SLICK − Scoring and Energy Functions for Protein−Carbohydrate Interactions

Andreas Kerzmann,*,† Dirk Neumann,‡ and Oliver Kohlbacher†

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 36.1, 66123 Saarbrücken, Germany

Received September 25, 2005

Protein−carbohydrate interactions are increasingly being recognized as essential for many important biomolecular recognition processes. From these, numerous biomedical applications arise in areas as diverse as drug design, immunology, or drug transport. We introduce SLICK, a package containing a scoring and an energy function, which were specifically designed to predict binding modes and free energies of sugars and sugarlike compounds to proteins. SLICK accounts for van der Waals interactions, solvation effects, electrostatics, hydrogen bonds, and CH···π interactions, the latter being a particular feature of most protein−carbohydrate interactions. Parameters for the empirical energy function were calibrated on a set of high-resolution crystal structures of protein−sugar complexes with known experimental binding free energies. We show that SLICK predicts the binding free energies of predicted complexes (through molecular docking) with high accuracy. SLICK is available as part of our molecular modeling package BALL (www.ball-project.org).

1. INTRODUCTION

A multitude of biologically important processes such as cell growth, differentiation, cell aggregation, and immunology rely on protein−carbohydrate interactions. These interactions have thus been studied in detail and have been used to address numerous biomedical applications. Lectins, a family of carbohydrate-binding proteins, have repeatedly been employed as mediating or functionalizing components in drug targeting[1,2] and drug-delivery[3] systems. Moreover, endogenous lectins can be targets for drugs[4] themselves. Sugars have been shown to enhance antitumor treatment in mouse models,[5] and a method for enhancing cancer treatment by coating particles with sugar was recently reported.[6] Consequently, a reliable method for the prediction of protein−carbohydrate binding would be very helpful in the understanding these processes and in the rational design of novel drugs based on sugars or drug-delivery systems based on lectins.[7]

Experimental studies report that lectin−sugar interactions[8] are strongly influenced by CH···π interactions, which are not covered by energy functions of general docking tools such as AutoDock,[9] FlexX,[10] or Gold.[11] Such interactions are now being incorporated into molecular force fields[12] and gaining more attention in molecular modeling. Previous efforts for predicting protein−carbohydrate complexes fitted existing functions to protein−carbohydrate calibration sets[13] or focused on reproducing geometric features of the structures[14,15] without incorporating new models for such interactions. This study will introduce a CH···π term into the calculation.

While the general importance of sugar-binding proteins is obvious, the domain of plant lectins offers new possibilities for drug-delivery systems. These lectins are very well researched and generally very stable against heat and digestion.[16,17] These features make plant lectins valuable for pharmaceutical applications where prodrugs or functionalized drug carrier systems have to survive the gastro-intestinal tract while retaining their function.

Molecular docking increasingly becomes a valuable tool for exploring pharmaceutically interesting targets. Docking programs address the following problem: Given the spatial structure of two individual molecular entities, calculate the binding geometry and binding free energy of the complex formed by these molecules. There are numerous approaches available for molecular docking (see, e.g., ref 18 for a review). The main concept behind most docking programs is the following (see Figure 1): Starting with the spatial structures of the binding molecules, a structure-generation step produces a great number of putative complex structures. These many complexes are then filtered by suitable fast filters or scoring functions in order to reduce the number of candidates. The top-scored structures are then evaluated energetically with an accurate energy function, thus yielding approximations of the binding mode and binding free energy of the complex.

Lectin−sugar interactions differ from general protein−ligand interactions in a number of features, making them hard to predict with standard tools. In addition to the aforementioned CH···π interaction, the large number of freely rotatable polar hydroxyl groups provides for increased hydrogen bonding and strong electrostatic interactions. Sugar binding sites are often rather shallow binding grooves,[19] making their identification difficult for ligand docking programs because the geometric complementarity of the ligand and the pocket is less pronounced. The glycosidic bond

* Corresponding author e-mail: kerzmann@informatik.uni-tuebingen.de.
† University of Tübingen.
‡ Saarland University.

10.1021/ci050422y CCC: $33.50
Published on Web 05/20/2006

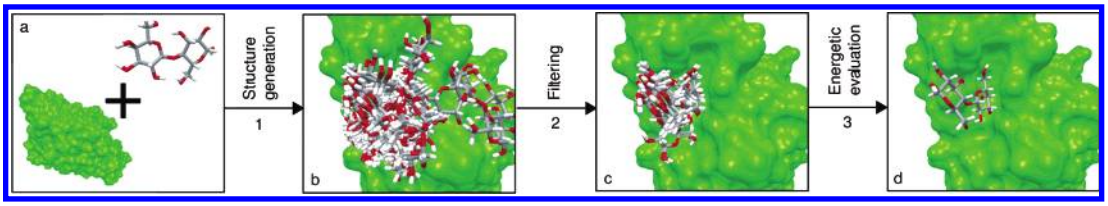

Figure 1. Schematic picture of the docking process.

of sugar oligomers is very flexible without pronounced rotameric states,[20] resulting in a huge conformational space. In addition, the interactions governing the protein−sugar binding are often dominated by solvation effects, which are largely caused by the polar nature of the many hydroxyl groups and are notoriously hard to model.

The goal of this work is the design of an empirical energy and scoring function suitable for the prediction of lectin−carbohydrate interactions in molecular docking environments. Because docking methods frequently employ a fast scoring function in the structure-generation process and a slower but more accurate energy function for predicting the binding free energy of the docked complex, we chose to separate our approach in the same way. We call this new set consisting of an energy and a scoring function SLICK, for **S**ugar−**L**ectin **I**nteractions and Do**CK**ing. To discriminate the scoring function from the energy function, the scoring function will be called SLICK/score while the energy function will be called SLICK/energy.

This article is structured as follows: First, we will detail the theoretical models and introduce experimental data we used in modeling and validating SLICK. Then, we will give results of applying the scoring and energy functions and a statistical analysis of these results. After that, a discussion of the results and an outlook into future improvements conclude this article.

2. MATERIALS AND METHODS

The computational models used by our scoring and energy functions cover interactions which prove to be very important for the prediction of binding modes and affinities of lectin−carbohydrate complexes. The next sections summarize the different components of our energy function and their incorporation into SLICK, followed by a description of the structures and docking methods used in this study.

The scoring function SLICK/score, which we use for rescoring structures generated by a docking program, is the weighted sum of four contributions. It consists of hydrogen bond and CH···π interactions, a softened van der Waals term, and electrostatic interactions based on Coulomb's law. The score S is given as

$$S = s_0 + s_{\mathrm{CH}\pi} S_{\mathrm{CH}\pi} + s_{\mathrm{hb}} S_{\mathrm{hb}} + s_{\mathrm{vdw}} \Delta E_{\mathrm{vdw}} + s_{\mathrm{es}} \Delta E_{\mathrm{es}}^{\mathrm{Coulomb}} \quad (1)$$

The computational effort needed for this scoring function is rather low and thus suited for structure-generation purposes within docking programs. The coefficients for SLICK/score were optimized manually on the calibration set (see below).

The energy function SLICK/energy for estimating ΔG is a weighted sum of five contributions and comprises hydrogen bonding and CH···π interactions between the protein and ligand, a softened van der Waals term, and a full solvation term including electrostatic interactions between the protein and ligand and the nonpolar solvation contribution.

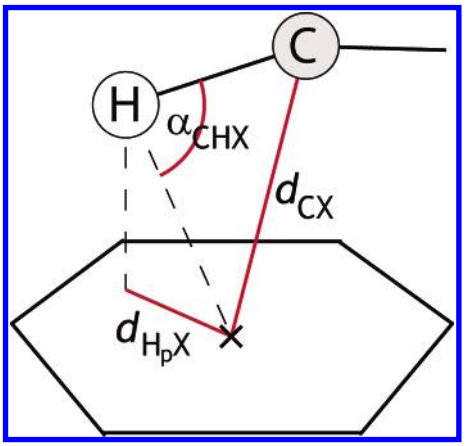

Figure 2. Geometry of the CH···π interaction: d_{CX} is the distance between the aliphatic carbon and the ring center, α_{CHX} denotes the angle between the CH bond and the connecting line from hydrogen to the ring center, and d_{H_pX} is the distance of H to the ring center projected into the ring plane.

$$\Delta G = c_0 + c_{\mathrm{CH}\pi} \Delta G_{\mathrm{CH}\pi} + c_{\mathrm{hb}} \Delta G_{\mathrm{hb}} + c_{\mathrm{vdW}} \Delta G_{\mathrm{vdW}} + c_{\mathrm{np}} \Delta G_{\mathrm{solv}}^{\mathrm{np}} + c_{\mathrm{es}} (\Delta G_{\mathrm{solv}}^{\mathrm{es}} + \Delta G_{\mathrm{int}}^{\mathrm{es}}) \quad (2)$$

Coefficients c_i are obtained by fitting calculated energies to experimental data employing multiple linear regression using the statistics package R.[21] The robustness of the energy function on a statistical level was assessed using cross-validation. We did a full leave-one-out test on the data set. Additionally, a randomized 5-fold cross-validation was performed, and the mean values over 1000 runs of the cross-validation procedure were analyzed.

2.1. CH···π Interactions. There are several models for CH···π interactions. We chose a simple geometric model by Brandl et al.[22] for our computations, which is very close to the geometric formulation of ordinary hydrogen bonds. Figure 2 shows the three values d_{CX}, α_{CHX}, and d_{H_pX}, which are considered by the CH···π model. d_{CX} denotes the distance between the center of the C atom and the center X of the aromatic ring. The angle α_{CHX} is measured between the CH bond and the connecting line between hydrogen atom H and ring center X. The last quantity d_{H_pX} is the distance of the hydrogen atom from the ring center projected into the ring plane.

Brandl et al. define limits for these three quantities. If all of the values are within the defined limits, a CH···π bridge is found. We score each of the three observables of this model on the basis of the original parameters and our own observations from crystallographic data. Our modified limits are $d_{CX} < 4.5$ Å, $\alpha_{CHX} > 110°$, and 0.7 Å $< d_{H_pX} <$ 1.7 Å.

A mere yes or no decision discriminating interactions by only considering hard limits is very susceptible to small

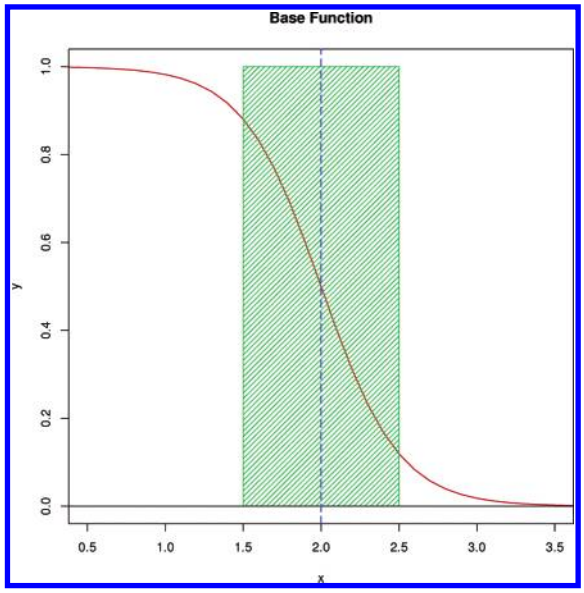

Figure 3. Switching function of the model. The slope of the function (red line) is calculated to switch smoothly using a tolerance interval (green area) around the limit (blue dashed line) given by the model. Because of the sigmoid nature of the base function, regions outside of the tolerance interval can still be scored with low scores, allowing for increased tolerance against small errors in data.

errors in the data. Therefore, we score these values by using a sigmoid function that switches smoothly from 0 to 1 depending on how far a quantity approaches the predefined limit. The analytical form of the CH···π term is based on the Fermi function

$$f(x,a,b) = \frac{1}{1 + \exp(-ax + b)} \tag{3}$$

The slope of this function is defined by the parameters a and b. For every observable, these parameters are calculated from their respective limits by using a tolerance interval. Figure 3 illustrates this definition. The total score $S_{CH\pi}$ of the CH···π contribution is the mean of the three individual scores:

$$S_{CH\pi} = \frac{1}{3}f(d_{CX},a_{CX},b_{CX}) + \frac{1}{3}f(\alpha_{CHX},a_{CHX},b_{CHX}) + \frac{1}{3}f(d_{H_pX},a^l_{H_pX},b^l_{H_pX})\, f(d_{H_pX},a^u_{H_pX},b^u_{H_pX}) \tag{4}$$

The parameters a_{CX}, b_{CX}, a_{CHX}, b_{CHX}, $a^l_{H_pX}$, $b^l_{H_pX}$, $a^u_{H_pX}$, and $b^u_{H_pX}$ represent the limits translated into parameters of the switching function.

Figure 4 shows a contour plot of the CH···π interaction of probe groups over the aromatic ring of phenylalanine. The probes were CH groups perpendicular to the ring plane.

2.2. Hydrogen Bonds. For scoring and evaluating hydrogen-bond interactions, we use a modified version of the hydrogen-bond model by Böhm[23] with the parametrization of Eldridge and co-workers.[24] Our modification turns the linear terms of the Böhm function into a smooth scoring function by employing the aforementioned sigmoid Fermi function. Scores for hydrogen bonds are defined as

$$S_{hb} = f(\Delta r,a^r_{hb},b^r_{hb})\, f(\Delta\alpha,a^\alpha_{hb},b^\alpha_{hb}) \tag{5}$$

with Δr being the deviation of the hydrogen-bond length from the ideal value of 1.85 Å and $\Delta\alpha$ being the deviation of the hydrogen-bond angle from the ideal value of 180°. The tolerance limits are defined as $a^r_{hb} = 0.25$ Å, $b^r_{hb} = 0.65$ Å, $a^\alpha_{hb} = 30°$, and $b^\alpha_{hb} = 80°$.

2.3. van der Waals Interactions. We compute the van der Waals interaction of the receptor and ligand with a modified softened Lennard-Jones potential. Softening the van der Waals potential can account for structural reorganization of the receptor to some extent[25] and makes the energy function less susceptible to small errors in ligand positioning. The repulsive part of the 6−12 term was softened by discarding individual contributions above a defined limit c_{vdw}. The van der Waals contribution was calculated with a modified version of our implementation of the AMBER force field.[26] Of the several strategies for softening the van der Waals potential, we chose this simple one as a starting point. The individual contribution e^{vdw}_{ij} by atoms i and j is defined as

$$e^{vdw}_{ij} = \frac{A^{ij}}{r_{ij}^{12}} - \frac{B^{ij}}{r_{ij}^{6}} \tag{6}$$

The coefficients A^{ij} and B^{ij} are parameters that are dependent on the types of atoms i and j and were taken from the Glycam2000a force field. The distance between atoms i and j is denoted by r_{ij}. The limit was chosen to be $c_{vdw} = 5$ kJ/mol individual contribution. The total van der Waals energy is given by

$$E_{vdw} = \sum \begin{cases} e^{vdw}_{ij} & \text{if } e^{vdw}_{ij} < c_{vdw} \\ c_{vdw} & \text{otherwise} \end{cases} \tag{7}$$

2.4. Electrostatics. Electrostatic interactions generally provide an important contribution to protein−ligand interactions. By approximating atoms as point charges, the electrostatic interactions of all atoms in a system of molecules can be determined by using Coulomb's law, which states that the electrostatic energy E_{es} of a system of point charges is

$$E_{es} = \frac{1}{4\pi\epsilon_0\epsilon_r}\sum_{i<j}\frac{q_iq_j}{r_{ij}} \tag{8}$$

with q_i and q_j being the charges of atoms i and j with distance r_{ij}. The constant ϵ_0 denotes the vacuum permittivity, while ϵ_r is the dielectric constant of the medium.

2.5. Solvation Effects. Electrostatics do not only contribute via direct interactions between the atoms of two molecules but also by means of solvation effects. Biological processes take place in physiological milieu, which essentially means water from the electrostatics point of view. Water is a highly polar solvent thus influencing the binding free energy through solvation effects.

Generally, solvation free energies can be calculated as the sum of polar contributions to ΔG_{solv} (resulting from electrostatic effects) and nonpolar contributions. The solvation free energy can thus be written as $\Delta G_{solv} = \Delta G^{es}_{solv} + \Delta G^{np}_{solv}$.

Electrostatic effects cannot be neglected in thorough energy calculations of protein−ligand interactions. This is

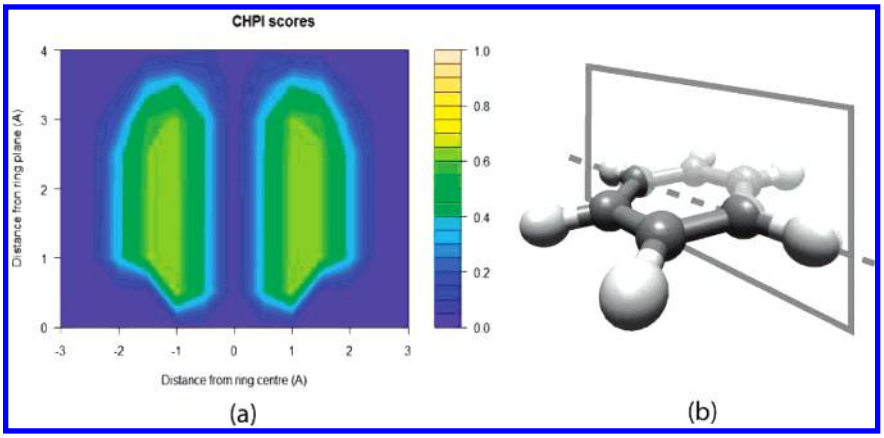

Figure 4. Cross section of the CH···π potential over the benzene ring of phenylalanine. Part a displays CH···π scores calculated for the plane indicated in part b. Scores were generated by sweeping the space with CH groups perpendicular to the ring plane.

especially true for lectin−sugar complexes, because sugar molecules do present a large number of polar hydroxyl groups.

We have chosen the solvation model of Jackson and Sternberg,[27] which was proven to achieve very good solvation free energies in both protein−protein and protein−ligand docking. This accuracy requires significantly more CPU time than the calculation of mere Coulomb interactions. Note that the Jackson−Sternberg model not only covers solvation free energies but also electrostatic interaction energies. Total electrostatic energy is given as $\Delta G^{es} = \Delta G^{es}_{solv} + \Delta G^{es}_{int}$.

We use a finite-difference Poisson−Boltzmann (FDPB) solver implemented within the BALL[28] framework to determine the electrostatic part of the solvation free energy. To keep computation times short, we only consider atoms within a rectangular bounding box around the ligand. This box extends 8 Å in every direction from the ligand. Our investigations have shown that this additional approximation does not reduce the overall quality of the energy calculation. FDPB calculations are done on a 0.5 Å grid with electrostatic focusing.[29] We use a relative dielectric constant of 80 for bulk water and harmonic smoothing of the dielectric constant on the grid.

Solvation free energies also depend on nonpolar effects which are of entropic as well as enthalpic nature. The main contribution of nonpolar models is the energy needed for creating a cavity in the solvent that contains the solvated molecule. Usually, these energies are calculated with the Uhlig model,[30] which is very simple and easy to calculate. A more sophisticated approach is the scaled particle theory by Reiss and Tully-Smith,[31] which we implemented in the formulation of Pierotti.[32]

Our nonpolar solvation model can be written as

$$\Delta G^{np}_{solv} \approx \Delta G_{cav} = \frac{1}{4\pi}\sum_i \frac{A_i}{r_i^2}\delta G_{cav}(r_i) \qquad (9)$$

where i denotes the index of a hard sphere with radius r_i, A_i is the molecular surface area occupied by that sphere, and $\delta G_{cav}(r_i)$ is the energy necessary for creating a spherical cavity of radius r_i in the solvent continuum. The molecular surface area is defined by the solvent-accessible surface of that molecule.

2.6. Docking. For this study, we chose AutoDock 3.0.5[9] as the docking program for testing the SLICK scoring and energy functions. AutoDock's docking strategy is based on a Lamarckian genetic algorithm which employs local optimization in its evolutionary strategy.

Ligands were docked to the protein using the standard AutoDock 3.0.5 parameters unless otherwise noted. The energy grids were centered on the geometric center of the respective ligand in the binding site. The grid dimensions were 65 × 65 × 65 points with a spacing of 0.375. Energies were scaled employing the free-energy model 140n coefficients. Nonpolar hydrogens were explicitly modeled. Each docking run started with a random population of 10 individuals per dihedral. The maximum number of energy evaluations during a single run was 30 000 per individual. For each carbohydrate, 200 runs were performed. The resulting final candidates of each run were then employed in the following calculations.

2.7. Experimental Data. The experimental data necessary for calibrating the energy function consist of structural data, that is, X-ray or NMR structures from a structural database, and experimentally determined binding affinities of carbohydrates bound to lectins. The latter is mainly provided through isothermal titration calorimetry (ITC) measurements reported in the literature.

The data available for investigating lectin−sugar interactions are very limited. While there are many structures of protein−carbohydrate complexes available in the Protein Data Bank (PDB),[33] there is only a very small set of data with reliable experimental binding energies available. Careful selection yielded a set of 20 complexes of lectins binding mono- and oligosugars with high-resolution crystal structures and high-quality ITC measurements. Table 1 lists all complexes of our data set. See the Supporting Information for their respective sources and further details.

2.7.1. Preparation of Crystal Structures. The crystal structures were prepared according to the following protocol: First, water molecules were removed from X-ray data and missing hydrogen atoms added. Then, atomic partial charges and atomic radii were assigned. Finally, hydrogen positions in the bound state were optimized.

Hydrogens were added using BALL for the protein and MOE[34] for the carbohydrate. Hydrogen positions were optimized with the Glycam2001a parameter set[35] for the AMBER force field implementation in BALL 1.1. Partial

Table 1. Data Set for Calibrating and Validating the Energy Function[a]

PDB ID	lectin	ligand	ΔG (kJ/mol)
	Calibration Set		
1J4U	AIA (*Artocarpus integrifolia* agglutinin)	Me-α-D-Man	−18.24
5CNA	ConA (concanavalin A)	Me-α-D-Man	−22.18
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
1AXZ	ECorL (*Erythrina corallodendron* lectin)	Gal	−18.2
1AX2	ECorL	LacNAc	−22.7
1BQP	PSL (pea lectin)	D-Man	−16.6
2PEL	PNA (peanut agglutinin)	Lac	−17.76
1K7U	WGA (wheat germ aggl.)	$(GlcNAc)_2$	−21.34
1EHH	UDA (*Urtica dioica* aggl.)	$(GlcNAc)_3$	−21.34
	Validation Set		
1GIC	ConA	Me-α-D-Glc	−19.25
1QDO	ConA	Me-2-O-(α-D-Man)-α-D-Man	−28.45
1DGL	DGL (*Dioclea grandiflora* lectin)	Me-3,6-di-O-(α-D-Man)-α-D-Man	−34.31
1AX0	ECorL	GalNAc	−17.9
1AX1	ECorL	Lac	−18.8
2BQP	PSL	D-Glc	−14.0
1QF3	PNA	Me-β-D-Gal	−16.96
1EN2	UDA	$(GlcNAc)_4$	−23.43
4GAL	hGal-7 (Human Galectin-7)	Lac	−19.25
5GAL	hGal-7	LacNAc	−18.11

[a] Sources for structures and energies are listed in more detail in the Supporting Information.

charges of the sugar were calculated as ESP charges at the 6-31G* level using GAMESS.[36] Radii and protein atomic partial charges were taken from the PARSE[37] parameter set.

2.8. Calibration and Validation. The data set of crystal structures given in Table 1 was divided into two equally large sets of 10 protein−carbohydrate complexes. The first set was used for calibrating SLICK. Validation was done on the second set. Structures were selected to represent lectin families and ligand sizes equally in each set.

SLICK/score and SLICK/energy were calibrated on the crystal structures of the calibration set. For SLICK/energy, cross-validation was applied for assessing its robustness against changes in the calibration set. Then, for each complex, 200 docking candidates were created with AutoDock. Each candidate was rescored with SLICK/score, giving a ranking of the docking candidates. The binding free energy of the highest ranked true positive of every complex was estimated with SLICK/energy. In this context, structures with a heavy-atom root-mean-square deviation (RMSD) less than 1.5 Å from the crystal structure were considered as true positives. For analysis, the absolute difference between estimated and measured energy was calculated. Rank, RMSD, and binding free energy were additionally computed using the AutoDock energy function.

3. RESULTS

Calibrating SLICK/energy produced a correlation coefficient of 0.94 with a maximum absolute error ΔE_{max} of 2.8 kJ/mol and a mean absolute error ΔE_{avg} of 1.0 kJ/mol. The energy function achieved a mean absolute error of 2.8 kJ/mol in leave-one-out cross-validation. Randomized 5-fold cross-validation yielded a mean absolute error of 3.3 kJ/mol averaged over 1000 runs.

Rescoring and energetic evaluation of the highest-ranked (first) true positive are given in Table 2. For the validation set, the highest-ranked candidate is at the same time the first true positive in four cases. While the candidates of the DGL complex are not ranked well, the R_{ftp} of the remaining structures is below or equal to 5. These numbers indicate that SLICK/score is able to identify binding conformations very well. Figure 5 shows exemplary plots of RMSD versus SLICK/score for two complexes of the validation set. Energies are estimated with a mean absolute error of 7.3 kJ/mol, while the mean RMSD is about 1.0 Å. Comparing these numbers to those of the calibration set expectedly reveals a slightly better performance except for the rank of the first true positive.

Table 2 suggests a correlation between ligand size and energy prediction quality. In the mean, energies of monomeric ligands in the validation set are estimated with an error of 4.7 kJ/mol, while the mean error for dimeric ligands is almost twice as high at 8.1 kJ/mol. Ranks also tend to be much better for monomers. Similar results can be observed in the calibration set.

AutoDock, which was used for docking the complexes, encounters large difficulties in identifying the binding conformation and estimating binding energies. In most cases, R_{ftp} is well-above 10, giving a mean R_{ftp} of almost 50. Additionally, the mean absolute error of the binding free energy is about 22 kJ/mol. AutoDock energies do not reveal a significant correlation between ligand size and prediction error.

Structural analysis revealed two main reasons for bad scoring of oligomers. First, the flexibility of sugars along the glycosidic bond connecting monomers leads to a very large conformational space, which easily overstrains the structure-generation method. In our case, the docking program used for creating candidates was not designed for the lectin sugar case. Consequently, its structure generator, which uses a more general energy function, tends to create clusters of candidates that do not cover the conformational space sufficiently. Second, water molecules play an important role in many lectin−sugar complexes. Comparing, for example, the binding modes of Gal and Lac in the binding site of ECorL reveals tightly bound water molecules in the crystal structure on the protein surface. These have to be

Table 2. Results of Rescoring and Evaluating AutoDock Candidates of Calibration and Validation Set with SLICK[a]

complex			SLICK				AutoDock			
PDB ID	*n*-mer	ΔG_{exp} (kJ/mol)	R_{ftp}	d_{ftp} (Å)	ΔG_{comp} (kJ/mol)	ΔE (kJ/mol)	R_{ftp}	d_{ftp} (Å)	ΔG_{comp} (kJ/mol)	ΔE (kJ/mol)
					Calibration Set					
1J4U	1	−18.24	1	1.11	−20.38	2.14	3	1.41	−37.28	19.04
5CNA	1	−22.18	1	0.46	−22.23	0.05	1	0.89	−39.91	17.73
1AXZ	1	−18.20	1	0.78	−21.57	3.37	1	1.20	−34.90	16.70
1BQP	1	−16.60	1	0.78	−21.10	4.50	40	0.81	−35.35	18.75
1AX2	2	−22.70	25	1.01	−30.44	7.74	93	1.01	−46.07	23.37
2PEL	2	−17.76	34	1.06	−33.90	16.14	119	1.18	−42.26	24.50
1ONA	3	−30.96	1	1.14	−36.87	5.91	3	1.14	−64.06	33.10
1EHH	3	−21.34	30	1.47	−36.48	15.14	91	1.47	−49.04	27.70
mean			11.75	0.98		6.87	43.88	1.14		22.61
					Validation Set					
1GIC	1	−19.25	1	0.59	−19.91	0.66	1	0.90	−39.25	20.00
1AX0	1	−17.90	1	0.89	−24.98	7.08	1	0.88	−41.59	23.69
2BQP	1	−14.00	2	0.45	−21.35	7.35	73	0.47	−35.27	21.27
1QF3	1	−16.96	5	0.88	−20.50	3.54	54	0.98	−32.51	15.55
1QDO	2	−28.45	2	1.49	−25.98	2.47	175	1.49	−42.93	14.48
1AX1	2	−18.80	1	0.60	−21.19	2.39	78	0.49	−39.91	21.11
4GAL	2	−19.25	2	1.27	−32.21	12.96	67	1.29	−37.61	18.36
5GAL	2	−18.41	3	1.13	−33.13	14.72	18	1.13	−41.17	22.76
1DGL	3	−34.41	26	1.04	−41.85	7.44	9	1.04	−63.85	29.44
1EN2	4	−23.43	1	1.25	−37.53	14.10	15	1.43	−60.33	36.90
mean			4.40	0.96		7.27	49.10	1.01		22.36

[a] R_{ftp} denotes the rank of the first true positive candidate (RMSD < 1.5 Å); d_{ftp} is the RMS deviation of the first true positive from the crystal structure; ΔG_{comp} denotes the computed binding free energy, and ΔE is the deviation of the predicted energy from the experimental binding free energy $\Delta G_{exp.}$ Numbers are given for SLICK and AutoDock. The complexes 1QDC and 1K7U of the calibration set are not shown. For orientation, the number of monomers in the ligand is given in column *n*-mer.

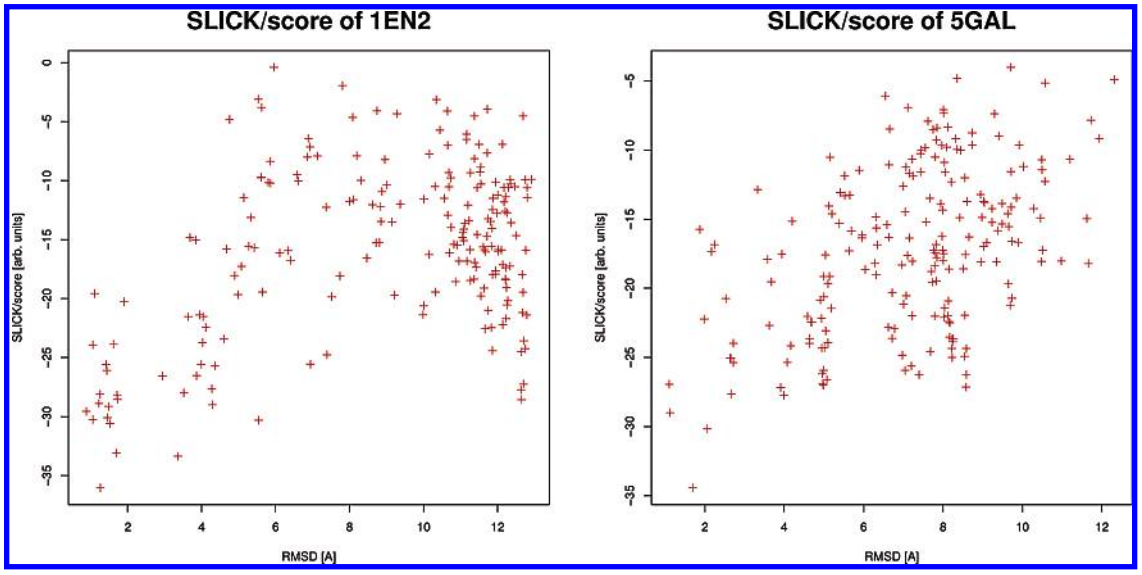

Figure 5. Rescoring docked structures with SLICK/score.

replaced by the larger ligand. Additionally, dimeric sugars often bind tightly with one ring, while the other extends into the solvent where it is stabilized by water-mediated hydrogen bonds. Neither AutoDock nor SLICK take such interactions into account at the present time.

Analysis of the individual contributions of SLICK showed clearly that the hydrogen-bonding and CH···π terms are most important for the identification of binding conformations. An example illustrating the coordination of the sugar in the binding site by hydrogen bonds and CH···π interactions is shown in Figure 6. In fact, several complexes could be scored correctly using only the hydrogen-bonding and CH···π terms. There are only a few examples where, especially, CH···π scores unspecifically when aromatic side chains are present in the binding site.

4. DISCUSSION

In this article, we have introduced SLICK, a package providing SLICK/score, a function for scoring structures in protein−carbohydrate docking, and SLICK/energy, an empirical energy function for predicting binding free energies of protein−carbohydrate complexes. We have shown that, with this combination of scoring and energy functions, the binding modes and affinities of various plant lectin complexes can be calculated with high accuracy. We have also shown that the energy function proves robust on the data

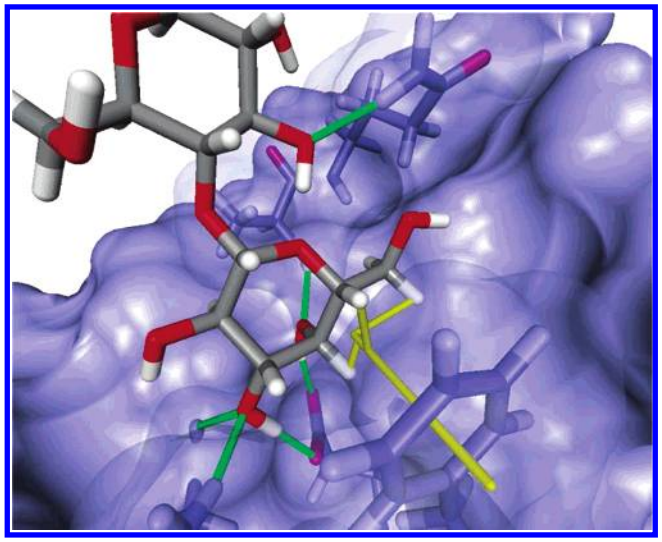

Figure 6. Hydrogen bonds (green sticks) and CH···π interactions (yellow sticks) in the crystal structure of the binding site of ECorL binding Lac. Hydrogen bonds are built with ASP 89, GLY 107, ASN 133, ALA 218, and GLN 229. Three CH···π interactions coordinate the Gal ring over PHE 131.

sets we used. Moreover, the applicability of SLICK to sugar−lectin complexes beyond the domain of plant lectins has been shown on the example of human galectin-7.

The CH···π interaction proved important for the prediction of carbohydrate binding modes. Together with hydrogen bonding, it strongly influences the binding conformation of carbohydrate ligands, which is consistent with earlier experimental studies.[8] Consequently, docking programs used for protein−carbohydrate interactions should incorporate CH···π models.

Generating structures with AutoDock using its own energy function clearly yields a set of ligand candidates that are not optimal with respect to the scoring function under investigation. Nevertheless, our results indicate that SLICK/score could prove generally useful on independent structure generators. The incorporation of SLICK/score into a docking program will be the next step in proving its effectiveness and increasing prediction quality.

Because of the small calibration data set, the danger of overfitting the energy function is undeniable. We are aware of that and plan to increase the size of the data set soon. New data is currently being generated.

The simple cutoff model for van der Waals interactions is surely not the best approach. There are several softened van der Waals potentials in use. Changing the repulsive exponent to 9 instead of 12 without cutting the potential is reportedly useful for identifying known ligands in rigid docking.[25]

The analysis revealed a great importance of water molecules for the binding mode of oligomeric sugars. Modeling water in the binding site of receptors still poses an unsolved problem in docking applications. More research is necessary in that field, especially for predicting correct binding modes of lectin−carbohydrate complexes. This problem has to be addressed through the structure-generation method, not the scoring or energy functions.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their valuable comments. This work was supported in part by Deutsche Forschungsgemeinschaft grants BIZ 1/1-3, BIZ 4/1-1, and LE 952/2-3.

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* (7), 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* (2), 95−104.

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

(4) Yamazaki, M.; Kojima, S.; Bovin, V.; André, N. S.; Gabius, S.; Gabius, H.-J. Endogenous lectins as targets for drug delivery. *Adv. Drug Delivery Rev.* **2000**, *43*, 225−244.

(5) Hong, F.; Yan, J.; Baran, J. T.; Allendo, D. J.; Hansen, R. D.; Ostroff, G. R.; Xing, P. X.; Cheung, N.-K. V.; Gordon, D. Ross 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.

(6) von Bubnoff, A. Sugar coating improves anticancer treatment. news@nature.com, DOI: 10.1038/news050418-6.

(7) 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.

(8) Muraki, M.; Ishimura, M.; Harata, K. Interactions of wheat-germ agglutinin with GlcNAcβ1,6Gal sequence. *Biochim. Biophys. Acta* **2002**, *1569*, 10−20.

(9) Morris, G. M.; Goodsell, D. S.; Halliday, R. S. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19* (14), 1639−1662.

(10) 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.

(11) Jones, G.; Willet, 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.

(12) Macias, A. T.; Mackerell, A. D., Jr. CH/π interactions involving aromatic amino acids: Refinement of the CHARMM tryptophan force field. *J. Comput. Chem.* **2005**, *26*, 1452−1463.

(13) Laederach, A.; Reilly, P. J. Specific empirical free energy function for automated docking of carbohydrates to proteins. *J. Comput. Chem.* **2003**, *24*, 1748−1757.

(14) Reiling, S.; Schlenkrich, M.; Brickmann, J. Force field parameters for carbohydrates. *J. Comput. Chem.* **1997**, *17*, 450−468.

(15) Glennon, T. M.; Merz, K. M., Jr. A carbohydrate force field for AMBER and its application to the study of saccharide to surface adsorption. *THEOCHEM* **1997**, *395*−*396*, 157−171.

(16) Pusztai, A. *Plant Lectins*; Cambridge University Press: Cambridge, U. K., 1991; pp 78−95.

(17) Pusztai, A.; Bardocz, S. Biological effects of plant lectins on the gastrointestinal tract: Metabolic consequences and applications. *Trends Glycosci. Glycotechnol.* **1996**, *8*, 149−165.

(18) 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.

(19) Taroni, C.; Jones, S.; Thornton, J. M. Analysis and prediction of carbohydrate binding sites. *Protein Eng.* **2000**, *13*, 89−98.

(20) Bohne, A.; Lang, E.; von der Lieth, C.-W. W3-SWEET: Carbohydrate modeling by internet. *J. Mol. Model.* **1998**, *4*, 33−43.

(21) R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2005; ISBN 3-900051-07-0.

(22) Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R. C-h...pi-interactions in proteins. *J. Mol. Biol.* **2001**, *307*, 357−377.

(23) Böhm, H.-J. The development of a simple empirical scoring function to estimate the binding constant for a protein−ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243−256.

(24) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast

empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.
(25) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* **2004**, *47*, 5076−5084.
(26) 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.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, *7* (19), 5179−5197.
(27) 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.
(28) Kohlbacher, O.; Lenhof, H.-P. BALL − Rapid software prototyping in computational molecular biology. *Bioinformatics* **2000**, *16* (9), 815−824.
(29) Gilson, M. K.; Sharp, K. A.; Honig, B. H. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J. Comput. Chem.* **1988**, *9*, 327−335.
(30) Uhlig, H. H. The solubilities of gases and surface tension. *J. Phys. Chem.* **1937**, *41* (9), 1215−1225.
(31) Reiss, H.; Tully-Smith, D. M. Further development of scaled particle theory for rigid spheres: Application of the statistical thermodynamics of curved surfaces. *J. Chem. Phys.* **1971**, *55*, 1674−1689.
(32) Pierotti, R. A. A scaled particle theory of aqueous and nonaqueous solutions. *Chem. Rev.* **1976**, *76*, 717−726.
(33) 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.
(34) The Chemical Computing Group. *The Molecular Operating Environment*; The Chemical Computing Group: Montreal, Canada. http://www.chemcomp.com/ (accessed Jun 2005).
(35) 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* (11), 3832−3846.
(36) 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.
(37) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **1994**, *98* (7), 1978−1988.

CI050422Y