BALLDock/SLICK: A New Method for Protein-Carbohydrate Docking

Andreas Kerzmann,[†] Jan Fuhrmann,[‡] Oliver Kohlbacher,[†] and Dirk Neumann*[,‡]

Division for Simulation of Biological Systems, Center for Bioinformatics, University of Tübingen, Sand 14, 72076 Tübingen, Germany, and Junior Research Group Drug Transport, Center for Bioinformatics, Saarland University, Building E1 1, 66123 Saarbrücken, Germany

Received March 25, 2008

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 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.

1. INTRODUCTION

Protein-carbohydrate interactions are gaining increasing attention in pharmaceutical research and drug design. They are known to play a pivotal role in many important biological processes. For example, protein-carbohydrate interactions affect cell growth, differentiation, and aggregation. One class of proteins, the so-called lectins, are of special interest in this context, because they are known to bind sugars with very high specificity. There have been successful attempts of employing sugars and/or lectins in pharmaceutical applications. Drug targeting[1,2] and drug delivery systems[3] based on lectins have been presented. Sugars have been shown to enhance antitumor treatment in mouse models,[4] and a method for enhancing cancer treatment by coating particles with sugar has been proposed in 2005.[5] In the same year, Werz et al.[6] published a promising lead for an anthrax vaccine based on sugar oligomers.

Predicting binding mode (pose) and binding affinity of protein-carbohydrate complexes is thus an interesting problem for computer-aided drug design. A successful computational model of these interactions can be applied in virtual screening and lead optimization for carbohydrate-like compounds. For both applications, ligand docking is the most common method. Docking methods hinge on two key components: a search algorithm for sampling the conformational space and an energy or scoring function employed for assessing the binding energy. Numerous algorithms and scoring functions have been proposed to this end, and their performance has been evaluated in various comparative studies[7−16] The search methods commonly employed may be divided into several classes:[17,18] fast shape matching,[19,20] Monte Carlo search,[21,22] simulated annealing,[23] distance geometry,[24] tabu search,[25] evolutionary algorithms,[26−30] or incremental construction.[31,32] During the docking calculations the interaction between ligand and receptor is usually assessed by a single scoring or energy function. Both scoring and energy functions aim at estimating the binding free energy from the sum of several contributions.[33]

Scoring functions are usually employed to screen large ligand databases for possible interactions with a given target and to rank the solutions. Therefore, some contributions to the binding free energy which are time-consuming to calculate are often deliberately omitted or approximated to faster evaluate the scoring function. Energy functions, in contrast, try to calculate the binding free energy much more accurately by including such terms. Since there are many different ways to partition the binding free energy,[34] scoring and energy functions may differ considerably in their individual design. Docked conformations identified with a single scoring function are often rescored by another single function or even a combination of two or more functions.[35−37]

There is a plethora of different docking codes described and available, and we will thus discuss only those relevant to this work in some detail here. AutoDock[29] employs a genetic algorithm for sampling the conformational space, an approach that has been used widely.[26,28,35] The AutoDock energy function accounts for explicit hydrogen bonding and includes an atom-pair based term for calculating the changes in the desolvation free energy. FlexX[31] is a program for flexible ligand docking with the receptor kept rigid throughout the computations. FlexX decomposes the ligand into rigid fragments which are then incrementally reassembled in the binding site. A very fast multigreedy heuristic tries to find good poses based on an empirical scoring function accounting for hydrogen bonds, salt bridges, aromatic interactions, lipophilic interactions, and entropic contributions. HADDOCK[38,39] is a docking program initially developed for constrained docking (i.e., docking incorporating biophysical or biochemical information into the docking process). It uses a multistage protocol employing energy minimization and simulated annealing. The latest version of HADDOCK also supports

* Corresponding author e-mail: d.neumann@bioinf.uni-sb.de.
[†] University of Tübingen.
[‡] Saarland University.

10.1021/ci800103u CCC: $40.75
Published on Web 07/23/2008

flexible docking of carbohydrate ligands,[39] which has been successfully used to study the ligand specificity of chicken galectins.[40]

In contrast to protein-peptide or general protein−ligand interactions, there are some peculiarities to protein-carbohydrate interactions that render the docking of carbohydrate ligands difficult. Since these peculiarities have already been discussed extensively in our previous work,[41,42] we will mention them only summarily. Carbohydrate binding sites are often shallow compared to other binding pockets,[43] which renders approaches relying primarily on geometric complementarity unsuitable. Carbohydrate ligands are often unusually flexible,[44] and the large number of hydroxyl groups leads to a large number of hydrogen bonds that significantly contribute to the overall binding free energy[45] and−due to their polar nature−are important for the ligand interaction with the solvent. A key problem in the modeling of the interactions, however, is the CH-π interactions occurring between aromatic side chains of the protein and C−H bonds of the carbohydrate. These interactions have been identified as crucial to protein-carbohydrate interactions and are thus essential for a correct modeling of their binding mode.[46]

In this work, we present an approach designed in particular for protein-carbohydrate docking. Part of the approach are a scoring and an energy function for protein-carbohydrate interactions. This algorithm samples the conformational space of the ligand efficiently and also allows for receptor side-chain flexibility. The interaction between a carbohydrate and its receptor is evaluated using our SLICK scoring function,[42] which has been specifically designed for protein-carbohydrate interactions. SLICK consists of a fast and reliable scoring function, SLICK/score for filtering putative complex structures, and a computationally more demanding empirical energy function, SLICK/energy, for the final assessment of the candidate poses. The approach has been implemented in our freely available docking program BALLDock (part of our molecular modeling package BALL).[47] We have calibrated the scoring function on a comprehensive test set of known carbohydrate-lectin complexes and validate the accuracy of the docking approach on an independent test data set. We achieve excellent predictions for these complexes and compare these to results obtained by FlexX.

2. MATERIALS AND METHODS

To thoroughly sample the vast conformational space in the small ligand docking problem numerous search algorithms have been employed. Among the most popular ones are the so-called meta-heuristics, high level methods that can be specialized to solve optimization problems.[48]

These algorithms are independent of the actual optimization problem. Once implemented a meta-heuristic is able to optimize the switching of traffic lights, task schedules, or in our case the binding modes of molecules. The actual field of application is defined by its objective function, which transforms one instance of the optimization method to an instance of the optimization problem. It returns a fitness score which allows for the estimation of the quality of a solution. But not only this modularity accounts for the widespread use of meta-heuristics but also their performance often compares favorably to more specialized algorithms. In the case of molecular docking, applications using meta-heuristics

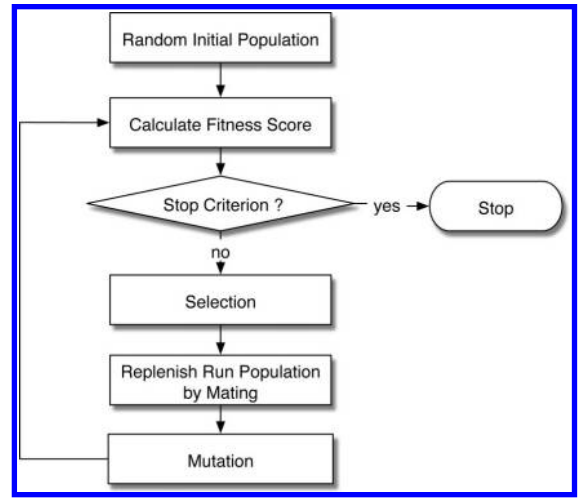

Figure 1. Flowchart of a genetic algorithm.

(e.g., GOLD[28] and AutoDock)[49] perform as well as comparable algorithms implemented in packages like Dock[50] and FlexX.[31] In this work, we use one of these meta-heuristics, a genetic algorithm, similar to the one used in AutoDock.

2.1. Genetic Algorithm. Within the family of meta-heuristics, genetic algorithms belong to the subclass of evolutionary algorithms. Common to all members of this subclass is the approach to imitate the principles of Darwinian evolutionary theory, particularly natural selection and reproduction. Consequently, another second common attribute is the existence of a *population*, a pool of individuals, each representing one possible solution for the problem at hand. A genetic algorithm uses a set of genetic operations to drive this population toward better solutions iteratively. The algorithm starts with the creation of an initial population which, in order to generate a very diverse set of conformations is often larger in size than subsequent populations (Figure 1). In the next step each individual is assigned a fitness score which is used to discard the worst (unfittest) members of the pool and to select the best individuals for creating progeny. Individuals which qualify to produce offspring are subject to *mating* to replenish the pool by producing new individuals, whereas *mutation* may modify existing individuals. To conserve the current best solutions *elitism* is being applied, which means that a number of top ranked individuals are protected from mutation. These steps are repeated until a preset number of iterations has been reached or until a predefined convergence criterion has been satisfied.

2.2. Encoding. When applying genetic algorithms to the problem of ligand docking, each individuum stands for a ligand pose, and this has to be encoded by set of genes. In our implementation the chromosome of each individuum is made up from a set of three genes which describe translation (translation gene) and rotation of the whole ligand molecule (quaternion gene) as well as torsions around single bonds (torsion genes) for flexible ligands. The gene for translation is composed of three independent real values, while the gene for the rotatation contains one real-value for each rotatable bond. The rotation gene is made up of a quaternion containing four interdependent real values (Figure 2). The *real-valued genes* and the *quaternion genes* assume values from the intervals [0, 1] and [−1, 1], respectively.

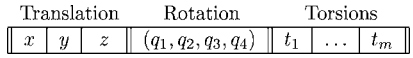

Figure 2. Ligand chromosome employed in our implementation. Each gene is made up of several values, e.g. translation is encoded by three independent values.

Table 1. Options Used for Docking with the Genetic Algorithm

parameter	value
iterations	600−1000
initial population	5000
pool size	200−400
survival rate	0.5
mutation rate	0.05
elitism	1

These types of genes differ the way they are mutated and mated. Mating of two real values a and b yields a uniformly distributed number c for the offspring as follows

$$c = (b + a)/2 + x(b - a) \quad (1)$$

where x is a uniformly distributed random number with $-1 \leq x \leq 1$. By employing a factor of 2, we allow for offspring values outside the range of the parents' values.

Real-valued genes are being considered cyclic, i.e. increasing a single value beyond the upper limit of one will result in values close to zero again and vice versa. It is thus possible to describe torsion values in a straightforward manner. Translations are constrained to a three-dimensional box whose position and size is predefined by the user and which usually centers on the putative binding site but may also encompass the whole receptor molecule.

Quaternion genes represent orientations in three-dimensional space. The gene consists of four values between −1.0 and 1.0. The four-dimensional vector representing the gene is normalized to unit length. Mating two quaternion genes requires an interpolation between their values that produces again a unit-length vector. In our implementation we use eq 1 for every single value of the quaternion followed by a normalization of the new quaternion gene. Thus, the offspring's quaternion gene is selected from a four-dimensional cube spanned by the parent quaternions. This represents a modification of the linear interpolation algorithm by J. Kuffner.[51]

To calculate the fitness score an individual's genes are mapped to the ligand's state variables, that is the values encoded in the genotype are transformed to the specific positions of the ligand's atoms relative to the receptor. This phenotype is then used to calculate the binding free energy (see below).

The user may alter the parameters for the genetic algorithm, like initial population size, pool size, number of populations, elitism, mutation, and survival rate. The values for those parameters used in this study are listed in Table 1.

The genetic docking algorithm described above, BALLDock, has been implemented using our molecular modeling framework BALL.[47] The latest version of BALLDock contains various additional improvements (e.g., flexible receptor side chains) that are not described here as they have not been used in this study.

2.3. Scoring and Energetic Evaluation. Scoring and energetic evaluation of docking candidates are done using the SLICK package[42] for protein-carbohydrate interactions. It contains SLICK/score, a fast and reliable scoring function for ranking docked conformations of protein-carbohydrate complexes. SLICK/score consists of four terms, considering CH···π interactions, hydrogen bonds, a softened van der Waals term, and electrostatic interactions. The score S of a complex is given by

$$S = s_0 + s_{CH\pi}S_{CH\pi} + s_{hb}S_{hb} + s_{vdw}\Delta G_{vdw} + s_{es}\Delta G_{es}^{int} \quad (2)$$

Electrostatic interactions are calculated with the Coulomb model. The weights s_i of the different contribution were manually optimized.

The binding energy of protein-carbohydrate complexes is computed with the empirical energy function SLICK/energy. This function is very similar to SLICK/score but takes solvation effects into account as well. The analytical form of SLICK/energy is given by

$$\Delta G = c_0 + c_{CH\pi}S_{CH\pi} + c_{hb}S_{hb} + c_{vdW}\Delta G_{vdW} + c_{np}\Delta G_{solv}^{np} + c_{es}(\Delta G_{solv}^{es} + \Delta G_{int}^{es}) \quad (3)$$

where $S_{CH\pi}$, S_{hb}, and ΔG_{vdW} are the same terms as in SLICK/score. Electrostatic solvation effects and interactions are covered by the Jackson-Sternberg model.[52] Nonpolar solvation contributions are computed with models from scaled particle theory.[53] The coefficients c_i are obtained by fitting SLICK/energy to experimentally available data using multiple linear regression. Scoring and energy function are calibrated on a thoroughly researched set of plant lectins binding different sugars (cf. Table 2). Please refer to ref 42 for details on the calibration procedure.

The different scoring and energy terms employed by SLICK need different parameter sets. In the case of hydrogen bonds and CH···π bridges the positions of the atom centers are sufficient for the calculation using the parameters given in ref. 42. van der Waals and Coulomb electrostatics calculations were performed with Glycam2000a[54] parameters using the AMBER[55] implementation of BALL, which was modified to compute a softened van der Waals term. Nonpolar solvation effects were predicted using Bondi[56] radii. Electrostatics calculations for the Jackson-Sternberg model[52] use PARSE[57] parameters as far as applicable. Ligand charges had to be calculated with *ab initio* methods. For that purpose, ligands were built manually and optimized geometrically with the GAMESS package[58] using the RHF/3−21G method. Then, ligand charges were computed with the same software at the RHF 6−31G* level and included in SLICK calculations.

2.4. Integration of SLICK into BALLDock. For this study we integrated SLICK/score into BALLDock, using it as scoring function during the docking process. SLICK/energy is then applied to the results of the docking run in order to predict binding energies.

The original energy function of BALLDock consists of three contributions. It comprises van der Waals energy, electrostatics, and conformational energy of the ligand. Like SLICK/score, BALLDock uses a softened form of the AMBER implementation in BALL to calculate van der Waals energies. Electrostatics are calculated with the Coulomb's law. The conformational energy of the ligand is its AMBER94 energy calculated with the BALL implementation using the softened van der Waals term.

The original BALLDock energy function was extended by incorporating the missing energy terms. Because the hydrogen position is crucial to the calculation of hydrogen

Table 2. Data Set for Calibrating and Validating the Energy Function[a]

PDB ID	lectin	ligand	ΔG_{exp} [kJ/mol]
1J4U	AIA (*Artocarpus integrifolia* agglutinin)	Me-α-D-Man	−18.24
5CNA	ConA (concanavalin A)	Me-α-D-Man	−22.18
1GIC	ConA	Me-α-D-Glc	−19.25
1QDO	ConA	Me-3-O-(α-D-Man)-α-D-Man	−28.45
1QDC	ConA	Me-6-O-(α-D-Man)-α-D-Man	−22.18
1ONA	ConA	Me-3,6-di-O-(α-D-Man)-α-D-Man	−30.96
1DGL	DGL (*Dioclea grandiflora* lectin)	Me-3,6-di-O-(α-D-Man)-α-D-Man	−34.31
1AXZ	ECorL (*Erythrina corallodendron* lectin)	Gal	−18.20
1AX0	ECorL	GalNAc	−17.90
1AX1	ECorL	Lac	−18.80
1AX2	ECorL	LacNAc	−22.70
2BQP	PSL (*Pisum sativum* lectin)	D-Glc	−14.00
1BQP	PSL	D-Man	−16.60
1QF3	PNA (Peanut agglutinin)	Me-β-D-Gal	−16.96
2PEL	PNA	Lac	−17.76
1EHH	UDA (*Urtica dioica* agglutinin)	$(GlcNAc)_3$	−21.34
1EN2	UDA	$(GlcNAc)_4$	−23.43
1K7U	WGA (wheat germ agglutinin)	$(GlcNAc)_2$	−21.34

[a] Structures are taken from the PDB. The experimental binding free energies ΔG_{exp} were taken from the literature (see the Supporting Information for details).

Table 3. Data Set for Testing the Docking Program[a]

PDB ID	lectin	ligand	ΔG_{exp} [kJ/mol]
	Plant Lectins		
1KJ1	ASA (*Allium sativum* agglutinin)	α-D-Man	-
1KUJ	AIA	Me-α-D-Man	−17.3
1MVQ	CML (*Cratylia mollis* lectin)	Me-α-D-Man	-
1WBL	PTL (*Psophocarpus tetragonolobus* lectin)	Me-α-D-Gal	-
1FNZ	RPL (*Robinia pseudoacacia* bark lectin)	GalNAc	-
1C3M	Heltuba	Manα1−3Man	-
1GZC	ECL (*Erythrina crista-galli* lectin)	Lac	−19.9
1JOT	MPA (*Maclura pomifera* agglutinin)	Galβ1−3GalNAc	-
1PUM	VAA (*Viscum album* lectin)	Gal	-
1PUU	VAA	Lac	-
1HKD	PSL	Me-α-D-Glc	−16.0
1RIN	PSL	Man_3	-
1OFS	PSL	sucrose	-
	Nonplant Lectins and Sugar Binding Proteins		
1DIW	tetanus toxin	Gal	-
1GLG	chemotactic protein receptor	Gal	-
1K12	AAA (*Anguilla anguilla* lectin)	Fuc	-
1NL5	engineered maltose binding protein	Mal	-
2GAL	hGal-7 (human galectin-7)	Gal	-
4GAL	hGal-7	Lac	−19.3
5GAL	hGal-7	LacNAc	−18.4
1C1L	congerin I	Lac	-
1SLT	S-lectin	LacNAc	-

[a] Structures are taken from PDB. Missing values for binding free energies were not available in the literature.

bond and CH$\cdots\pi$ interactions, both terms were integrated into the docking part rather than using them during the creation of the energy grid.

2.5. Experimental Data. The data used for evaluating BALLDock/SLICK consist of two sets. The first data set consists of the 18 high-quality plant lectin complex structures used in calibrating SLICK[42] and is listed in Table 2. Experimental binding free energies are available for all complexes in this set. Because SLICK itself was calibrated on that set, the danger of biased results is evident. Consequently, a second, larger data set (listed in Table 3) is used for validating the docking method on a more general selection of lectins and sugar binding proteins. The data set contains plant and animal lectins. Furthermore, the variety of sugar ligands extends over the SLICK calibration set. Unfortunately, binding free energies are only known for five complexes of this set (cf. Table 3). The structural data of all sets stem from the Protein Data Bank.[59] Energies were taken from the literature. Please see the Supporting Information for a complete list of references for the experimental data used here.

Structural data were prepared for calculations by removing water molecules from the structures, adding missing hydrogen atoms, and optimizing the positions of hydrogens. Hydrogens of proteins were added employing the BALL[47] fragment database. The sugar hydrogens were added with the Molecular Operating Environment.[60] Selected ligands were cleaned up using the OpenEyes tools.[61]

2.6. Docking Parameters. After integrating SLICK/score into BALLDock, the calibration data set was docked.

Table 4. Results of Docking the Calibration Set with BALLDock/SLICK[a]

PDB ID	*n* -mer	ΔG_{exp} [kJ/mol]	R_{ftp}	d_{ftp} [Å]	ΔG_{ftp} [kJ/mol]	ΔE_{ftp} [kJ/mol]
1J4U	1	−18.24	1	0.54	−17.30	0.94
5CNA	1	−22.18	1	0.91	−20.33	1.85
1GIC	1	−19.25	1	0.48	−17.85	1.40
1QDO	2	−28.45	8	0.83	−22.05	6.40
1QDC	2	−22.18	2	0.80	−19.60	2.58
1ONA	3	−30.96	1	1.32	−28.56	2.40
1DGL	3	−34.41	2	1.43	−26.14	8.27
1AXZ	1	−18.20	1	0.47	−15.79	2.41
1AX0	1	−17.90	1	0.51	−21.06	3.16
1AX1	2	−18.80	1	0.86	−19.39	0.59
1AX2	2	−22.70	1	0.76	−19.12	3.58
2BQP	1	−14.00	1	0.72	−16.30	2.30
1BQP	1	−16.60	1	0.68	−15.65	0.95
1QF3	1	−16.96	1	0.52	−18.26	1.30
1EHH	3	−21.34	1	1.47	−32.09	10.75
1EN2	4	−23.43	1	1.17	−29.54	6.11
1K7U	2	−21.34	1	1.04	−27.47	6.13
mean	-	-	1.53	0.85	-	3.60

[a] Column *n*-mer denotes the number of sugar monomers in the ligand, ΔG_{exp} is the experimental binding free energy of the complex, R_{ftp} denotes the rank of the first true positive structure (RMSD < 1.5 Å), d_{ftp} is the RMSD of the first true positive, ΔG_{ftp} is the SLICK prediction of the binding free energy for the first true positive, and ΔE_{ftp} denotes the absolute energy difference between ΔG_{exp} and ΔG_{ftp}.

Docking runs were performed with a large initial population of 5,000 individuals, while the population size during the docking was limited to 200 chromosomes for monomers and to 400 for larger sugars. The mutation rate was set to 0.05. Docking was performed in a docking box which extended 8 Å from the bounding box of the ligand known from crystal structure. For every monomer complex, 600 runs were performed and analyzed. For larger sugars, the number of runs was raised to 1,000. The larger number of parameters for oligo-carbohydrates were necessary because of the high flexibility of the ligands along the glycosidic bonds and the resulting large conformational space. With these parameters, a sufficient coverage of the binding site was possible (data not shown).

2.7. Comparison of BALLDock/SLICK with FlexX. The first step in comparing FlexX with BALLDock/SLICK on protein-carbohydrate complexes was docking the calibration set of SLICK with FlexX release 2.0.2. In these docking experiments, FlexX standard parameters were used. Structures were taken from the PDB entries. The default protonation states of the amino acids were used as suggested by FlexX. Atomic parameters for the energy function were automatically assigned. Ring conformations of the sugar rings of the ligand were taken from the crystal structure instead of using CORINA conformations, which is reasonable given the very rigid nature of sugar-rings. Receptor binding sites were defined with the default method by using spheres of 6.5 Å around ligand atoms. The base fragment was chosen automatically as well as the placement of the base fragment in the binding site.

3. RESULTS

Table 4 summarizes the results of docking the calibration set with BALLDock/SLICK. The highest ranked candidate structure with an rmsd below 1.5 Å, which will be referred to as the *first true positive* (ftp), is in almost every case the top ranked structure. The average deviation of all candidates is only 0.85 Å, and the mean absolute error of the binding free energy for the first true positive is below 4 kJ/mol.

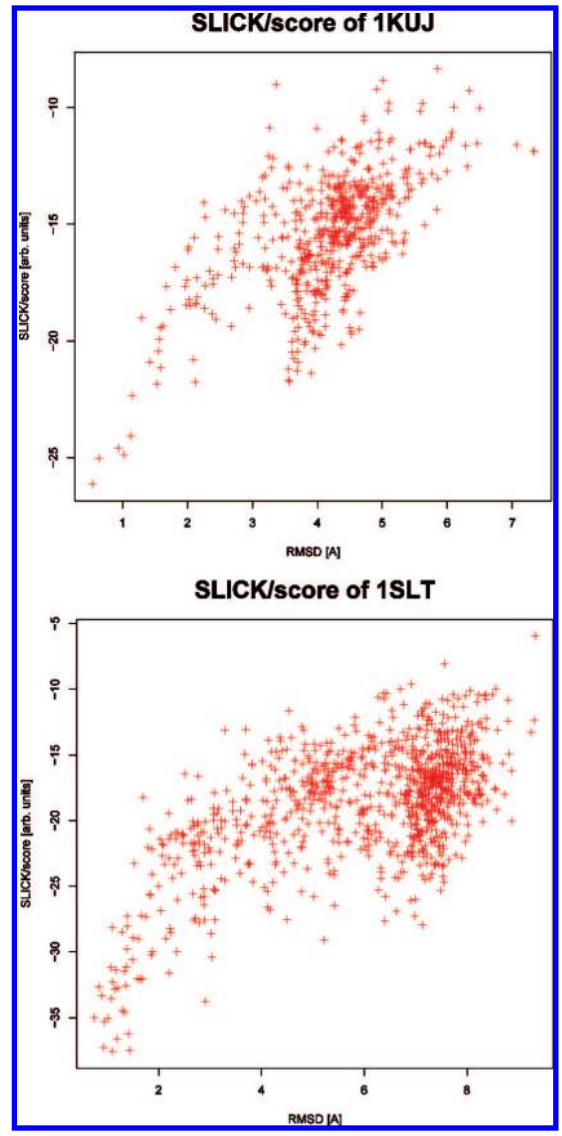

Figure 3. Results of BALLDock/SLICK docking runs: Complexes 1KUJ and 1SLT.

Figure 3 shows exemplary docking plots of BALLDock/SLICK results. All candidates with the exception of 2PEL show the correct tendency for scoring less deviating conformations better. But in almost every docking run, a cluster of numerous deviating structures is observable.

Analysis of the binding candidates produced by BALLDock/SLICK revealed the reasons for the deviating clusters. In the monomer case, one encounters many "rotated" rings. Compared to the crystal structure, these rings are rotated around the axis of symmetry of the ring plane. The scoring function is able to distinguish the correct pose from the incorrect one. The ConA/Me-Glc complex (1GIC) illustrates that very well. The scoring function identifies two highly scored clusters, one at about 1 Å rmsd and one at about 4 Å deviation. The latter cluster contains conformations with flipped rings and is thus correctly identified as worse than the native pose. Figure 4 shows an example of such flipped monomers.

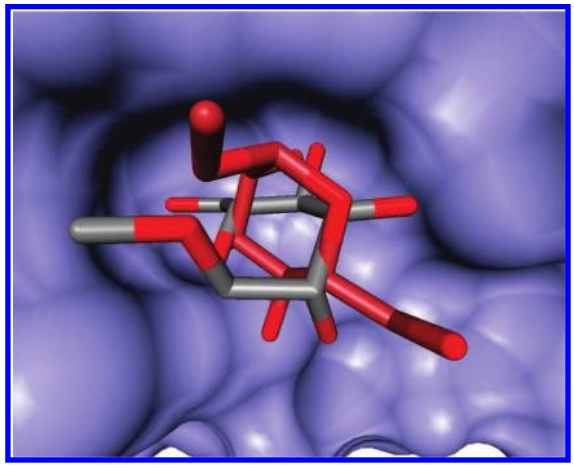

Figure 4. Rotated monomer ring (see text) in the binding site of ConA. The best ranked structure (normal colors) has an rmsd of below 1 Å, while the rmsd of the rotated monomer (reddish colors) is about 4 Å. Note the position of the ring oxygen. The figure was produced with BALLView[63,64] and POV-Ray.[65]

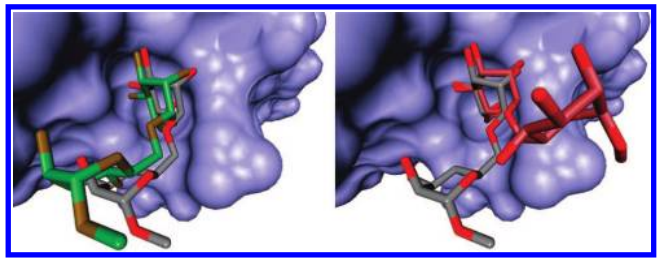

Figure 5. Mannose-dimer binding to ConA. Left: comparison between crystal structure (normal colors) and best scored docking candidate (greenish colors, rmsd 1.70 Å, pivotal ring rmsd 0.96 Å). Right: crystal structure (normal colors) and one of many conformations with the second ring rotated out of the binding site (reddish colors, rmsd 4.50 Å, pivotal ring rmsd 0.67 Å). The figure was produced with BALLView[63,64] and POV-Ray.[65]

In the case of dimers, clusters of large deviation are caused by sugar residues that extend into the solvent. Figure 5 shows an example for this behavior (ConA binding Me-3-O-(α-D-Man)-α-D-Man). One mannose ring is binding tightly into the binding site. The second mannose ring, which is surrounded by solvent, is sterically limited by the highly flexible glycosidic bond only. Analysis of the crystal structure shows that the second ring is mainly coordinated by water-mediated hydrogen bonds. This behavior is consistent with experimental findings[62] and seems to be typical for such complexes. BALLDock/SLICK creates many conformations with the tightly binding first ring in perfect position, while the pose of the second ring is rotated away from the binding mode. The high flexibility of the glycosidic bond connecting these two rings makes thorough scanning of the ligand's conformational space imperative.

There remains only one problematic case, which is PNA binding Lac (2PEL). This is surprising because other complexes with Lac or LacNAc seem to work well. Additionally, the second PNA complex in the calibration set (1QF3) gives perfect results. The ligand in 1QF3 is Me-Gal. In most Lac or LacNAc complexes, Gal is the ring that binds directly to the receptor, while the Glc ring extends into the solvent. Assuming this behavior in the PNA/Lac complex as well, it is not obvious why docking Lac with its Gal ring into the binding site of PNA should fail.

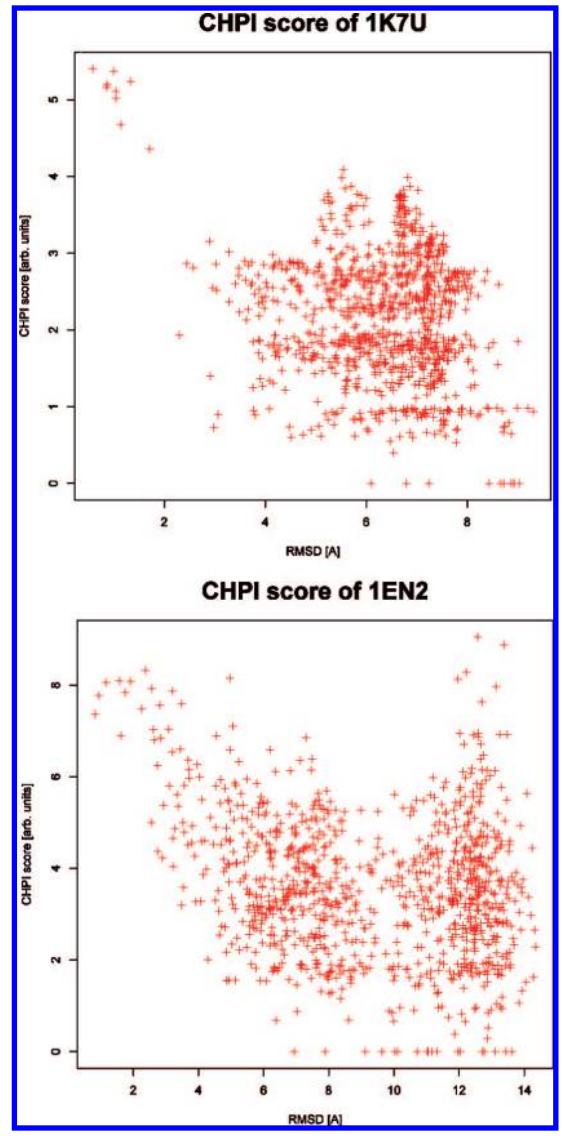

Figure 6. CH···π scores of WGA and UDA complexes.

A nice result is the good performance of the largest ligands in the calibration set. Docking a GlcNAc trimer and tetramer to UDA (complexes 1EHH and 1EN2) results in top ranked structures below 1.5 Å. The lowest deviations achieved are at 0.93 and 0.82 Å rmsd, respectively. These complexes are strongly influenced by CH···π interactions, which is reproduced by the model very well (cf. Figure 6).

3.1. Docking the Validation Set. For further assessment of the ability of BALLDock/SLICK to identify binding conformations of protein-carbohydrate complexes, a set of 22 lectins and sugar-binding proteins was docked. The docking set consists of two parts. One contains only plant lectins, while the other encompasses animal lectins and other sugar-binding proteins. For these complexes only a few binding free energies were found in the literature. Consequently, the analysis presented here focuses on structural features of the results. The results of these docking runs are summarized in Table 5.

The docking set not only covers nonplant lectins but also contains sugar ligands that were not included in the calibration set. The latter only contains sugars built from Man, Glc, and Gal monomers and their methylated and acetylated derivatives. In the docking set we find fructose, fucose, and maltose. While Man, Glc, and Gal are pyranoses in

Table 5. Comparison of BALLDock/SLICK and FlexX on the Extended Docking Set[a]

		BALLDock/SLICK				FlexX			
PDB ID	*n*-mer	R_{ftp}	d_{ftp} [Å]	R_{min}	d_{min} [Å]	R_{ftp}	d_{ftp} [Å]	R_{min}	d_{min} [Å]
				Plant Lectins					
1KJ1	1	1	0.92	11	0.46	10	1.41	17	1.37
1KUJ	1	1	0.54	1	0.54	1	0.94	2	0.53
1MVQ	1	1	0.51	4	0.29	1	0.67	81	0.59
1WBL	1	1	0.31	3	0.31	-	-	5	1.84
1FNZ	1	1	1.00	2	1.00	1	1.21	1	1.21
1C3M	2	1	0.27	1	0.27	1	1.25	5	0.43
1GZC	2	1	0.93	2	0.45	-	-	44	3.48
1JOT	2	1	1.43	12	0.93	25	1.25	197	0.91
1PUM	1	10	0.74	114	0.51	6	1.06	23	1.02
1PUU	2	17	0.43	29	0.40	1	1.07	14	0.98
1HKD	1	1	1.25	11	0.70	1	0.66	10	0.50
1RIN	3	2	0.86	3	0.44	9	1.39	9	1.39
1OFS	2	7	1.25	7	1.25	1	1.07	13	0.53
mean	-	3.46	0.80	15.38	0.58	5.18	1.09	32.38	1.14
				Nonplant Lectins					
1DIW	1	17	1.13	202	1.01	-	-	179	2.58
1GLG	1	1	0.42	75	0.23	1	1.14	23	0.50
1K12	1	3	0.74	7	0.21	3	0.67	3	0.67
1NL5	1	1	1.03	1	1.03	1	1.47	10	0.57
2GAL	1	(117)	(0.54)	152	0.26	-	-	17	1.97
4GAL	1	1	1.04	1	1.04	-	-	17	2.38
5GAL	1	2	1.33	34	1.20	-	-	-	-
1C1L	2	1	0.68	1	0.68	1	0.84	1	0.84
1SLT	2	1	1.10	8	0.75	1	1.44	1	1.44
mean	-	3.38	0.93	53.44	0.71	1.40	1.11	33.42	1.37
total mean	-	3.42	0.85	30.95	0.63	4.00	1.10	32.00	1.23

[a] Column *n*-mer denotes the size of the oligosugar, R_{ftp} is the rank of the first true positive (FTP), d_{ftp} is the RMSD of the FTP, R_{min} is the rank of the candidate with minimal RMSD, and d_{min} denotes the minimal RMSD. Docking runs with no true positive below rank 100 were counted as unsuccessful. Numbers are given in parentheses.

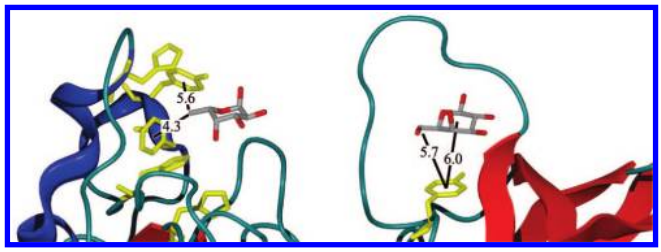

Figure 7. Left: binding site of 1K12 with many aromatic side chains. No CH$\cdots\pi$ interactions can be observed. Right: binding site of 1DIW. TYR 1124 is the only aromatic side chain in the vicinity, but it seems to contribute electrostatically only. A CH$\cdots\pi$ interaction cannot be built. Note the flexible coil above the sugar. Distances between ligand and aromatic side chains are given in Å.

Table 6. Results for Those Structures of the Test Set with Known Binding Affinity

PDB ID	ΔG_{exp} [kJ/mol]	R_{ftp}	d_{ftp} [Å]	ΔG_{ftp} [kJ/mol]	ΔE_{ftp} [kJ/mol]
1KUJ	−17.3	1	0.54	−22.46	5.18
1GZC	−19.9	1	0.93	−20.76	0.86
1HKD	−16.0	1	1.25	−18.88	2.88
4GAL	−19.25	1	1.04	−24.05	4.80
5GAL	−18.41	2	1.33	−21.18	2.77
mean	-	1.20	0.85	-	2.75

D-configuration, fructose is a furanose, and fucose is a deoxy-L-galactose; maltose is a Glc dimer. Since these ligands differ greatly from the calibration set, docking these sugars will permit the assessment of whether BALLDock/SLICK is able to cope with a broad range of sugar ligands or whether SLICK/score is biased too much by the choice of the calibration set. In addition, some of the proteins in the docking set have very deep binding pockets in contrast to the rather shallow binding sites of most plant lectins in the calibration set.

In this section, the structural analysis of the docking runs is of great importance, especially when it comes to nonplant lectins, because only on this basis insights for the further development of BALLDock/SLICK and SLICK in general are possible. Therefore, problem cases that occur in these docking runs are analyzed in depth in order to gain knowledge on the interactions and the errors made by BALLDock/SLICK.

For the 13 plant lectins, the docking runs confirm the results achieved with the calibration set. In nine cases, the first true positive is also the top ranked structure. For two complexes the first true positive is at least found among the top ten candidates. The mean rmsd of first true positives is only 0.8 Å. The two *Viscum album* lectin complexes (1PUM, 1PUU) seem to be harder to dock. Their lowest rmsd structures rank at 114 and 29, respectively, although in both cases candidates with very small deviations exist. In these cases, the docking plots reveal two highly scored regions. One is scoring good approximations of the binding mode well, while the other clusters at about 3.5 Å deviation. The latter cluster is dominated by electrostatic energy. Structural analysis shows that in the close vicinity of the ligands five aspartic acid side chains (ASP 23, 26, 27, 28, and 45) and one asparagine (ASN 47) are located. This could explain the strong electrostatic interactions dominating the scoring. Although there is a tryptophane in the binding site, which in the crystal structure clearly is participating in CH$\cdots\pi$

Table 7. Ranking and Calculated Energies for Candidates Created during FlexX Docking[a]

			FlexX				SLICK re-evaluation			
PDB ID	*n*-mer	ΔG_{exp} [kJ/mol]	R_{ftp}	d_{ftp} [Å]	ΔG_{ftp} [kJ/mol]	ΔE_{ftp} [kJ/mol]	R_{min}	d_{min} [Å]	ΔG_{min} [kJ/mol]	ΔE_{min} [kJ/mol]
1J4U	1	−18.24	1	0.76	−12.82	5.42	1	0.67	−17.71	0.53
5CNA	1	−22.18	1	0.89	−9.23	12.95	26	1.27	−21.17	1.01
1GIC	1	−19.25	1	0.84	−15.11	4.14	1	0.63	−19.36	0.11
1QDC	2	−22.18	1	1.21	−13.62	8.56	2	0.80	−23.46	1.28
1ONA	3	−30.96	3	1.47	−9.57	21.39	1	1.47	−29.91	1.05
1DGL	3	−34.41	1	0.75	−18.11	16.30	1	1.03	−27.89	6.52
1AX0	1	−17.90	32	1.22	−8.02	9.88	4	1.22	−18.12	0.22
2PEL	2	−17.76	70	1.39	−5.55	12.21	5	0.88	−24.16	6.40
1K7U	2	−21.34	6	1.19	−9.28	12.06	1	1.40	−32.38	11.04
mean	-	-	12.89	1.08	-	11.43	4.67	1.04	-	3.13

[a] Only complexes with candidates of low RMSD (<1.5 Å) are listed here. The candidates were docked with FlexX, and the generated structures were re-evaluated with SLICK. Column *n*-mer denotes the size of the sugar ligand, ΔG_{exp} is the experimental binding free energy, R_{ftp} is the rank of the first true positive (FTP), d_{ftp} is the RMSD of the FTP, ΔG_{ftp} is the calculated binding free energy for the FTP, and ΔE_{ftp} denotes the absolute difference between ΔG_{exp} and ΔG_{ftp}. See text for details.

interactions, the score for these interactions is too low to compensate for the strong electrostatic interaction.

Two docking runs produced top ranked structures with large deviations from the native binding mode. In the case of 1RIN this is an outlier in an otherwise perfect plot. In 1OFS, analyzing the data of each energy contribution showed that the false positives were dominated by van der Waals energies. Structural analysis revealed then that the dimeric ligand, sucrose, was docked with the wrong sugar ring into the binding site. Sucrose consists of fructose and glucose. In contrast to all other monomers encountered so far, fructose is a furanose, which means that the ring consists of four carbons and an oxygen instead of five carbons and an oxygen. Thus, fructose is smaller than glucose and fits better into the pocket sterically. Consequently, conformations which place fructose into the binding site are scored much better because of the drastically larger van der Waals energies. The crystal structure, however, reveals that in the bound conformation the glucose ring binds to the receptor while the fructose ring extends into the solvent.

In summary, the results for the plant lectin part of the docking set are very satisfactory. In comparison, the nonplant lectins results are slightly inferior. In five out of nine cases the bound conformation of the ligand is on the top rank. The mean values of d_{ftp} and d_{min} are 0.93 and 0.71 Å, respectively, which compare well to the plant lectin docking set.

In the case of 1K12 the tendency of the scoring function is correct, but the monomer is found in many twisted conformations. In addition, there are several aromatic groups in the vicinity of the binding site. Astonishingly, these groups seem to have little influence on the actual binding conformation, judging from the pose found in the crystal structure (Figure 7). Twisted rings that build CH$\cdots\pi$ bridges are thus scored higher than those resembling the bound state, which is compensated only partially by the other interactions.

The same seems true for the 1DIW complex (Figure 7). Here, a tyrosine is the only aromatic side chain in the binding site. But this side chain does rather contribute electrostatically than as a CH$\cdots\pi$ partner in the crystal structure. Located below the sugar ring is an aspartic acid that can act as a hydrogen bond acceptor. But in the crystal structure, the Gal seems to form hydrogen bonds to the backbone. Consequently, twisted conformations forming hydrogen bonds to the ASP are ranked very high, which makes it difficult for SLICK/score to identify the real binding pose. Additionally, the binding site is partially made up from a seemingly very flexible coil, shown in Figure 7. The question remains whether the crystal structure represents the only binding pose of Gal in this case.

3.2. Comparing BALLDock/SLICK to FlexX. One of the most popular docking programs is FlexX,[31] which uses an incremental construction algorithm for creating putative binding conformations. FlexX is known for its speed and often used in screening large substance libraries for possible binding candidates. But it is also known for very accurate predictions of binding conformations of drug-like substances, which is the reason for choosing this program for comparison purposes.

Table 7 shows the results of these docking attempts. As in all previous analyses, only solutions with a heavy-atom rmsd of below 1.5 Å were considered as true positives. FlexX was only able to create nine solutions of this quality. While most ligand conformations are at least in the vicinity of this limit, two oligomers (GlcNAc trimer and tetramer binding to UDA) are very far away from the native conformation with 5.7 Å and 9.4 Å rmsd, respectively. While the ranking of true positives is acceptable, the binding free energy estimates calculated by FlexX for the first true positives are deviating by 11.4 kJ/mol in the mean, which is according to our experience better than standard AutoDock (data not shown) but still does not reach the prediction quality of SLICK.

The results obtained from FlexX indicate that it performs well in ranking the lectin-sugar complexes as long as reasonable candidates are produced. In five of nine cases, the first true positive is also the best ranked conformation. However, in two cases (1AX0 and 2PEL) the rank of the near-native conformation is very poor. Judging from these results, the FlexX energy function seems to cope with protein-carbohydrate interactions only partially. The question remains whether the overall performance is dominated by the structure generator or the energy function.

In order to assess the influence of FlexX's energy function, the structures generated by FlexX were re-evaluated with SLICK, shown in Table 7. From the structures generated during the FlexX docking, the 200 best structures were taken. Again, from the nine candidates below 1.5 Å, the five first

true positives are at the same time the highest ranked structures. Interestingly, the complexes with that high prediction quality differ between the FlexX scoring and the SLICK scoring. Considering 5CNA, the FlexX score ranked a true positive at the top of the list, while SLICK/score only achieves rank 26. On the other hand, 1AX0 and 2PEL are still under the top ten according to SLICK/score, while FlexX ranks a true positive at 32 and 70, respectively. The average first true positive rank of the SLICK/score is at 4.7, which is considerably better than the mean rank of 12.9 achieved by FlexX. The biggest difference is found in the energy estimates. The energy difference between calculated and experimental values is at 3.1 kJ/mol when using SLICK/energy for energy calculations, which is well within the estimate obtained from the calibration of SLICK/energy.

Finally, the performance of FlexX was compared to BALLDock/SLICK on the docking set that was already employed earlier. Table 5 shows a summary of the results obtained by this comparison. Apparently, FlexX achieves significantly better results on the docking set than on the calibration set. Of the 22 structures of the docking set, 16 could be docked. The mean rank of the first true positive is 4.0, and the mean rmsd of the first true positive is as low as 1.1 Å. BALLDock/SLICK produces three more successful docking runs and provides a better ranking and deviation, but the differences are rather small.

4. DISCUSSION

In this article we have introduced the docking method BALLDock/SLICK designed for predicting the binding mode and binding affinity of protein-carbohydrate complexes. This method, which is based on the genetic docking method BALLDock and the SLICK package of scoring and energy functions for protein-carbohydrate interactions, yields high quality predictions both structurally and energetically. On the SLICK calibration set, 17 out of 18 complexes could be successfully redocked with an average rmsd of 0.85 Å and an average absolute error of 3.6 kJ/mol in the binding free energy estimate. On a larger and more diverse validation set, BALLDock/SLICK achieves very good results in 14 of 22 complexes, that is for 14 ligands the energetically best ranked solution featured an rmsd lesser than 1.5 Å.

We have also shown that BALLDock/SLICK outperforms a general docking program on both the calibration and the validation data. We have chosen FlexX for comparison, which is frequently used in drug design. In direct comparison on the calibration set, FlexX was only able to successfully dock 9 of the 18 plant lectin complexes. Using SLICK for re-evaluating the structures created by FlexX docking, the results improved noticeably, especially considering the binding free energy estimates.

On the validation set, FlexX achieved much better results but still does not reach the prediction accuracy of BALLDock/SLICK. With FlexX, 16 of the 22 protein-carbohydrate complexes could be predicted with reasonable accuracy. However, it has to be noted that FlexX was designed for general ligand docking, while the development of BALLDock/SLICK focused on protein-carbohydrate interactions. Due to limited availability of experimental data, binding free energy estimates was possible for only five of the complexes in the validation set.

Although the results of BALLDock/SLICK are encouraging, there still remain problems. Oligosugars are built from several rigid subunits connected by a very flexible glycosidic bond. When binding to proteins, one of the monomeric subunits tends to bind very tightly into the active site of the protein, whereas the remaining rings often extend into the solvent. These solvent-surrounded rings are frequently coordinated by water-mediated hydrogen bonds to the protein surface. Although BALLDock/SLICK is able to handle hydrogen bonds, indirect bonds are not treated explicitly at this stage.

Water-mediated H-bonds are still an unsolved problem in protein−ligand docking and have to be addressed through the structure generator of a docking program. A possible improvement of the current method is the introduction of explicit water molecules into the docking process. This problem will be addressed in future research. It is reasonable to assume that treatment of water-mediated hydrogen bonds will improve results noticeably.

ACKNOWLEDGMENT

The authors would like to thank the Deutsche Forschungsgemeinschaft for supporting J. Fuhrmann and D. Neumann (DFG grant BIZ 4/1). We gratefully acknowledge the use of software from OpenEye Scientific Software, Inc.

Supporting Information Available: References to structures and binding free energies used in the data set. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Wirth, M.; Fuchs, A. Lectin-mediated drug targeting: Preparation, binding characteristics and antiproliferative activity of wheat germ agglutinin conjugated doxorubicin in Caco-2 cells. *Pharm. Res.* **1998**, *15*, 1031–1037.

(2) Wirth, M.; Hamilton, G.; Gabor, F. Lectin-mediated drug targeting: Quantification of binding and internalization of wheat germ agglutinin and solanum tuberosum lectin using Caco-2 and HT-29 cells. *J. Drug Targeting* **1998**, *6*, 95–104.

(3) Clark, M. A.; Hirst, B. H.; Jepson, M. A. Lectin-mediated mucosal delivery of drugs and microparticles. *Adv. Drug Delivery Rev.* **2000**, *43*, 207–223.

(4) Hong, M. A.; Yan, J.; Baran, J. T.; Allendorf, D. J.; Hansen, R. D.; Ostroff, G. R.; Xing, P. X.; Cheung, N. K. V.; Ross, G. D. Mechanism by which orally administered β-1,3-glucans enhance the tumoricidal activity of antitumor monoclonal antibodies in murine tumor models. *J. Immunol.* **2004**, *173*, 797–806.

(5) von Bubnoff, A. Sugar coating improves anticancer treatment. Nature News [Online] 2005. http://dx.doi.org/10.1038/news050418-6 (accessed May 23, 2008).

(6) Werz, D. B.; Seeberger, P. H. Total synthesis of antigen bacillus anthracis tetrasaccharide - creation of an anthrax vaccine candidate. *Angew. Chem., Int. Ed.* **2005**, *44*, 6315–6318.

(7) Westhead, D. R.; Clark, D. E.; Murray, C. W. A comparison of heuristic search algorithms for molecular docking. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 209–228.

(8) Vieth, M.; Hirst, J. D.; Dominy, B. N.; Daigler, H.; Brooks, C. L., III. Assessing search strategies for flexible docking. *J. Comput. Chem.* **1998**, *19*, 1623–1631.

(9) Diller, D. J.; Verlinde, C. L. M. J. A critical evaluation of several global optimization algorithms for the purpose of molecular docking. *J. Comput. Chem.* **1999**, *20*, 1740–1751.

(10) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein-small molecule docking methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151–166.

(11) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L., III. Comparative study of several algorithms for flexible ligand docking. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 755–763.

(12) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L., III. Assessing energy functions for flexible docking. *J. Comput. Chem.* **1998**, *19*, 1612–1622.
(13) Ha, S.; Andreani, R.; Robbins, A.; Muegge, I. Evaluation of docking/scoring approaches: A comparative study based on MMP3 inhibitors. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 435–448.
(14) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035–1042.
(15) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins: Struct., Funct, Genet.* **2002**, *47*, 409–443.
(16) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–2303.
(17) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759–4767.
(18) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.
(19) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langdridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
(20) Perola, E.; Xu, K.; Kollmeyer, T. M.; Kaufmann, S. H.; Prendergast, F. G.; Pang, Y. P. Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J. Med. Chem.* **2000**, *43*, 401–4089.
(21) Caflish, A.; Fischer, S.; Karplus, M. Docking by Monte Carlo minimization with a solvation correction: application to an FKBP-substrate complex. *J. Comput. Chem.* **1997**, *18*, 723–743.
(22) Liu, M.; Wang, S. MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 435–451.
(23) Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins: Struct., Funct., Genet.* **1990**, *8*, 195–202.
(24) Essex, J. W.; Taylor, R. D.; Jewsbury, P. J. Flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *J. Comput. Chem.* **2003**, *24*, 1637–1656.
(25) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 367–382.
(26) Judson, R. S.; Jaeger, E. P.; Treasurywala, A. M. A genetic algorithm based method for docking flexible molecules. *J. Mol. Struct. (Theochem)* **1994**, *308*, 191–206.
(27) Clark, D. E.; Westhead, D. R. Evolutionary algorithms in computer-aided molecular design. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 337–358.
(28) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
(29) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
(30) Thomsen, R.; Christensen, M. H. MolDock: a new technique for high-accuracy molecular docking. *J. Comput. Chem.* **2003**, *24*, 1748–1757.
(31) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
(32) Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **1996**, *3*, 449–462.
(33) Gohlke, H.; Klebe, G. Statistical potentials and scoring functions applied to protein-ligand binding. *Curr. Opin. Struct. Biol.* **2001**, *11*, 231–235.
(34) Ajay; Murcko, M. A. Computational methods to predict binding free energy in ligand-receptor complexes. *J. Med. Chem.* **1995**, *38*, 4953–4967.
(35) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.
(36) Paul, N.; Rognan, D. ConsDock: a new program for the consensus analysis of protein-ligand interactions. *Proteins* **2002**, *47*, 521–533.
(37) Betzi, S.; Suhre, K.; Chetrit, B.; Guerlesquin, F.; Morelli, X. GFscore: a general nonlinear consensus scoring function for high-throughput docking. *J. Chem. Inf. Model.* **2006**, *46*, 1704–1712.
(38) Dominguez, D.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737.
(39) de Vries, S. J.; van Dijk, A. D. J.; Krzeminski, M.; van Dijk, M.; Thureau, A.; Hsu, V.; Wassenaar, T.; Bonvin, A. M. J. J. HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins: Struct., Funct., Bioinformatics* **2007**, *69*, 726–733.
(40) Wu, A. M.; Singh, T.; Liu, J.-H.; Krzeminski, M.; Russwurm, R.; Siebert, H.-C.; Bonvin, A.; Andre, S.; Gabius, H.-J. Activity-structure correlations in divergent lectin evolution: fine specificity of chicken galectin CG-14 and computational analysis of flexible ligand docking for CG-14 and the closely related CG-16. *Glycobiology* **2007**, *17*, 165–184.
(41) Neumann, D.; Lehr, C.-M.; Lenhof, H.-P.; Kohlbacher, O. Computational modeling of the sugar-lectin interaction. *Adv. Drug Delivery Rev.* **2004**, *56*, 437–457.
(42) Kerzmann, A.; Neumann, D.; Kohlbacher, O. SLICK - Scoring and energy functions for protein-carbohydrate interactions. *J. Chem. Inf. Model.* **2006**, *46*, 1635–1642.
(43) Taroni, C.; Jones, S.; Thornton, J. M. Analysis and prediction of carbohydrate binding sites. *Protein Eng.* **2000**, *13*, 89–98.
(44) Bohne, A.; Lang, E.; von der Lieth, C.-W. W3-SWEET: Carbohydrate modeling by internet. *J. Mol. Model.* **1998**, *4*, 33–43.
(45) Solis, D.; Fernandez, P.; Diaz-Maurino, J.; Jimenez-Barbero, J.; Martin-Lomas, M. Hydrogen-bonding pattern of methyl beta-lactoside binding to the Ricinus communis lectins. *Eur. J. Biochem.* **1993**, *214*, 677.
(46) Fernandez-Alonso, M.; Canada, F. J.; Jimenez-Barbero, J.; Cueva, G. Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions. *J. Am. Chem. Soc.* **2005**, *127*, 7379–7386.
(47) Kohlbacher, O.; Lenhof, H.-P. BALL - Rapid software prototyping in computational molecular biology. *Bioinformatics* **2000**, *16*, 815–824.
(48) Black, P. E. Metaheuristic. In Dictionary of Algorithms and Data Structures; Black, P. E., Ed.;U.S. National Institute of Standards and Technology: 2005.http://www.nist.gov/dads/HTML/metaheuristic.html (accesssed May 23, 2008).
(49) Laederach, H.-P.; Reilly, P. J. Specific empirical free energy function for automated docking of carbohydrates to proteins. *J. Comput. Chem.* **2003**, *24*, 1748–1757.
(50) Shoichet, B. K.; Bodian, D. L.; Kuntz, I. D. Molecular docking using shape descriptors. *J. Comput. Chem.* **1992**, *13*, 380–397.
(51) Kuffner, J. J. Effective sampling and distance metrics for 3D rigid body path planning. Proceedings of the IEEE Internationa Conference on Robotics and Automation, IEEE, 2004.
(52) Jackson, R. M.; Sternberg, M. J. E. A continuum model for protein protein interactions: Application to the docking problem. *J. Mol. Biol.* **1995**, *250*, 258–275.
(53) Pierotti, R. A. A scaled particle theory of aqueous and nonaqueous solutions. *Chem. Rev.* **1976**, *76*, 717–726.
(54) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. GLYCAM_93 parameter development. *J. Phys. Chem.* **1995**, *99*, 3832–3846.
(55) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(56) Bondi, A. Van der Waals volumes and radii. *J. Phys. Chem.* **1964**, *68*, 441–451.
(57) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **1994**, *98*, 1978–1988.
(58) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
(59) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(60) The Chemical Computing Group: Montreal, Canada, 2006. http://www.chemcomp.com (accessed May 23, 2008).
(61) The Chemical Computing Group: Santa Fe, NM, U.S.A., 2005. http://www.eyesopen.com (accessed May 23, 2008).
(62) Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G.-J. Involvement of water in carbohydrate-protein binding. *J. Am. Chem. Soc.* **2001**, *123*, 12238–12247.
(63) Moll, A.; Hildebrandt, A.; Lenhof, H.-P.; Kohlbacher, O. BALLView: An object-oriented molecular visualization and modeling framework. *J. Comput.-Aided Mol. Des.* **2006**, *19*, 791–800.
(64) Moll, A.; Hildebrandt, A.; Lenhof, H.-P.; Kohlbacher, O. BALLView: A tool for research and education in molecular modeling. *Bioinformatics* **2006**, *22*, 365–366.
(65) *Persistence of Vision Raytracer, version 3.6*; Persistence of Vision Pty. Ltd.: Williamstown, Victoria, Australia. http://www.povray.org/download/ (accessed May 23, 2008).

CI800103U