput. Chem.*, 14 (1993) 1347.
44. Boys, S.F. and Bernardi, F., *Mol. Phys. (UK)*, 19 (1970) 553.
45. Jorgensen, W.L. and Tirado-Rives, J., *J. Am. Chem. Soc.*, 110 (1988) 1657.
46. van Gunsteren, W.F., Billeter, S., Eising, A.A., Hübenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. and Tironi, I.G., *Biomolecular Simulations: The GROMOS 96 Manual and User Guide* VDF Zürich, 1996.
47. MacKerell, Jr., A.D., Bashford, D., Bellott, M., Dunbrack, R.L. Jr., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, III., W.E., Roux, B., Schlenker, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. and Karplus, M., *J. Phys. Chem. B.*, 102 (1998) 3586.
48. Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.R., Cheatham, T.E., DeBolt, III., S., Ferguson, D., Seibel, G. and Kollman, P., *Comp. Phys. Commun.*, 91 (1995) 1.
49. Allinger, N.L., Kok, R.A. and Imam, M.R., *J. Comput. Chem.*, 9 (1988) 591.
50. Lii, J.-H. and Allinger, N.L., *J. Am. Chem. Soc.*, 111 (1989) 8576.
51. Pappu, R.V., Hart, R.K. and Ponder, J.W., *J. Phys. Chem. B.*, 102 (1998) 9725.
52. Guex, N. and Peitsch, M.C., *Electrophoresis*, 18 (1997) 2714.
53. Turnbull, W.B., Precious, B.L. and Homans, S.W., *J. Am. Chem. Soc.*, 126 (2004) 1047.
54. Coutinho, P.M. and Henrissat, B., *Carbohydrate-Active Enzymes server*, 1999: <http://afmb.cnrs-mrs.fr/CAZY/>.
55. Coutinho, P.M. and Henrissat, B., In Gilbert, H.J., Davies, G., Henrissat, B. and Svensson, B. (Eds.), *Recent Advances in Carbohydrate Bioengineering*. The Royal Society of Chemistry, Cambridge, 1999, pp. 3–12.
56. Johnson, E.R., Wolkow, R.A. and DiLabio, G.A., *Chem. Phys. Lett.*, 394 (2004) 334.
57. Grimme, S., *J. Comput. Chem.*, 25 (2004) 1463.
58. Morales, J.C. and Penadés, S., *Angew. Chem. Int. Edit.*, 37 (1998) 654.