

STRUCTURAL GLYCOMICS – MOLECULAR DETAILS OF PROTEIN-CARBOHYDRATE INTERACTIONS AND THEIR PREDICTION

DIRK NEUMANN¹ AND OLIVER KOHLBACHER^{1,2,*}

¹neumann scientific computing & consulting, Berlin, Germany

²Center of Bioinformatics, Division for Simulation of Biological Systems,
Sand 14, University of Tübingen, Tübingen, Germany

E-Mail: * oliver.kohlbacher@uni-tuebingen.de

Received: 8th April 2010 / Published: 10th December 2010

ABSTRACT

Protein-ligand docking is an essential technique in computer-aided drug design. While generally available docking programs work well for most drug classes, carbohydrates and carbohydrate-like compounds are often problematic for docking. We discuss the peculiarities of protein-carbohydrate interactions and their impact on protein-carbohydrate docking and review the state of the art in docking of carbohydrates to proteins. Finally, we give an overview of carbohydrate docking studies and present a new docking method specifically designed to handle docking of carbohydrate-like compounds. BALLDock/SLICK combines an evolutionary docking algorithm for flexible ligands and flexible receptor side chains with carbohydrate-specific scoring and energy functions. The scoring function has been designed to identify accurate ligand poses, while the energy function yields accurate estimates of the binding free energies of these poses. On a test set of known protein-sugar complexes we demonstrate the ability of the approach to generate correct poses for almost all of the structures and achieve very low mean errors for the predicted binding free energies.

INTRODUCTION

The interactions between carbohydrates and proteins are of special interest, because glycoproteins and glycolipids play numerous fundamental roles in biological processes such as cell fertilization, differentiation, and development. Realizing the importance of carbohydrates has even led to the term “glycomics” to indicate a new era of carbohydrate research [1].

One of the main reasons for the observed versatility of carbohydrates in transmitting information is the complex structure of carbohydrates. The huge number of possible carbohydrate structures, – glycospace – can encode huge amounts of information [2]. In mammals, however, only a small part of glycospace seems to be used [3]. This potential for transmitting information has been exploited by nature in various forms, where protein-carbohydrate interactions are crucial for many biological processes including cell-cell recognition, proliferation, and differentiation. A prominent example is the interaction of leukocyte L-selectin with its endothelial ligand, which is crucial for initiating leukocyte recruitment to sites of inflammation [4]. Glycosylation determines antibody activity, too, and is responsible for mediating triggered inflammatory responses [5 – 7]. It has also been shown that the glycosylation patterns differ between normal cells and cancer cells [8, 9].

While the protein-carbohydrate interaction may be used for defensive purposes, it has also been hijacked by other organisms, bacteria and viruses [10 – 12]. For example, *Helicobacter pylorus* binds to carbohydrates expressed on gastric epithelial cells [13]. Bacterial toxins like cholera toxin, *Escherichia coli* enterotoxin pertussis toxin, and tetanus toxin attach themselves to carbohydrates on cell surfaces and subsequently enter the cell.

Of course, the protein-carbohydrate interaction may also be exploited in novel therapeutic approaches [14, 15]. In tumour therapy for example, galectins are promising targets [16]. In the case of HIV, several proteins were reported to bind to the highly glycosylated envelope protein gp120 – an interaction which may be the basis for a vaccine [17 – 19]. More generally, protein-carbohydrate interactions may be used to enhance the efficacy of drug delivery systems [20, 21] or to define the specificity of drug targeting systems for certain, possibly malignant, cells or tissues [22, 23]. The large potential of such approaches was shown in several studies [24, 25].

A thorough understanding of the factors governing the affinity between proteins and carbohydrates might help to design tailor-made carbohydrates or proteins to better exploit this special type of interaction. However, experimental assessment of the carbohydrate-recognition by NMR spectroscopy or X-ray crystallography is impeded by difficulties of co-crystallizing proteins and carbohydrates. Here, computational modelling might help to increase our

understanding of the different contributions to the binding energy. Recent developments allow searching the conformational space efficiently and yield reliable estimates of the binding free energy.

PROTEIN-CARBOHYDRATE INTERACTION

Properties of Carbohydrates

While current docking methods are quite successful when dealing with drug-like structures, they are having difficulties with carbohydrate-like structures due to their different physico-chemical properties. The average log P(octanol/water) of drugs is around 2.5. They typically have only few hydrogen bond acceptors or donors, and a small number of rotatable bonds [26, 27]. In contrast, carbohydrates and their derivatives possess many hydroxyl groups and thus a large number of rotatable bonds. Due to the many hydroxyl groups, these compounds are usually highly water soluble and their log P is often negative (Tab. 1).

Table 1. Selection of carbohydrates, carbohydrate derivatives, and drug molecules with high similarity to sugar molecules.

Compound/ Generic Name	XlogP3	#H-bond donors	#H-bond acceptors
Glucose	-2.6	5	6
Cellobiose	-4.7	8	11
4-nitrophenyl- β -D-glucoside	-0.4	4	8
Phenylglucoside	-0.9	4	6
Acarbose	-8.5	14	19
Ademoetionine	-12.2	16	49
Anthelmycin	-8.5	13	17
Digoxin	1.3	6	14
Paromomycin	-8.7	13	19
Propikacin	-8.0	13	17
Streptomycin	-8.0	12	19
Tobramycin	-6.2	10	14

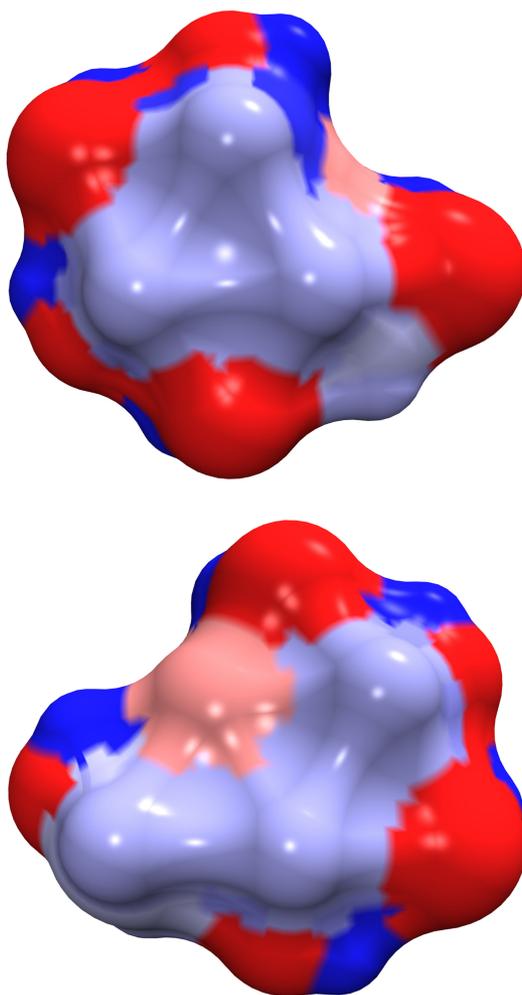


Figure 1. Electrostatic potential of β -D-glucopyranose. Color code: blue – positive charge, red – negative charge.

The surface of carbohydrates and their derivatives is composed of hydrophobic and hydrophilic patches formed by nonpolar aliphatic protons and polar hydroxyl groups (Fig. 1), which leads to anisotropic solvent densities around carbohydrate molecules [28]. In aqueous solution, favourable interaction of water molecules with the hydrophilic patches results from electrostatic interactions and hydrogen bonding. At the same time, the interaction of water with hydrophobic surface patches is unfavourable. The balance between hydrophobic and hydrophilic patches is essential for carbohydrate solubility, but also for molecular recognition.

Another essential property of carbohydrates is their conformational flexibility. Compared to drug-like molecules, carbohydrates are typically much more flexible. Similar to amino acid dimers, the energetically favourable conformations of carbohydrate dimers may be easily shown by Ramachandran plots. Figure 2 shows the Ramachandran plots for a dipeptide (with end caps) and cellobiose, a glucose dimer. At a cursory glance, both plots look very similar featuring multiple minima, with the separating energy barriers being higher in the case of cellobiose. However, carbohydrates in complex were found to adopt conformations belonging to different minima. These observations underline the necessity for thoroughly sampling the conformational space of carbohydrate oligomers during docking. While this may be feasible for glycosidic bonds, the number of degrees of freedom increases rapidly when additionally taking into account the orientation of the hydroxyl groups.

Carbohydrate Binding Proteins

The binding partners of carbohydrates encompass a wide variety of proteins including periplasmic receptor proteins, antibodies, lectins and enzymes. Lectins are proteins which bind carbohydrates with high specificity and which have been found both in plants and living organisms. While the biological roles of plant lectins are not fully understood by now, the functions of many lectins in higher organisms and especially in mammals have been elucidated in detail giving rise to the following short definition: lectins are molecular tools to decipher sugar-encoded messages [29].

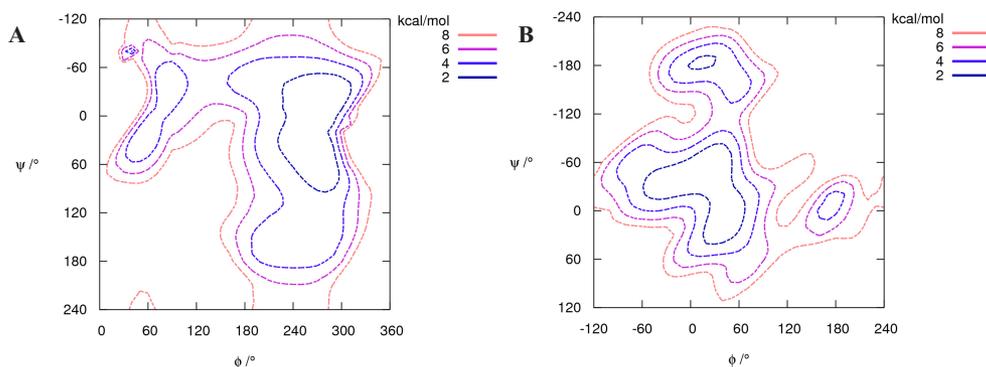


Figure 2. Comparison of peptide and carbohydrate flexibility. (A) Ramachandran plot for ACE-VAL-VALNME. (B) Energy map for cellobiose. Energies were calculated using MM3 and are relative to minimum energy found.

A variety of lectins is involved in the innate immune system: For example, the binding of the mannose-binding lectin is one of the first steps in opsonization of pathogens. Galectins, β -binding animal lectins which are involved in intracellular signalling pathways, also play important roles in immune and inflammatory responses [30]. Interestingly, galectins are not only localized in the cell membrane or intracellularly, but have been detected in extracellular

space, too – very probably exerting different biological functions. The binding sites of lectins are usually rather shallow, located near the protein surface and thus accessible to solvent, whereas the binding sites of periplasmic receptors are buried in these proteins [31].

In addition to the lectins, enzymes like isomerases, glycoside hydrolases and glycosyltransferases belong to the class of carbohydrate binding proteins, too. While glycoside hydrolases (*e.g.*, lysozymes and chitinases) catalyze the hydrolysis of the glycosidic linkage, glycosyltransferases catalyze the transfer of a monosaccharide unit to, *e.g.*, a lipid or protein.

However, irrespective of its biological effect and in analogy to the surface of carbohydrates the protein surface may be divided in patches which interact favourably with water molecules and patches where rearrangement of the water molecules leads to an unfavourable interaction.

Contributions to the Binding Free Energy

Macroscopically, the formation of complexes between carbohydrates and proteins may be driven by favourable changes in enthalpy or entropy.

Usually, the binding free energies ΔG of carbohydrate monomers binding to proteins are quite small. For dimers or higher oligomers, ΔG increases significantly [32], but usually even these binding free energies are smaller than the binding free energies of other complexes such as, *e.g.*, MHC-peptide complexes [33]). In spite of the small binding free energies, the interaction between proteins and carbohydrates plays a crucial role in biological processes and may be used for pharmaceutical purposes, *e.g.* for modifying drug delivery systems with lectins, due to the polyvalent effect which leads to a high avidity [34] (for a single lectin molecule, there is a large number of possible interaction partners on a single cell surface). In addition to this obvious statistical effect, the simultaneous binding of, *e.g.*, two or more lectins with a glycocalyx may anchor a drug delivery system to a cell.

When performing docking studies, however, the variable of interest is the binding free energy between a single carbohydrate and a single carbohydrate binding protein. Usually, ΔG is assumed to be composed of independent contributions. Typically, they include hydrogen bonding, van-der-Waals forces, and consequences of the hydrophobic effect.

The exact terms used to approximate ΔG or ΔH may differ, because they inherently depend on the underlying model used for energy decomposition. In the following we will shortly outline several contributions to the binding free energy.

Hydrophobic Effect

Protein-ligand complexes may be stabilized by the so-called hydrophobic effect. In protein-carbohydrate interactions, this effect often leads to the stacking of aromatic residues against the faces of sugars. Two variations of the hydrophobic effect are discussed in literature: the “classical” hydrophobic effect and the “non-classical” hydrophobic effect [35]. The classical hydrophobic effect results mainly from a highly favourable entropy of formation. Here, small hydrophobic solutes induce an ordering of the water molecules at the solvent-surface interface. Decrease of the hydrophobic surface area upon complexation leads to a decrease in solvent ordering and hence to favourable changes in entropy. In the case of the “non-classical” hydrophobic effect, the complex formation is mainly enthalpy driven due to favourable interactions between the solute molecules forming the complex as well as favourable interactions between the solvent molecules.

Such non-classical hydrophobic effects have been reported for lectin-carbohydrate complexes [36], where the solvent interactions may provide 25–100% of the enthalpy of binding [37].

CH/ π Interactions

The enthalpy for carbohydrates binding to proteins is characterized by a special type of interaction, the so called CH/ π interaction. The CH/ π interaction was defined as a hydrogen bond formed between a hydrogen attached to a carbon atom and the π system of arenes. Although much weaker than classical hydrogen bonds, CH/ π interactions may contribute significantly to the enthalpy of binding. On the one hand, the presence of the electronegative oxygens bound to the carbon atoms probably increases the enthalpy resulting from the formation of a single CH/ π interaction. On the other hand, multiple CH/ π interactions may be formed already in the case of a methyl group interacting with a benzene. The number of interactions may be even higher for the stacking of glucose or galactose with aromatic sidechains.

The calculation of this effect using quantum mechanics is impeded by the high level of theory necessary [38]. Apart from enhancing the affinity between carbohydrate and protein, CH/ π interactions probably determine the specificity of the carbohydrate binding proteins as well.

Hydrogen Bonds

Hydrogen bonds may be established between polar hydrogen atoms and lone pairs of hydrogen bond acceptors. Due to the large number of hydrogen bond donors and acceptors present in carbohydrates, they tend to form hydrogen bonds when in complex with a protein.

Here, both binding partners compete with water molecules for the hydrogen bonds. As a result, the overall enthalpic gain from a hydrogen bond formed between carbohydrate and protein may be small.

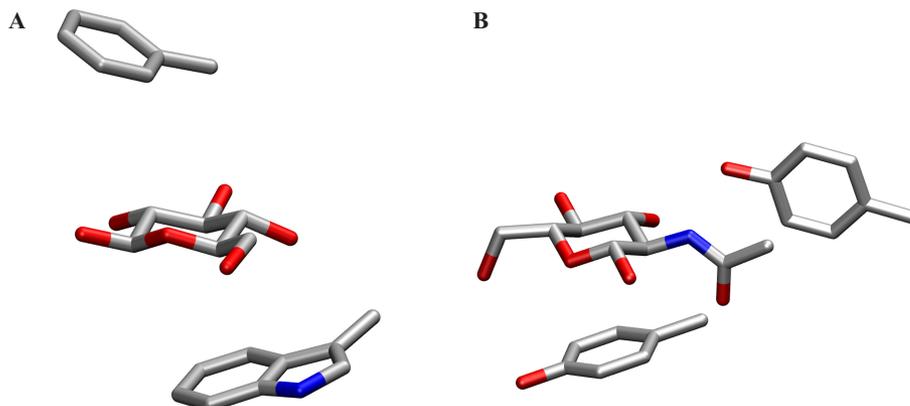


Figure 3. Stacking of (A) β -D-glucose and aromatic sidechains of the glucose/galactose binding protein (PDB accession code: 2FVY) and (B) of *N*-acetylglucosamine and tyrosine residues in wheat germ agglutinin (PDB accession code: 2UVO).

Although carbohydrates often displace all water molecules in the binding site, in a number of cases conserved water molecules are observed in the binding site. These water molecules mediate protein-carbohydrate interactions, especially if no or only few direct hydrogen bonds are established [39, 40]. However, water molecules may also help in stabilizing oligosaccharide conformations [41]. Although there exist some approaches for taking into account solvent molecules in the binding site [42], calculating water mediated hydrogen bonds of the carbohydrate with itself is still very difficult.

Electrostatic Interactions and van-der-Waals Interactions

Electrostatic interactions take place between partially charged atoms. The effects are higher for charged molecules such as sialic acids, sulfatides or protonated or deprotonated protein sidechains. Electrostatic forces are part of many force fields used in molecular modelling and thus routinely computed using Coulomb's law. The calculation of the necessary atomic partial charges of carbohydrates for use in established force fields has been repeatedly addressed in literature [43]. For docking purposes, often a reduced dielectric constant or a distance-dependent dielectric function is employed.

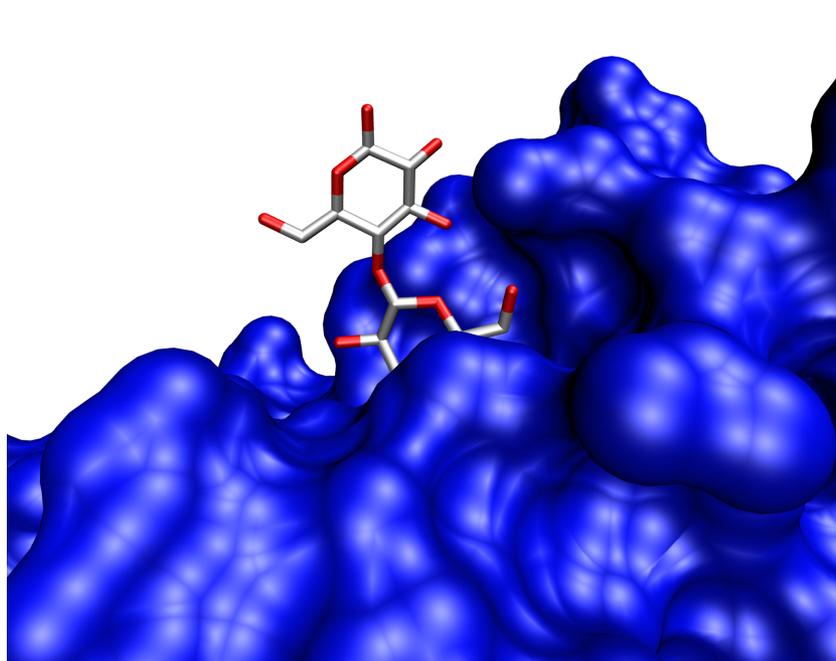


Figure 4. Example of a carbohydrate dimer extending into the solvent instead of forming a closely packed complex (PDB accession code: 1UZY).

In contrast to electrostatic interactions, van-der-Waals forces are, roughly speaking, weak attractive forces arising from the interactions between – not necessarily permanent – dipoles. At short ranges, however, overlap between electron orbitals results in a strong repulsion between atoms. The effect of both interactions on complex stability is often modelled using the Lennard-Jones potential. While the energetic contribution of a single pair of atoms to the total binding free energy may be small, summation over all pairwise interactions occurring between protein and ligand may result in a comparatively high favourable interaction.

Solvation/Desolvation

A typical problem that arises in docking carbohydrate dimers and oligomers is the tendency of docking programs to maximize atomic contacts between ligand and protein. This results in structures with small volume, where the carbohydrate lies more or less flat on the protein surface. However, several X-ray structures show contradictory structures with carbohydrate residues extending well into the surrounding solvent (Fig. 4).

Possibly, these structures might be correctly predicted, if the impact of solvation and desolvation on the binding free energy was computed with sufficient accuracy. The more accurate methods, in particular Poisson-Boltzmann methods, are unfortunately computationally rather expensive. While they can be used in re-scoring of docking conformation, their use in the inner loop of docking, the pose generation, is still prevented by large computation times.

PROTEIN-CARBOHYDRATE DOCKING

Force Fields and Scoring Functions for Carbohydrate Docking

To simulate the behaviour of carbohydrate *in vacuo* or in solution (e.g., to study ring puckering [44] or rotational barriers of oligosaccharides), either established force fields or special parameterizations may be used [45–49]. Such force fields allow for investigating the deformation of carbohydrate rings as well as predicting. These special force fields (as well as previously established ones) have been employed repeatedly for molecular dynamics simulations of protein-carbohydrate complexes [50, 51]. In some cases, the simulations were even used for estimating successfully binding free energies [40, 52–54].

Despite the many possible advantages of established force fields, they were not designed to predict binding free energies or enthalpies in protein-ligand docking. Since solvent molecules are usually modelled explicitly, force fields do not need to include extra terms for hydrophobic effects. The special CH/ π interactions are not taken into account. Some force fields do model for hydrogen bonds explicitly, while others regard it as part of the electrostatic interaction. Irrespective of the approach, displacement of water molecules competing for hydrogen bonds is not accounted for.

Although some force fields correlate well with *ab initio* calculations for *ab initio* optimized geometries [55], a scoring function for docking must be able to distinguish between decoy structures and true hits. A recent comparison of the results of *ab initio* and force field calculations underlines the difficulties in predicting binding enthalpies in protein-carbohydrate complexes using existing force fields: the stabilizing interaction energy for the interaction between fucose and tryptophane is heavily overestimated by the AMBER* force field [56].

Search Algorithms and Docking Programs

A major problem in docking carbohydrates is the possibly large number of degrees of freedom to be optimized during the docking calculation. In many force fields and consequently in most docking programs, polar hydrogens are modelled explicitly, which leads to many rotatable bonds and, hence, numerous degrees of freedom when docking carbohy-

drates. Since well-established docking algorithms are able to handle only a limited number of degrees of freedom with high efficiency, they may fail due to limited sampling of the conformational space.

Case Studies

Structures of proteins in complex with carbohydrates or carbohydrate derivatives have been available from the Protein Data Bank for some time and thus have been part of the data sets used for developing docking algorithms and docking programs.

A protein that has been repeatedly investigated in this context is neuraminidase. Native docking experiments of neuraminidase ligands performed with PRO_LEADS showed that these ligands are more difficult to dock than others [57]. The failure was attributed mainly to a failure of the energy function used: for two neuraminidase derivatives, decoy structures were ranked better than the crystal structure. However, another interesting aspect is the small size of the two top ranking clusters obtained in the docking experiments with sialic acid (PDB code: 1NSC): only 6 of the 100 docking results populate the top ranked cluster located in the binding site. For two of the complexes, experimental binding affinities are available and can be compared with predicted energies: in one case, the binding energy is overestimated by 5 kJ/mol, whereas in the other it is underestimated by the same amount.

A complex very similar to 1NSC (pdb code: 2SIM) was successfully reproduced by the program GOLD [58]. Like PRO_LEADS, the GOLD energy function honours the formation of hydrogen bonds. Successful docking to the L-arabinose-binding protein (PDB codes: 1ABE, 1ABF) was reported repeatedly for several docking algorithms and energy/scoring functions including AutoDock 3.0.5 and MOE 2004.03 [59–61].

For the hemagglutinin-sialic acid complex 4HMG, however, docking using AutoDock and MOE 2004.03 failed [59]. Despite the success of AutoDock in identifying the binding pose, it overpredicted the binding free energy by more than 10 kJ/mol [62].

In comparison to AutoDock, GOLD and ICM, the docking programs FlexX and DOCK, which both use an incremental reconstruction algorithm, failed to identify the correct binding pose for some complexes of neuraminidase and L-arabinose binding protein [63]. This hints at a problem of these algorithms with selecting and placing the first ligand fragments. In a series of publications, the glucosyl saccharides were docked into the glucoamylase active site using AutoDock [64, 65]. For the 9 investigated disaccharides, the solutions in highest ranked cluster corresponded to the crystal structures. These results underlined the potential of the AutoDock program and its scoring function in docking carbohydrates.

In 1999, Minke *et al.* published a method to predict water interactions for use in carbohydrate docking studies [66]. The authors employed AutoDock 2.2 to develop their methodology using known X-ray structures of heat-labile enterotoxin. Subsequently, they performed docking calculations with a number of carbohydrate derivatives without water molecules and with predicted water molecules. In some cases, inclusion of the predicted water molecules enabled the program to better identify the binding pose in the blind docking tests. Despite the partial success, the AutoDock scoring function was not able to rank true hits better than decoy structures in all cases. In summary, the success of this scoring function seems to depend on the properties of the protein-carbohydrate complex investigated.

Since most docking programs seem to neglect the peculiar interactions in protein-carbohydrate complexes, specific scoring functions for carbohydrate docking have been developed.

Starting from the AutoDock scoring function, Reilly and co-workers identified new coefficients as well as additional terms for use in carbohydrate docking with AutoDock 3.0.6. In a first study, a training set of 30 protein-carbohydrate complexes was employed to fit the coefficients and to investigate the effect of two different hydrogen bonding terms on the calculated binding free energy. The test set consisted of 17 complexes with experimentally determined binding free energies. The smallest error in predicted binding free energies was obtained, when docking was performed using a molecular mechanics potential energy function and the results were re-evaluated using the newly developed free energy model [67]. In a subsequent study, a much larger training set was used to re-fit the parameters and to further extend the previous free energy model [68]. The best model featured a root mean squared error less than 8.5 kJ/mol. However, the authors conclude that the solvation and entropic terms are still partially poorly modelled.

Kerzmann *et al.* chose a similar approach [69, 70]. However, the authors used two different functions throughout their studies: one function for (re-)scoring putative solutions (SLICK/score) and another, much more elaborate function to calculate the binding free energies of promising docking results (SLICK/energy). Both scoring and energy function contained special terms for hydrogen bonding and CH/ π interactions accounting not only for the distance-dependence but also for the directionality of these interactions (Fig. 5).

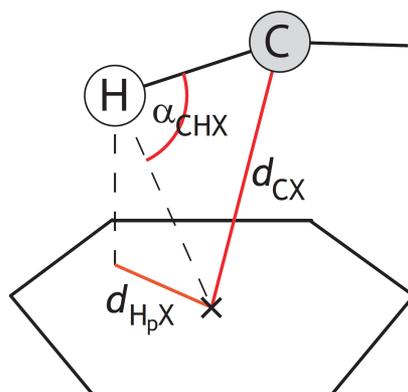


Figure 5. The geometry of the CH/ π interaction employed in SLICK. d_{CX} is the distance between aliphatic carbon and ring centre, α_{CHX} denotes the angle between CH bond and the connection between the hydrogen atom and the ring centre. d_{H_pX} is the distance of H to the ring centre projected into the ring plane.

Van-der-Waals and electrostatic interactions were calculated essentially using the AMBER/GLYCAM200a force field [71]. In addition, the energy function contained special terms to take into account changes in ΔG due to solvation/desolvation of the ligand and protein. Originally, scoring and energy function were trained on a set of 18 protein-carbohydrate complexes. Re-scoring docking results for this training set showed the potential of the new scoring function. For this purpose, the protein-carbohydrate complexes were first re-docked using AutoDock 3. Then the average rank of the first true positive, that is the first hit obtained during docking with a root-mean square deviation (RMSD) less than 1.5 Å, was calculated. The average rank for the AutoDock energy function and SLICK/score was 33.1 and 8.3, respectively. These results show that although AutoDock was able to generate solutions close to the crystal structures, its energy function ranked decoy structures higher. Using SLICK/score during docking with a genetic algorithm resulted in an average rank of the first true positive of 1.2 for the training set and 3.6 for a test set consisting of 20 lectin-carbohydrate complexes. Binding free energies were calculated with SLICK/energy for all docking results of the training set. The Spearman rank correlation coefficients were $\rho_{fp} = 0.71$ and $\rho_{min} = 0.85$ for the energies of the first true positives and of the hits with minimal RMSD, respectively. This result shows that even small deviations from the crystal structure may lead to deterioration of the calculated binding free energy and underlines the necessity to thoroughly sample energy minima during docking. The quality of SLICK/energy was further assessed by re-docking an external energy validation set. For these structures, the free energy model by Kerzmann *et al.* featured a root mean squared error for the binding free energy less than 4 kJ/mol. Although both training and test set used in the studies by Kerzmann *et al.* were quite small, the results are very encouraging.

In summary, in view of the good performance of the energy models presented here, the modification of existing functions for application in carbohydrate docking is certainly worthwhile. Although the number of groups active in this field is limited, the importance of protein-carbohydrate interactions in biological processes will definitely stimulate further research.

CONCLUSION

There exist already some promising approaches to protein-carbohydrate docking. However, to calculate binding free energies or enthalpies with the necessary accuracy, increasing the understanding of the interactions on an atomic level is crucial. Simple experimental models in conjunction with high-level *ab initio* calculations will certainly help in gaining additional insight and thus form the basis for future scoring function to be used in protein-carbohydrate docking [72]. Finally, recently established methods for handling many rotatable bonds in flexible docking will allow for docking even oligosaccharides with high efficiency [73, 74]. Ultimately, applications for carbohydrate-docking will allow for exploiting protein-carbohydrate interactions – especially for therapeutic purposes. Well-designed carbohydrate mimetics or lectinomimetics might be employed as new therapeutic agents. Similarly, the effectiveness of drug-delivery systems and the specificity of drug-targeting systems may be increased considerably.

REFERENCES

- [1] Feizi, T. and Mulloy, B. (2003) Carbohydrates and glycoconjugates. Glycomics: the new era of carbohydrate biology. *Curr. Opin. Struct. Biol.* **12**:602 – 604.
doi: <http://dx.doi.org/10.1016/j.sbi.2003.09.001>.
- [2] Solís, D., Jiménez-Barbero, J., Kaltner, H., Romero, A., Siebert, H.-C., von der Lieth, C.-W. and Gabius, H.-J. (2001) Towards defining the role of glycans as hardware in information storage and transfer: Basic principles, experimental approaches and recent progress. *Cells Tissues Organs* **168**:5 – 23.
doi: <http://dx.doi.org/10.1159/000016802>.
- [3] Werz, D.B., Ranzinger, R., Herget, S., Adibekian, A., von der Lieth, C.-W., and Seeberger, P.H. (2007) Exploring the structural diversity of mammalian carbohydrates (“glycospace”) by statistical databank analysis. *ACS Chem. Biol.* **2**:685 – 691.
doi: <http://dx.doi.org/10.1021/cb700178s>.
- [4] Khan, A.I., Landis, R.C. and Malhotra, R. (2003) L-selectin ligands in lymphoid tissues and models of inflammation. *Inflammation* **27**:265 – 280.
doi: <http://dx.doi.org/10.1023/A:1026056525755>.
- [5] Burton, D.R. and Dwek, R.A. (2006) Sugar determines antibody activity. *Science* **313**:627 – 628.
doi: <http://dx.doi.org/10.1126/science.1131712>.
- [6] Kaneko, Y., Nimmerjahn, F. and Ravetch, J.V. (2006) Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science* **313**:670 – 673.
doi: <http://dx.doi.org/10.1126/science.1129594>.
- [7] Altmann, F. (2007). The role of glycosylation in allergy. *Int. Arch. Allergy Immunol.* **142**:99 – 115.
doi: <http://dx.doi.org/10.1159/000096114>.
- [8] Moore, R., King, N. and Alroy, J. (1988) Differences in cellular glycoconjugates of quiescent, inflamed, and neoplastic colonic epithelium in colitis and cancer-prone tamarins. *Am. J. Pathol.* **131**:484 – 489.
- [9] Gorelik, E., Galili, U. and Raz, A. (2001). On the role of cell surface carbohydrates and their binding proteins (lectins) in tumor metastasis. *Cancer Metastasis Rev.* **20**:245 – 277.
doi: <http://dx.doi.org/10.1023/A:1015535427597>.
- [10] Ofek, I. and Sharon, N. (1988) Lectinophagocytosis: a molecular mechanism of recognition between cell surface sugars and lectins in the phagocytosis of bacteria. *Infect. Immun.* **56**:539 – 547.
-

- [11] Bomsel, M. and Alfsen, A. (2003) Entry of viruses through the epithelial barrier: pathogenic trickery. *Nat. Rev. Mol. Cell. Biol.* **4**:57–68.
doi: <http://dx.doi.org/10.1038/nrm1005>.
- [12] Vigerust, D.J. and Shepherd, V.L. (2007) Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol.* **15**:211–218.
doi: <http://dx.doi.org/10.1016/j.tim.2007.03.003>.
- [13] Kobayashi, M., Lee, H., Nakayama, J. and Fukuda, M. (2009). Roles of gastric mucin-type O-glycans in the pathogenesis of *Helicobacter pylori* infection. *Glycobiology* **19**:453–461.
doi: <http://dx.doi.org/10.1093/glycob/cwp004>.
- [14] Hong, F., Yan, J., Baran, J.T., Allendorf, D.J., Hansen, R.D., Ostroff, G.R., Xing, P.X., Cheung, N.-K.V. and Ross, G.D. (2004) Mechanism by which orally administered β -1,3-glucans enhance the tumoricidal activity of antitumor monoclonal antibodies in murine tumor models. *J. Immunol.* **173**:797–806.
- [15] Oppenheimer, S.B., Alvarez, M., Nnoli, J. (2008) Carbohydrate-based experimental therapeutics for cancer, HIV/aids and other diseases. *Acta Histochem.* **110**:6–13, 2008.
doi: <http://dx.doi.org/10.1016/j.acthis.2007.08.003>.
- [16] Ingrassia, L., Camby, I., Lefranc, F., Mathieu, V., Nshimyumukiza, P., Darro, D. and Kiss, R. (2006) Anti-galectin compounds as potential anti-cancer drugs. *Curr. Med. Chem.* **13**:3513–3527.
doi: <http://dx.doi.org/10.2174/092986706779026219>.
- [17] Ji, X., Gewurz, H. and Spear, G.T. (2005) Mannose binding lectin (MBL) and HIV. *Mol. Immunol.* **42**:145–152.
doi: <http://dx.doi.org/10.1016/j.molimm.2004.06.015>.
- [18] Balzarini, J. (2006) Inhibition of HIV entry by carbohydrate-binding proteins. *Antivir. Res.* **71**:237–247, 2006.
doi: <http://dx.doi.org/10.1016/j.antiviral.2006.02.004>.
- [19] Sirois, S., Touaibia, M., Chou, K.C. and Roy, R. (2007) Glycosylation of HIV-1 gp120 V3 loop: towards the rational design of a synthetic carbohydrate vaccine. *Curr. Med. Chem.* **14**:3232–3242.
doi: <http://dx.doi.org/10.2174/092986707782793826>.
- [20] Clark, M.A., Hirst, B.H. and Jepson, M.A. (2000) Lectin-mediated mucosal delivery of drugs and microparticles. *Adv. Drug Deliv. Rev.* **43**:207–223.
doi: <http://dx.doi.org/10.2174/092986707782793826>.
-

- [21] Russell-Jones, G.J. (2001) The potential use of receptor-mediated endocytosis for oral drug delivery. *Adv. Drug Deliv. Rev.* **46**:59–73.
doi: [http://dx.doi.org/10.1016/S0169-409X\(00\)00127-7](http://dx.doi.org/10.1016/S0169-409X(00)00127-7).
- [22] Minko, T. (2004) Drug targeting to the colon with lectins and neoglycoconjugates. *Adv. Drug. Deliv. Rev.* **56**:491–509.
doi: <http://dx.doi.org/10.1016/j.addr.2003.10.017>.
- [23] Nishikawa, M. (2005) Development of cell-specific targeting systems for drugs and genes. *Biol. Pharm. Bull.* **28**:195–200.
doi: <http://dx.doi.org/10.1248/bpb.28.195>.
- [24] Osborn, H.M.I, Evans, P.G., Gemmell, N. and Osborne, S.D. (2004) Carbohydrate-based therapeutics. *J. Pharm. Pharmacol.* **56**:691–702.
doi: <http://dx.doi.org/10.1211/0022357023619>.
- [25] McReynolds, K.D. and Gervay-Hagues, J. (2007) Chemotherapeutic interventions targeting HIV interactions with host-associated carbohydrates. *Chem. Rev.* **107**:1533–1552. 107:1533–1552.
doi: <http://dx.doi.org/10.1021/cr0502652>.
- [26] Ghose, A.K., Viswanadhan, V.N. and Wendoloski, J.J. (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* **1**:55–68.
doi: <http://dx.doi.org/10.1021/cc9800071>.
- [27] Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**:3–26.
doi: [http://dx.doi.org/10.1016/S0169-409X\(00\)00129-0](http://dx.doi.org/10.1016/S0169-409X(00)00129-0).
- [28] Liu, Q. and Brady, J.W. (1996) Anisotropic solvent structuring in aqueous solutions. *J. Am. Chem. Soc.* **118**:12276–12286.
doi: <http://dx.doi.org/10.1021/ja962108d>.
- [29] Gabius, H.J., Siebert, H.C., André, S., Jiménez-Barbero, J. and Rüdiger, H. (2004) Chemical biology of the sugar code. *ChemBioChem* **5**:740–764.
doi: <http://dx.doi.org/10.1002/cbic.200300753>.
- [30] Liu, F.T. and Rabinovich, G.A. (2010) Galectins: regulators of acute and chronic inflammation. *Ann. N.Y. Acad. Sci.* **1183**:158–82.
doi: <http://dx.doi.org/10.1111/j.1749-6632.2009.05131.x>.
-

- [31] Vyas, N.K. (1991) Atomic features of protein-carbohydrate interactions. *Curr. Opin. Struct. Biol.* **1**:732 – 740.
doi: [http://dx.doi.org/10.1016/0959-440X\(91\)90172-P](http://dx.doi.org/10.1016/0959-440X(91)90172-P).
- [32] Bains, G., Lee, R.T., Lee, Y.C. and Freire, E. (1992) Microcalorimetric study of wheat germ agglutinin binding to *N*-acetylglucosamine and its oligomers. *Biochemistry* **31**:12624 – 12628.
doi: <http://dx.doi.org/10.1021/bi00165a012>.
- [33] Logean, A., Sette, A. and Rognan, D. (2001) Customized versus universal scoring functions: application to class I MHC-peptide binding free energy predictions. *Biorg. Med. Chem. Lett.* **11**:675 – 679.
doi: [http://dx.doi.org/10.1016/S0960-894X\(01\)00021-X](http://dx.doi.org/10.1016/S0960-894X(01)00021-X).
- [34] Mammen, M., Choi, S.-K. and Whitesides, G.M. (1998) Polyvalent interactions in biological systems: Implications for design and use of multivalent ligands and inhibitors. *Angew. Chemie Int. Ed.* **37**(20):2754 – 2794.
doi: [http://dx.doi.org/10.1002/\(SICI\)1521-3773\(19981102\)37:20<2754::AID-AN-IE2754>3.3.CO;2-V](http://dx.doi.org/10.1002/(SICI)1521-3773(19981102)37:20<2754::AID-AN-IE2754>3.3.CO;2-V).
- [35] Meyer, E.A., Castellano, R.K. and Diederich, F. (2003) Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem. Int. Ed.* **42**:1210 – 1250.
doi: <http://dx.doi.org/10.1002/anie.200390319>.
- [36] Poveda, A., Asensio, J.L., Espinosa, J.F., Martin-Pastor, M., Cañada, J. and Jiménez-Barbero, J. (1997) Applications of nuclear magnetic resonance spectroscopy and molecular modeling to the study of protein-carbohydrate interactions. *J. Mol. Graph. Model.* **15**:9 – 17.
doi: [http://dx.doi.org/10.1016/S1093-3263\(97\)00012-0](http://dx.doi.org/10.1016/S1093-3263(97)00012-0).
- [37] Chervenak, M. and Toone, E. (1994) A direct measure of the contribution of solvent reorganization to the enthalpy of ligand binding. *J. Am. Chem. Soc.* **116**:10533 – 10539.
doi: <http://dx.doi.org/10.1021/ja00102a021>.
- [38] Spiwok, V., Lipovová, P., Skálová, T., Buchtelová, E., Hašek, J. and Králová, B. (2004) Role of CH/π interactions in substrate binding by *Escherichia coli* β-galactosidase. *Carbohydr. Res.* **339**:2275 – 2280.
doi: <http://dx.doi.org/10.1016/j.carres.2004.06.016>.
- [39] Bourne, Y., Rougé, P. and Cambillau, C. (1990) X-ray structure of a (α-Man(1 – 3)β-Man(1 – 4)GlcNAc)-lectin complex at 2.1 Å resolution. *J. Biol. Chem.* **265**:18161 – 18165.
-

- [40] Clarke, C., Woods, R.J., Gluska, J., Cooper, A., Nutley, M.A. and Boons, G.-J. (2001) Involvement of water in carbohydrate-protein binding. *J. Am. Chem. Soc.* **123**:12238 – 12247.
doi: <http://dx.doi.org/10.1021/ja004315q>.
- [41] Çarçabal, P., Jockusch, R.A., Hünig, I., Snoek, L.C., Kroemer, R.T., Davis, B.G., Gamblin, D.P., Compagnon, I., Oomens, J. and Simons, J.P. (2005) Hydrogen bonding and cooperativity in isolated and hydrated sugars: Mannose, galactose, glucose, and lactose. *J. Am. Chem. Soc.* **127**:11414 – 11425.
doi: <http://dx.doi.org/10.1021/ja0518575>.
- [42] Rarey, M., Kramer, B. and Lengauer, T (1999) The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins* **34**:17 – 28.
doi: [http://dx.doi.org/10.1002/\(SICI\)1097-0134\(19990101\)34:1<17::AID-PROT3>3.0.CO;2-1](http://dx.doi.org/10.1002/(SICI)1097-0134(19990101)34:1<17::AID-PROT3>3.0.CO;2-1).
- [43] Woods, R.J. and Chappelle, R. (2000) Restrained electrostatic potential atomic partial charges for condensed phase simulations of carbohydrates. *J. Mol. Struct. (Theochem)* **527**:149 – 156.
doi: [http://dx.doi.org/10.1016/S0166-1280\(00\)00487-5](http://dx.doi.org/10.1016/S0166-1280(00)00487-5).
- [44] Dowd, M.K., French, A.D. and Reilly, P.J. (1994) Modeling of aldopyranosyl ring puckering with mm3(92). *Carbohydr. Res.* **264**:1 – 19.
doi: [http://dx.doi.org/10.1016/0008-6215\(94\)00185-5](http://dx.doi.org/10.1016/0008-6215(94)00185-5).
- [45] Homans, S.W. (1990) A molecular mechanical force field for the conformational analysis of oligosaccharides: comparison of theoretical and crystal structures of Man α 1 – 3Man β 1 – 4GlcNAc. *Biochemistry* **29**:9110 – 9118.
doi: <http://dx.doi.org/10.1021/bi00491a003>.
- [46] Woods, R.J., Dwek, R.A. and Edge, C.J. (1995) Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. GLYCAM_93 parameter development. *J. Phys. Chem.* **99**:3832 – 3846.
doi: <http://dx.doi.org/10.1021/j100011a061>.
- [47] Glennon, T.M. and Merz, K.M., Jr. (1997) A carbohydrate force field for AMBER and its application to the study of saccharide to surface adsorption. *J. Mol. Struct. (Theochem)* **395**:157 – 171.
doi: [http://dx.doi.org/10.1016/S0166-1280\(96\)04949-4](http://dx.doi.org/10.1016/S0166-1280(96)04949-4).
- [48] Momany, F.A. and Willett, J.L. (2000) Computational studies on carbohydrates: in vacuo studies using a revised AMBER force field, AMB99C, designed for α -(1 \rightarrow 4) linkages. *Carbohydr. Res.* **326**:194 – 209.
doi: [http://dx.doi.org/10.1016/S0008-6215\(00\)00042-2](http://dx.doi.org/10.1016/S0008-6215(00)00042-2).
-

- [49] Kirschner, K.N. Youngye, A.B., Tschampel, S.M. González-Outeiriño, J., Daniels, C.R. Foley, B.L. and Woods, R.J. (2008) GLYCAM06: a generalizable biomolecular force field. *carbohydrates. J. Comput. Chem.* **29**:622 – 655, 2008.
doi: <http://dx.doi.org/10.1002/jcc.20820>.
- [50] Bradbrook, G.M., Forshaw, J.R. and Pérez, S. (2000) Structure/thermodynamics relationships of lectinsaccharide complexes. the *Erythrina corallodendron* case. *Eur. J. Biochem.* **267**:4545 – 4555.
doi: <http://dx.doi.org/10.1046/j.1432-1327.2000.01505.x>.
- [51] Tempel, W., Tschampel, S. and Woods, R.J. (2002) The xenograft antigen bound to *Griffonia simplicifolia* lectin 1-B₄. *J. Biol. Chem.* **8**:6615 – 6621.
doi: <http://dx.doi.org/10.1074/jbc.M109919200>.
- [52] Liang, G., Schmidt, R.K., Yu, H.-A., Cumming, D.A. and Brady, J.W. (1996) Free energy simulation studies of the binding specificity of mannose binding protein. *J. Phys. Chem.* **100**:2528 – 2534.
doi: <http://dx.doi.org/10.1021/jp952911e>.
- [53] Pathiaseril, A. and Woods, R.J. (2000) Relative energies of binding for antibody–carbohydrate-antigen complexes computed from free-energy simulations. *J. Am. Chem. Soc.* **122**:331 – 338.
doi: <http://dx.doi.org/10.1021/ja9914994>.
- [54] Bryce, R.A., Jillier, I.H. and Naismith, J.H. (2001) Carbohydrate-protein recognition: molecular dynamics simulations and free energy analysis of oligosaccharide binding to concanavalin A. *Biophys. J.* **81**:1373 – 1388.
doi: [http://dx.doi.org/10.1016/S0006-3495\(01\)75793-1](http://dx.doi.org/10.1016/S0006-3495(01)75793-1).
- [55] Spiwok, V., Lipovová, P., Skálová, T., Vondráčková, E., Dohnálek, J., Hašek, J. and Králová, B. (2006) Modelling of carbohydrate-aromatic interactions: *ab initio* energetics and force field performance. *J. Comput.-Aided. Mol. Des.* **19**:887 – 901.
doi: <http://dx.doi.org/10.1007/s10822-005-9033-z>.
- [56] Vandebussche, S., Díaz, D., Fernández-Alonso, M.C., Pan, W., Vincent, S.P., Cuevas, G., Cañada, F.J. Jiménez-Barbero, J. and Bartik, K. (2008) Aromatic-carbohydrate interactions: an NMR and computational study of model systems. *Chemistry* **14**:7570 – 7578.
doi: <http://dx.doi.org/10.1002/chem.200800247>.
- [57] Murray, C.W., Baxter, C.A. and Frenkel, A.D. (1999) The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. *J. Comput.-Aided Mol. Des.* **13**:547 – 562.
doi: <http://dx.doi.org/10.1023/A:1008015827877>.
-

- [58] Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**:727–748.
- [59] Wiley, E.A., MacDonald, M., Lambropoulos, A., Harriman, D.J. and Deslongchamps, G. (2006) LGA/EM-Dock – exploring Lamarckian genetic algorithms and energy-based local search for ligand-receptor docking. *Can. J. Chem.* **84**:384–391.
doi: <http://dx.doi.org/10.1139/V06-012>.
- [60] Thomsen, R. and Christensen, M.H. (2006) MolDock: a new technique for high-accuracy molecular docking. *J. Med. Chem.* **49**:3315–3321.
doi: <http://dx.doi.org/10.1021/jm051197e>.
- [61] Chen, H.M., Liu, B.F., Huang, H.L., Hwang, S.F. and Ho, S.Y. (2007) SODOCK: swarm optimization for highly flexible protein-ligand docking. *J. Comput. Chem.* **28**:612–623.
doi: <http://dx.doi.org/10.1002/jcc.20542>.
- [62] Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**:1639–1662.
doi: [http://dx.doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14<1639::AID-JCC10>3.0.CO;2-B](http://dx.doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B).
- [63] Bursulaya, B.D., Totrov, M., Abagyan, R. and Brooks, C.L. III. (2003) Comparative study of several algorithms for flexible ligand docking. *J. Comput.-Aided. Mol. Des.* **17**:755–763.
doi: <http://dx.doi.org/10.1023/B:JCAM.0000017496.76572.6f>.
- [64] Coutinho, P.M., Dowd, M.K. and Reilly, P.J. (1997) Automazed docking of glucosyl disaccharides in the glucoamylase active site. *Proteins* **28**:162–173, 1997.
doi: [http://dx.doi.org/10.1002/\(SICI\)1097-0134\(199706\)28:2<162::AID-PROT5>3.3.CO;2-0](http://dx.doi.org/10.1002/(SICI)1097-0134(199706)28:2<162::AID-PROT5>3.3.CO;2-0).
- [65] Coutinho, P.M., Dowd, M.K. and Reilly, P.J. (1998) Automazed docking of α -(1,4)- and α -(1,6)-linked glucosyl trisaccharides in the glucoamylase active site. *Ind. Eng. Chem. Res.* **37**:2148–2157.
doi: <http://dx.doi.org/10.1021/ie9706976>.
- [66] Minke, W.E., Diller, D.J., Hol, W.G. and Verlinde, C.L. (1999) The role of waters in docking strategies with incremental flexibility for carbohydrate derivatives. *J. Med. Chem.* **42**:1778–1788.
doi: <http://dx.doi.org/10.1021/jm980472c>.
-

- [67] Laederach, A. and Reilly, P.J. (2003) Specific empirical free energy function for automated docking of carbohydrates to proteins. *J. Comput. Chem.* **24**:1748–1757. doi: <http://dx.doi.org/10.1002/jcc.10288>.
- [68] Hill, A.D. and Reilly, P.J. (2008) A Gibbs free energy correlation for automated docking of carbohydrates. *J. Comput. Chem.* **29**:1131–1141. doi: <http://dx.doi.org/10.1002/jcc.20873>.
- [69] Kerzmann, A., Neumann, D. and Kohlbacher, O. (2006) SLICK – scoring and energy functions for protein-carbohydrate interactions. *J. Chem. Inf. Model.* **46**:1635–1642. doi: <http://dx.doi.org/10.1021/ci050422y>.
- [70] Kerzmann, A., Fuhrmann, J., Kohlbacher, O. and Neumann, D. (2008) BALLDock/SLICK: a new method for protein-carbohydrate docking. *J. Chem. Inf. Model.* **48**:1616–1625. doi: <http://dx.doi.org/10.1021/ci800103u>.
- [71] Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**:5179–5197. doi: <http://dx.doi.org/10.1021/ja00124a002>.
- [72] Terraneo, G., Potenza, D., Canales, A., Jiménez-Barbero, J., Baldrige, K.K. and Bernardi, A. (2007) A simple model system for the study of carbohydrate-aromatic interactions. *J. Am. Chem. Soc.* **129**:2890–2900. doi: <http://dx.doi.org/10.1021/ja066633g>.
- [73] Trott, O. and Olson, A.J. (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**:455–461.
- [74] Fuhrmann, J., Lenhof, H.P., Rurainski, A. and Neumann, D. (2010) A new Lamarckian genetic algorithm for flexible ligand-receptor docking. *J. Comput. Chem.*, page in press.
-