In Silico Mutagenesis and Docking Studies of *Pseudomonas aeruginosa* PA-IIL Lectin − Predicting Binding Modes and Energies

Jan Adam,[†,||] Zdeněk Kříž,[†,||] Martin Prokop,[†] Michaela Wimmerová,[†,‡] and Jaroslav Koča*[,†,§]

National Centre for Biomolecular Research, Department of Biochemistry, and Department of Chemistry, Faculty of Science, Kotlářská 2, Masaryk University, 611 37 Brno, Czech Republic

Received June 24, 2008

This article is focused on the application of two types of docking software, AutoDock and DOCK. It is aimed at studying the interactions of a calcium-dependent bacterial lectin PA-IIL (from *Pseudomonas aeruginosa*) and its *in silico* mutants with saccharide ligands. The effect of different partial charges assigned to the calcium ions was tested and evaluated in terms of the best agreement with the crystal structure. The results of DOCK were further optimized by molecular dynamics and rescored using AMBER. For both software, the agreement of the docked structures and the provided binding energies were evaluated in terms of prediction accuracy. This was carried out by comparing the computed results to the crystal structures and experimentally determined binding energies, respectively. The performance of both docking software applied on a studied problem was evaluated as well. The molecular docking methods proved efficient in identifying the correct binding modes in terms of geometry and partially also in predicting the preference changes caused by mutation. Obtaining a reasonable *in silico* method for the prediction of lectin-saccharide interactions may be possible in the future.

1. INTRODUCTION

Ligand−receptor recognition processes are very important in all living systems. As observed in the past decade, the lectin-saccharide interaction plays a very important role within this respect.

Lectins are proteins that are ubiquitous in the realm of the living organisms. They display a characteristic ability to bind carbohydrates specifically, but they have no catalytic activity overall.[1] They are a large family of proteins whose unifying feature is the ability to decode the sugar code information.[2] As a consequence, lectins play a major role in processes such as the cell-cell signaling, bioadhesion, and host recognition in pathogen infection.

In relation to saccharide binding, various architectures of the saccharide binding site have been reported to be used by lectins. Binding sites have been found and described in numerous possible protein folds.[3] One of the most common ways of binding a saccharide ligand is the combination and balance of hydrogen bonds with stacking of aliphatic rings of the monosaccharides and aromatic rings of sidechains of some amino acids.[4]

Another major type of lectin-saccharide binding uses the presence of one or more bridging cations (usually calcium) in the binding site. These so-called C-type lectins were characterized for the first time in the animal kingdom and reported to play important roles in the process of cell adhesion, among others.[5−7] Despite the fact that there are variations in the structures of lectins, the amino acids of the binding site are conserved. The binding itself is usually realized by a combination of two hydroxyl oxygen atoms coordinated to the calcium ion, combined with hydrogen bonds between the saccharide and the surrounding amino acids.[8]

Recent reports have been made about new calcium dependent lectins isolated from several opportunistic bacteria, first and foremost the *Pseudomonas aeruginosa*.[9] This bacterium is capable of causing high mortality and morbidity rates in patients with the cystic fibrosis disease. These lectins may play a crucial role in the processes of host recognition and the formation of biofilm during the colonization of the host tissue.[10−12] Recently it was proposed that there may be also a role in pilus biogenesis and proteolysis.[13] *Pseudomonas aeruginosa* produces two lectins that are related to its virulence, PA-IL and PA-IIL. The PA-IL lectin is a tetrameric, galactose-specific lectin with one calcium ion in the binding site and a binding mode analogical to the known C-type lectins.[14] The PA-IIL lectin is also a tetrameric lectin (see Figure 1), but it expresses an unusually high (micromolar) affinity to monosaccharides, namely L-fucose, L-galactose, and D-arabinose, respectively, and their methylated forms.[15] It displays a previously unobserved sugar binding site architecture with two close calcium ions (3.7 Å) embedded in an acidic pocket. The ions' coordination spheres are completed by three oxygen atoms from the hydroxyl groups of the saccharide. In this case, both atoms participate in the saccharide coordination, a fact previously unobserved in the lectin realm. Aside from the binding via calcium coordination, the amino acids interact with other atoms of the saccharide by hydrogen bonds as well as other hydrophilic and hydrophobic interactions. This is very closely related to the composition of the binding site, and a slight change usually leads to remarkable changes in binding

* Corresponding author phone: +420549494947; fax: +420549492556; e-mail: jkoca@chemi.muni.cz.
† National Centre for Biomolecular Research.
|| These authors contributed equally to this work.
‡ Department of Biochemistry.
§ Department of Chemistry.

10.1021/ci8002107 CCC: $40.75
Published on Web 10/21/2008

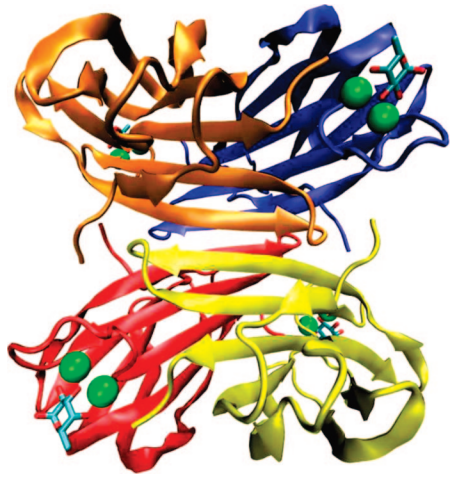

Figure 1. Tetrameric structure of the PA-IIL lectin with bound α-L-fucose ligand (PDB code 1GZT).

behavior, as it was proved by comparing binding properties of PA-IIL and its homologues. Whereas the PA-IL type lectin was found only in the *Pseudomonas aeruginosa* bacterium, several PA-IIL-like lectins were found by genome searching in other opportunistic bacteria like *Ralstonia solanacearum* and *Chromobacterium violaceum* and characterized to different extents.[16−19] These lectins display similar structure as that of the PA-IIL. They express similarly high affinities to monosaccharides (using the same binding mode), yet they differ from each other in sugar preferences - RS-IIL from *Ralstonia solanacearum* is a strongly mannose-preferring lectin, whereas the CV-IIL lectin from *Chromobacterium violaceum* prefers both L-fucose and D-mannose to comparable extent. The recent studies revealed the importance of composition of the binding site, namely the 22−23−24 "specificity-binding loop", and described the relationship between structure and thermodynamics of binding.[20] The aim of this study is to describe the effects of the specificity-binding loop composition by means of molecular modeling. Molecular modeling has proven to be a useful tool for the rationalization of specificity and related issues, especially with regards to the fact that the computational time needed for such applications is substantially shorter than performing the whole process from conception of a mutation to a solved crystal structure or functional characterization of the mutant *in vitro*. Molecular modeling can also assist in solving related applied research problems, like the design of glycomimetic drugs.

This work focuses on mutating residues present in the fucose-specific PA-IIL to the residues of mannose-specific RS-IIL. Figure 2 shows an alignment of the sequences of both lectins. The residues mutated in this study were serine 22 to alanine (mutant S22A), serine 23 to alanine (S23A), and glycine 24 to asparagine (G24N).

The modeling of lectin-saccharide interactions involves several important factors, which need to be considered in the calculations. Though the saccharides have a rigid cyclic structure, the flexibility of the functional groups is very high and has to be accounted for in the docking procedure, even for monosaccharides. The conformational behavior of the glycosidic bond in oligosaccharides is an important factor also. Successful application of docking to protein-saccharide interactions using the AutoDock program have been previously reported.[21−23] Nevertheless, the presence of two calcium ions in the binding site directly participating in the binding requires special attention.

This study focuses on the comparison of two docking programs, AutoDock[24] and DOCK.[25] Both programs are capable of docking flexible ligands. AutoDock uses a hybrid search that combines the general search by Lamarckian genetic algorithm with a local search derived from Solis and Wets algorithm.[26] Each AutoDock run provides a docked energy, state variables, geometry of the docked complex, and the estimated free energy of binding. The DOCK program uses a different approach based on creating a negative image of the potential binding site and then superimposing the ligand upon it. The docking is based upon the anchor-and-grow system where the anchor (i.e., the largest rigid substructure of the ligand) is placed into the binding site.[27] The best orientation is identified and optimized using the scoring function based on the AMBER force field *ff99*[28] and energy minimization according to the Nelder and Mead algorithm.[29] The orientations are then clustered and prioritized according to their score. The rest of the flexible ligand is then "grown" on best anchor orientations. This program only takes into account the ligand-protein interactions based on intermolecular electrostatic and van der Waals forces.

In-house developed software, TRITON, was used to prepare the computational projects for AutoDock and MODELLER software as well as to visualize and analyze the results. By providing an attractive and user-friendly graphical user interface, TRITON is of great help for both beginner and advanced users of the aforementioned software.

Both programs were tested, and the accuracy and efficiency was evaluated in terms of geometry and energy estimation. This comparison was made in order to evaluate their usefulness for developing an *in silico* method capable of predicting changes in lectin affinity and specificity caused by minor mutations in the structure of the binding site, that are proven to play an important role in reality. Such a method would provide valuable information for experimental biochemists and molecular biologists and allow them to identify the right directions leading to desired changes in protein behavior, essentially enabling an effective protein engineering approach. Implementation of such a specialized method into TRITON would further enhance the performance and attractiveness of the software for potential users.

2. METHODS

2.1. Sources. The structure of the PA-IIL complex with the fucose ligand was solved by Mitchell et al.[30] The initial structure for this work was obtained from the PDB database under the code 1GZT and used as a starting structure. The parameters suitable for use in molecular dynamics and energy minimization are available for different force fields.[31,32] The structures of ligands were either obtained from the CERMAV saccharide database (http://www.cermav.cnrs.fr/glyco3d/)[33] or modeled by using the Glycam biomolecule builder (http://www.glycam.com/CCRC/biombuilder/biomb_index.jsp).[34]

2.2. Macromolecule Preparation. Only a half of the tetrameric molecule, i.e. dimeric structure of two mutually complementary monomers containing two complete binding sites, was used for docking. This has been done in order to

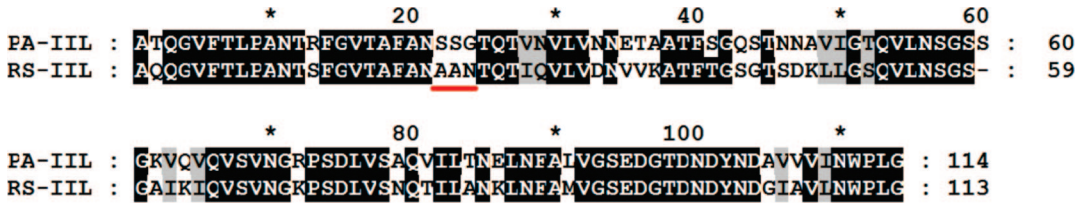

Figure 2. The alignment of the peptide sequences of PA-IIL and RS-IIL. The conserved amino acids are shaded black, similar amino acids are shaded gray, and the differing are left unshaded. The specificity-binding loop, i.e. amino acids in positions 22−23−24 is underlined in red.

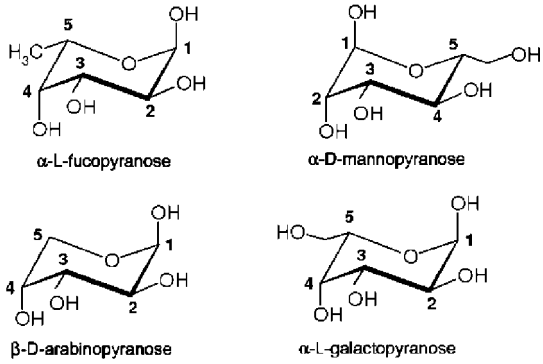

Figure 3. Examples of monosaccharides with the configuration required for binding the binding site of PA-IIL family lectins.

save computational time without any detrimental consequences for the studied binding site.

All water molecules, cocrystallized molecules, and ligands were removed from the structures. The hydrogen atoms were added by WHAT IF software.[35] The structures of single point mutants were modeled also *in silico* by the homology modeling-using program TRITON interfaced with MODELLER v. 6.2.[36] The active site of the PA-IIL lectin α-L-fucose complex was examined. The calcium ions charge was changed stepwise from +1.5e to +2.0e in a 0.1e step in order to find the optimal charge assignment.

2.3. Ligand Set Preparation. The ligands used for docking computations were chosen with regard to the available experimental structure and energy data. The PA-IIL family mutants are known to bind primarily monosaccharides with $^1C_4$ L-galacto and $^4C_1$ D-manno pyranose configurations, because of the specific configuration and orientation of the hydroxyl groups, necessary for proper binding into the site by completing the coordination spheres of both calcium ions. Examples of such "PA-IIL-family" compatible saccharides are shown in Figure 3. In computational experiments, it is possible to include only one of the (α- or β-) anomeric forms of the saccharide although in real solutions both of the anomeric forms are present in a dynamic equilibrium. The methylation of the O1 ensures the conservation of the desired anomeric form in the experiment. Several of these monosaccharides were prepared *in silico* in the O1-methylated form.

The geometries of all the structures were optimized using the Hartree−Fock method with a 6−31G(d) basis set as implemented in the Gaussian 03 program.[37] The ESP charges were calculated, and the RESP procedure of the Antechamber program from the AMBER program suite was used to generate input files with charges for docking programs.[28,38] The result was a Sybyl *mol2* file that was used as a valid input for the AutoDock ligand preparation procedure.

2.4. Docking. *AutoDock v.3.0.* Docking was performed by the academic software AutoDock 3[24] under the in-house developed TRITON software.[39] For the protein partial atoms charges, the Kollman united atom charges were used[40] as suggested in the AutoDock user manual. The Lamarckian genetic algorithm was used for the search of the energetically favored binding modes. The calcium ions' charge was determined (several settings were tested, see Results). A box set with its center topping a hypothetical equilateral triangle with the two calcium atoms, the dimensions being 40 × 40 × 40 grid points with 0.375 Å spacing, with 100 runs of Lamarckian algorithm and 500 000 energy evaluations per run were used. The population size was 50 individuals. During the algorithm, all exocyclic groups were set as rotatable. The evaluation of scoring function energies and the free energies using the cluster analysis with a 2 Å rmsd threshold was performed by AutoDock. Input data were prepared, and the results were visualized by TRITON software.

DOCK v. 6.1. The hydrogen atoms were added by the Chimera program[41] with respect to optimization of the H-bond network inside the protein. The AMBER charges were added to protein atoms according to the AMBER *ff99* force field using residue and atom names. The docking was performed by the DOCK 6.1 software.[42] The spheres generated to the distance of 12.0 Å around calcium ions in the active site were selected. A grid density of 0.25 Å was calculated for nonbonded interactions. The flexible docking procedure was used for searching the possible monosaccharide position. All noncyclic torsions of monosaccharides were changed in step by 10.0 degrees during the flexible docking procedure. Fifty thousand different orientations of monosaccharide molecules were generated, and the geometries of all positions were optimized. The docked structures were clustered with a 2.0 Å rmsd threshold. The top five of the best docked structures, from the standard scoring function point of view, were used for further optimization via AMBER molecular dynamics under the following conditions. The molecules were converted to the AMBER format, and the van der Waals radii were rescaled according to the used GB 5 model.[43] For simulation, the *ff99* force field[28] was used for the protein, and the general AMBER force field[44] was used for the saccharides. The system was minimized in 800 cycles, and then 25000 molecular dynamic steps were run at 300 K, followed by 2000 minimization cycles. These runs were performed for the macromolecule, ligand, and for the complex. The resulting energy was calculated according to the thermodynamics laws.

Table 1. Root Mean Square Deviations (RMSD) of TRITON/MODELLER-Produced Mutants and Corresponding Crystal Structures[a]

structure	rmsd [Å]
S22A	0.13
S23A	0.10
G24N	0.15

[a] The RMSD is calculated for all atoms belonging to 10 Å surroundings of the calcium ions.

3. RESULTS

3.1. *In Silico* Mutagenesis of PA-IIL. The dimeric structure of PA-IIL obtained under the conditions described above was used as a template structure for the homology modeling of three one-point mutants S22A, S23A, and G24N using TRITON, interfaced with MODELLER software. The X-ray structures of S22A, S23A, and G24N mutants and their complexes with monosaccharides from the Brookhaven Protein Database (PDB codes: 1GZT, 2JDM, 2JDN, 2JDP, 2JDU, and 2JDY) were solved by Adam et al.[20] Table 1 shows a comparison of the crystal structures and the MODELLER produced mutants in terms of rmsd.

It can be seen that the deviation between modeled mutants and crystal structures is minimal. The greater difference in the latter part is explained by the presence of several amino acids that can exist in several conformations. None of these amino acids is present neither in immediate or extended vicinity of the binding site. As such, the difference between *in silico* mutants and the X-ray structures was considered to be acceptable for treating the X-ray and MODELLER *in silico* constructed mutants as equal.

3.2. The Calcium Ions' Charge Variation. One of the first questions posed was the assignment of charge to the calcium ions in the binding site. The *ab initio* calculation on the system made by Mitchell et al.[15] resulted in charges of about +1.5 on the calcium ions. In our docking, the value of the calcium ion charge was changed from 1.5 to 2.0 in 0.1 increments. Docking of the α-L-fucose ligand into the macromolecule was performed by both types of software under conditions described in this article previously. The effect of the charge assigned to calcium atoms was measured in terms of both energy and geometry, and results from AutoDock are summarized in Table 2, whereas Table 3 provides results generated by DOCK software. The effect of the charge assignment to the ligand position can be seen in Figure 4. The different positions of the ligand dependent on the assigned calcium charge for the DOCK results are shown. The graphical representation of AutoDock results is not included, as the visual comparison shows no significant correlation between the charges and the ligand position.

The effect of charge evaluation was based on three criteria—binding energy, number of the clusters resulting from each cluster analysis in each program run, and interatomic distances. The number of clusters represents the relative accuracy of the docking procedure, while the interatomic distances represent the agreement with crystal structures.

3.3. Docking of the Monosaccharide Molecules. The monosaccharides were docked into the prepared macromolecules using both docking software. The docked saccharide conformations with the lowest binding energy were chosen and compared to the experimental data of the isothermal titration microcalorimetry and X-ray structures, where available. The comparison of the geometry of docked structures for saccharide-lectin complexes can be seen in Figures 5−7.

4. DISCUSSION

4.1. The Effect of Calcium Ions Charge Assignment. Calcium ion charge assignment was investigated by a series of docking experiments on the scale from 1.5 to 2.0 in 0.1 increments using both programs.

For the AutoDock software, increasing the charge of the calcium ion in the complex resulted in more favorable values of binding energy. The count of clusters reached its minimum at values of 1.8 and 2.0. Overall, the distances of the calcium ions from the coordinated oxygen atoms displayed lower values than observed in the crystal structure but without any relevant observable trend.

Results from the DOCK software are represented by two data sets: one for the standard DOCK scoring function and the other for the AMBER-rescored results. The standard scoring function divides the total free energy into its electrostatic and van der Waals contributions. The increase of the ions' charge is reflected by an increase in the electrostatic part of the total binding energy, as anticipated. The cluster count reached a minimum 5 for both 1.8 and 2.0 values. The AMBER optimization provided energy values with a minimum of 1.8. The distances obtained when using the standard scoring function display a notable difference from the experimental X-ray values. After the AMBER optimization, the distances were lowered as to closer agreement with the crystallographic data. Consequently the whole ligand was repositioned closer to both calcium ions (see Figure 4). Once again, the closest agreement reached was for values 1.8 and 2.0. The value of 1.8 was chosen as a compromise between the formal charge of the calcium ion and the lower charge that the calcium ion, surrounded by several negatively charged oxygen atoms, adopts in reality (about 1.5). Overall, the result of this experiment suggested two important conclusions: the standard scoring function of DOCK provides docked conformations of the ligand that are less tightly bound to protein and application of the AMBER scoring function provides conformations that are very close to the crystal structures.

We believe that the used charges are transferable to other systems with similar structure.

Besides the variation of the charges, it is possible to investigate the effect of assigning other parameters, as described by Nurisso et al.[45]

4.2. Docking of Saccharide Ligands. The docking provides a series of free binding energies for the possible lectin-saccharide complexes. The results are summarized in Table 4. The achieved results can be evaluated from several points of view, especially with regards to the experiment. Concerning AutoDock results, it is apparent that the absolute values of the binding energy differ somewhat from the experimental values; however, considering the level of simplification used here, this is not surprising. Albeit the absolute values are different in the first iteration of the prediction method, it is acceptable to achieve at least semiquantitative results with agreement in trends. Generally, the O1-methylated glycosides display better values of binding energy than their non-

Table 2. Effect of the Calcium Ion Charge in PA-IIL/α-L-Fuc Complex as Calculated by AutoDock[a]

charge (e)	E	number of clusters	*d* (O4−Ca1)	*d* (O3−Ca1)	*d* (O3−Ca2)	*d* (O2−Ca2)
1.5	−8.71	5	2.09	2.30	2.22	**2.27**
1.6	−9.03	5	2.19	2.27	2.22	2.12
1.7	−9.08	4	2.29	2.05	**2.46**	2.07
1.8	−9.33	**3**	2.06	2.31	2.20	2.15
1.9	−9.66	4	**2.45**	2.03	2.54	1.96
2.0	−9.74	**3**	2.25	**2.33**	2.22	2.09
X-ray distances			2.48	2.47	2.47	2.54

[a] Energy in [kcal.mol^{-1}], distances in ångström. The clusters are represented by the minimum energy conformation.

Table 3. Effect of the Calcium Ion Charge in PA-IIL/α-L-Fuc Complex as Calculated by DOCK[a]

charge (e)		E [kcal.mol^{-1}]	number of clusters	*d* (O4−Ca1)	*d* (O3−Ca1)	*d* (O3−Ca2)	*d* (O2−Ca2)
1.5	D[b]	−34.58 (−6.40/-28.18)	19	4.56	2.59	3.99	2.65
	A[b]	−27.42		2.98	2.74	2.69	2.84
1.6	D	−35.64(−7.61/-28.03)	7	4.85	2.60	4.09	2.52
	A	−30.40		2.79	2.74	2.66	2.81
1.7	D	−35.35 (−7.77/-27.88)	8	4.57	2.54	4.15	2.66
	A	−28.99		2.79	2.75	2.66	2.83
1.8	D	−35.92(−8.24/-27.68)	**5**	4.53	2.55	4.03	2.68
	A	−31.92		2.80	2.71	2.66	2.85
1.9	D	−37.57(−9.51/-28.06)	6	4.77	2.62	3.88	2.57
	A	−26.43		2.83	2.78	2.65	2.91
2.0	D	−37.33 (−10.57/-26.76)	**5**	4.79	2.57	4.05	2.54
	A	−30.32		2.80	2.72	2.65	2.82
X-ray distances				2.48	2.47	2.47	2.54

[a] Energy in [kcal.mol-1], distances in ångström. The clusters are represented by the minimal energy conformation. [b] D stands for energies obtained from DOCK (electrostatic/van der Waals), and A stands for energies obtained from DOCK combined with AMBER rescoring procedure.

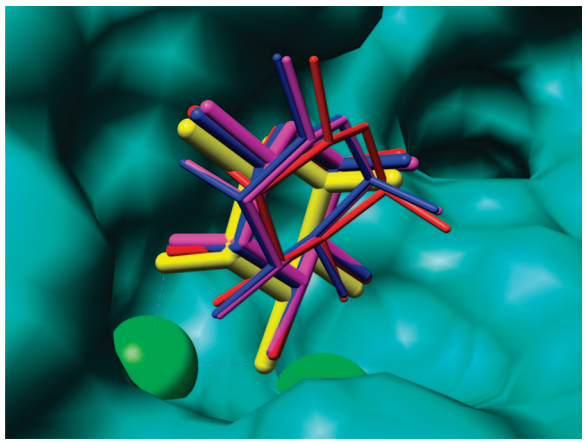

Figure 4. Comparison of DOCK-docked and DOCK-docked/AMBER-rescored structures of α-L-fucose into the PA-IIL lectin binding site and the crystal structure of the complex, for different assignments of charge to calcium ions. Crystal structure of the ligand is shown in yellow, charges 1.5 = red, 1.8 = magenta, 2.0 = navy blue. Thin sticks = DOCK, thick sticks = DOCK/AMBER rescoring.

methylated counterparts. This fact was observed in the experiment and described in a previously published article.[20] The reason may be an additional hydrophobic interaction caused by the presence of the methyl group. Furthermore, the ranks of the monosaccharides from docking calculations and experiments were compared to each other. Whereas the full comparison cannot be characterized as matching, the main features of the mutants that were measured experimentally are also observable in the computed complexes. Overall, the most important feature is the best compatibility (and therefore binding energy) of methylfucoside for PA-IIL lectin and the sugar preference switch from methylfucoside to methylmannoside caused by the serine 22 to the alanine mutation in the S22A mutant. With regards to mutants S23A and G24N, the order of preference was again similar to the experimental values, with methylfucoside and methylgalactoside preceding methylmannoside. However, with the exception of α-L-fucose in complex with PA-IIL and the α-L-Gal/S23A complex, the nonmethylated monosaccharides expressed worse free binding energies than the methylmannoside in all investigated cases. Visualization of the preference switch and the correspondence of computed and measured energies can be seen in Figure 8.

The DOCK software provided analogical sets of complex structures and related energies. In the case of DOCK, the initial docking procedure was not always successful in finding the proper spatial position of the monosaccharide ligand. Whereas the orientation of the ligand molecule was roughly correct, the whole ligand was significantly displaced in comparison to the crystal structures, as seen in Figures 5−7. It could be seen that the initial energies produced by DOCK (using the standard scoring function) very roughly reflect the AutoDock results trendwise, but the values of the energies are approximately four times higher than in the case of AutoDock. Alike the AutoDock results, the methylated saccharides expressed more favorable interaction energies than the nonmethylated counterparts. The energy difference may be caused by the improper parametrization of the saccharide molecules. Compared to AutoDock results, the DOCK does not seem to be quite the ideal software for the chosen application, whereas AutoDock more or less preserves the order of preference known from the ITC experiments.

An explanation as to why the use of the standard scoring function of DOCK provides a relatively poor result may lie

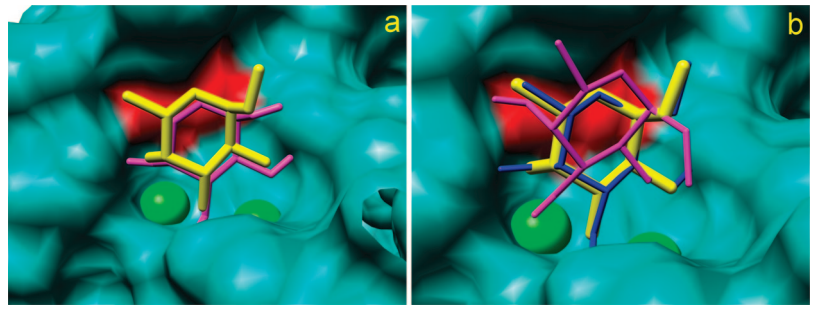

Figure 5. The docked structures of methyl-α-L-fucoside into the binding site of the S22A mutant of PA-IIL, crystal structure shown in yellow, docked structure in magenta. For DOCK results, the AMBER-rescored structure is shown in navy blue, additionally: a) S22A mutant with methyl-α-L-fucoside, AutoDock and b) S22A mutant with methyl-α-L-fucoside, DOCK,DOCK/AMBER rescoring. The mutated residue is highlighted by red color. The crystal structure does not contain hydrogen atoms. For docked sugar molecules, to keep the clarity of presentation, only hydrogens on hydroxyl groups are shown.

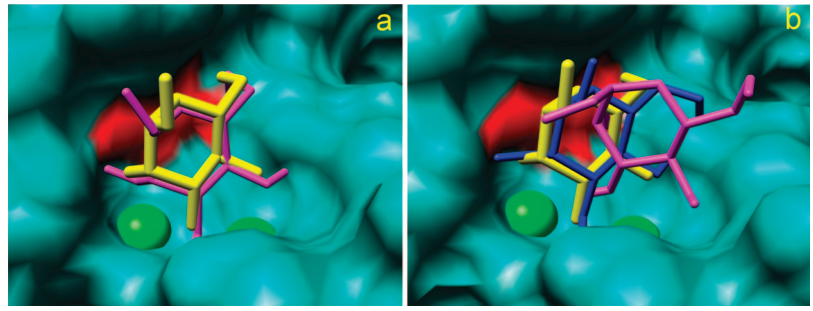

Figure 6. The docked structures of methyl-α-D-mannoside into the binding site of the S22A mutant of PA-IIL, crystal structure shown in yellow, docked structure in magenta. For DOCK results, the AMBER-rescored structure is shown in navy blue, additionally: a) S22A mutant with methyl-α-D-mannoside, AutoDock and b) S22A mutant with methyl-α-D-mannoside, DOCK,DOCK/AMBER rescoring. The mutated residue is highlighted by red color. The crystal structure does not contain hydrogen atoms. For docked sugar molecules, to keep the clarity of presentation, only hydrogens on hydroxyl groups are shown.

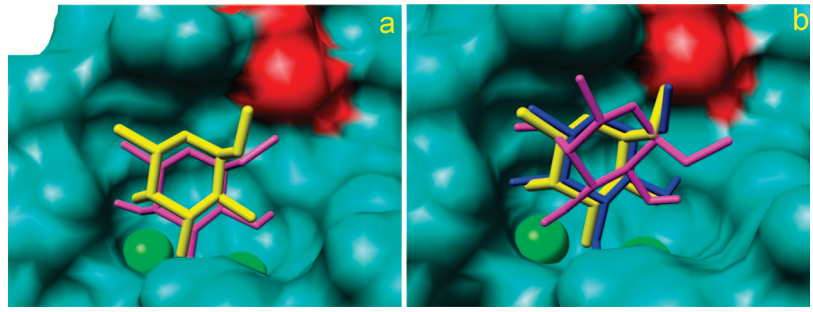

Figure 7. The docked structures of methyl-α-L-fucoside into the binding site of the G24N mutant of PA-IIL, crystal structure shown in yellow, docked structure in magenta. For DOCK results, the AMBER-rescored structure is shown in navy blue, additionally: a) G24N mutant with methyl-α-L-fucoside, AutoDock and b) G24N mutant with methyl-α-L-fucoside, DOCK,DOCK/AMBER rescoring. The mutated residue is highlighted by red color. The crystal structure does not contain hydrogen atoms. For docked sugar molecules, to keep the clarity of presentation, only hydrogens on hydroxyl groups are shown.

in the fact that the DOCK scoring function does not incorporate the solvent effect. Moreover, the anchor-and-grow algorithm in this case may be inappropriate. The algorithm places the rigid part of the molecule first and evaluates placement without considering the flexible parts. It then builds the flexible section according to the best orientations found. Proper positioning of the flexible parts of the molecule (e.g., the hydroxyl groups) is crucial for finding the appropriate binding mode, and the molecule must be taken as a whole when docked. In the case of AMBER optimization, the results still significantly differ from the experimental values, although the geometry of the binding mode is improved in relation to the crystal structure—the oxygen atoms coordinating the calcium ions are brought to more proper positions. Another issue that must be addressed is the inclusion of solvent interactions. It is known that for the lectin-saccharide systems, water molecules often participate in the interaction and mediate binding. The used software is not able to include the water molecules into the docking as active participants, and this problem is solved in different ways by different software. The AutoDock includes the solvent interaction through some parameters of the scoring function. The DOCK standard scoring neglects the solvent interaction altogether. However, the DOCK-AMBER rescoring process uses implicit solvent approach. In our calculations, the contribution of the solvent was performed as allowed by the software. As for the energy estimation, it is known that calculations using a continuous solvent field neglect the specific effect of the solvent molecules, that may result in the fact that the free energy calculations are only crude estimations of the real state (if such interactions are present), as reported by Spackova et al.[46] The use of

Table 4. Comparison of the Energy Values [kcal.mol^{-1}] Obtained from Docking Using AutoDock and DOCK Software with the Experimental Energy Values Determined by Isothermal Titration Calorimetry[a]

PA-IIL	AutoDock	DOCK (standard)	DOCK (AMBER resc.)	ITC
Me-α-L-Fuc	−10.7	−40.01	−32.22	−8.71
Me-α-L-Gal	−9.93	−40.02	−28.92	−8.09
α-L-Fuc	−9.33	−35.92	−27.56	−6.94
Me-α-D-Man	−9.23	−43.50	−26.88	−5.97
α-L-Gal	−8.85	−36.57	−20.58	−
α-D-Man	−7.37	−41.23	−17.52	−
β-D-Ara	−9.12	−32.37	−9.25	−
S22A	**AutoDock**	**DOCK (standard)**	**DOCK (AMBER resc.)**	**ITC**
Me-α-D-Man	−10.47	−32.60	−22.30	−7.58
Me-α-L-Fuc	−9.26	−30.60	−23.47	−7.39
α-L-Fuc	−8.96	−28.81	−19.85	−6.84
Me-α-L-Gal	−9.49	−31.70	−20.34	−6.60
α-L-Gal	−8.59	−29.59	−15.57	−
α-D-Man	−7.67	−31.57	−15.96	−
β-D-Ara	−8.59	−32.60	−6.00	−
S23A	**AutoDock**	**DOCK (standard)**	**DOCK (AMBER resc.)**	**ITC**
Me-α-L-Fuc	−10.49	−37.90	−20.82	−9.04
Me-α-L-Gal	−10.79	−38.83	−19.86	−8.11
α-L-Fuc	−8.98	−35.43	−18.81	−7.32
Me-α-D-Man	−9.15	−39.95	−16.80	−5.84
α-L-Gal	−9.20	−38.33	−14.02	−
α-D-Man	−6.80	−38.31	−24.26	−
β-D-Ara	−8.56	−34.23	−11.62	−
G24N	**AutoDock**	**DOCK (standard)**	**DOCK (AMBER resc.)**	**ITC**
Me-α-L-Fuc	−9.65	−39.84	−28.19	−9.19
Me-α-L-Gal	−9.83	−38.42	−22.26	−8.06
α-L-Fuc	−9.29	−37.10	−22.48	−7.18
Me-α-D-Man	−9.15	−45.03	−25.16	−5.96
α-L-Gal	−9.20	−36.97	−21.83	−
α-D-Man	−6.80	−37.01	−22.51	−
β-D-Ara	−8.63	−35.10	−10.80	−

[a] The AutoDock free energies were obtained from the in-built evaluation process, and the DOCK free energies come from the inbuilt evaluation process and from the re-evaluation by AMBER. Data were ordered by the ITC energies. However, a total agreement of the docking energies and experimental ITC data is not expected. It is important that docking energies give a good qualitative prediction. It is clearly seen that this is really the case for our data as it is shown in Figure 8. Also comparison of the energy values rank gives an excellent agreement for the DOCK docking energies rescored by AMBER and correlated with the ITC data. The only little exception is S22A mutant. This is caused by a different binding mode.

molecular dynamics with an explicit solvent is an obvious next step in future development and optimization of the method.

5. CONCLUSIONS

This work focused on the application of molecular modeling for the investigation of lectin-saccharide interactions and on the evaluation of the performance of two types of docking software applied to the problem. The AutoDock and DOCK software were used for docking a set of monosaccharides into the binding sites of PA-IIL lectin and its mutants. By means of varying the charge of the calcium atoms for the docking experiments and evaluating the results

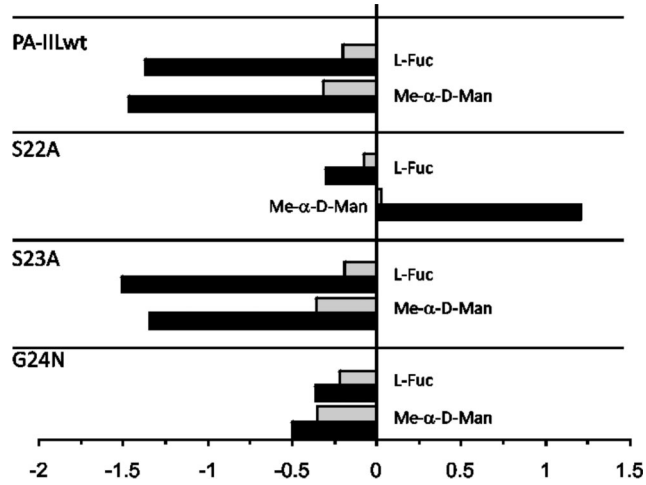

Figure 8. Visualization of the differences in AutoDock-computed energies (black) and ITC-measured energies (gray) of protein complexes of wild-type PA-IIL and its mutants with L-fucose and methyl-α-D-mannoside, compared to corresponding protein and methyl-α-L-fucoside complexes. (E(prot-saccharide) − E(prot-Me-α-L-Fuc)).

on the base of accuracy in relation to experimental data, an optimal charge value of 1.8 was chosen to be assigned to the AutoDock software and was able to identify the correct binding mode as the lowest energy conformation. Compared to experimental results, the AutoDock results reproduced the experimental order of preference to a large extent, keeping the desired mutual preference order between fucose and mannose in the mutants. The DOCK program provided docked structures that slightly differed from the crystal structures. This suboptimal performance could be explained by the nature of the docking process used, which puts the DOCK software at a slight disadvantage. After AMBER optimization and rescoring, structural agreement was greatly improved. The DOCK software produced results with higher absolute values of interaction energies, and rescoring by AMBER produced more reasonable energy values. However, the order of preference was conserved to a lesser extent compared to AutoDock. In order to get values closer to the real interaction energies, an additional step of molecular dynamics free energy estimation would be needed. Nevertheless, even in the current state, the automated docking method proved quite capable of identifying preference trends, and, therefore, it could be used as a preplanning step for the *in vitro* experiments, to help to identify the best potential candidates for mutagenesis, thus reducing the experimental time needed. Future work will be directed to further study of *in silico* mutagenesis (i.e., random mutations of the specificity-binding loop) and followed by *in vitro* mutagenesis verification of the predictions. Also, the use of limited receptor-flexible docking has been planned as soon as the in-house developed software TRITON is accommodated to implement the use of such a feature as introduced with AutoDock 4.

ACKNOWLEDGMENT

The authors thank the Ministry of Education, Youth, and Physical Training of the Czech Republic (Grant Numbers: MSM0021622413 and LC06030) and Grant Agency of the Czech Republic (Grant Number: GA303/06/0570).

REFERENCES AND NOTES

(1) Lis, H.; Sharon, N. Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **1998**, *98* (2), 637–674.

(2) Gabius, H. J.; Andre, S.; Kaltner, H.; Siebert, H. C. The sugar code: functional lectinomics. *Biochim. Biophys. Acta, Gen. Subj.* **2002**, *1572* (2−3), 165–177.

(3) Loris, R. Principles of structures of animal and plant lectins. *Biochim. Biophys. Acta, Gen. Subj.* **2002**, *1572* (2−3), 198–208.

(4) Rini, J. M. Lectin Structure. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 551–577.

(5) Cambi, A.; Koopman, M.; Figdor, C. G. How C-type lectins detect pathogens. *Cell. Microbiol.* **2005**, *7* (4), 481–488.

(6) Drickamer, K.; Fadden, A. J. Genomic analysis of C-type lectins. *Glycogenomics: Impact Genomics Informatics Glycobiol.* **2002**, (69), 59–72.

(7) Drickamer, K.; Taylor, M. E. Biology of Animal Lectins. *Annu. Rev. Cell Biol.* **1993**, *9*, 237–264.

(8) Drickamer, K. Ca2+-dependent sugar recognition by animal lectins. *Biochem. Soc. Trans.* **1996**, *24* (1), 146–150.

(9) Gilboa-Garber, N. *Pseudomonas aeruginosa* lectins. *Methods Enzymol.* **1982**, *83*, 378–385.

(10) Rhim, A. D.; Stoykova, L. I.; Trindade, A. J.; Glick, M. C.; Scanlin, T. F. Altered terminal glycosylation and the pathophysiology of CF lung disease. *J. Cyst. Fibrosis* **2004**, *3*, 95–96.

(11) Roussel, P.; Lamblin, G. The glycosylation of airway mucins in cystic fibrosis and its relationship with lung infection by Pseudomonas aeruginosa. *Adv. Exp. Med. Biol.* **2003**, *535*, 17–32.

(12) Tielker, D.; Hacker, S.; Loris, R.; Strathmann, M.; Wingender, J.; Wilhelm, S.; Rosenau, F.; Jaeger, K.-E. *Pseudomonas aeruginosa*lectin LecB is located in the outer membrane and is involved in biofilm formation. *Microbiology* **2005**, *151*, 1313–1323.

(13) Sonawane, A.; Jyot, J.; Ramphal, R. Pseudomonas aeruginosa LecB is involved in pilus biogenesis and protease IV activity but not in adhesion to respiratory mucins. *Infect. Immun.* **2006**, *74* (12), 7035–7039.

(14) Cioci, G.; Mitchell, E. P.; Gautier, C.; Wimmerova, M.; Sudakevitz, D.; Perez, S.; Gilboa-Garber, N.; Imberty, A. Structural basis of calcium and galactose recognition by the lectin PA-IL of Pseudomonas aeruginosa. *FEBS Lett.* **2003**, *555* (2), 297–301.

(15) Mitchell, E. P.; Sabin, S.; Šnajdrová, L.; Budová, M.; Perret, S.; Gautier, C.; Hofr, C.; Gilboa-Garber, N.; Koca, J.; Wimmerová, M.; Imberty, A. High affinity fucose binding of Pseudomonas aeruginosa lectin PA-IIL: 1.0 L resolution crystal structure of the complex combined with thermodynamics and computational chemistry approaches. *Proteins: Struct., Funct., Bioinf.* **2005**, *58*, 735–748.

(16) Perret, S.; Sabin, C.; Dumon, C.; Pokorná, M.; Gautier, C.; Galanina, O.; Ilia, S.; Bovin, N.; Nicaise, M.; Desmadril, M.; Gilboa-Garber, N.; Wimmerova, M.; Mitchell, E. P.; Imberty, A. Structural basis for the interaction between human milk oligosaccharides and the bacterial lectin PA-IIL of Pseudomonas aeruginosa. *Biochem. J.* **2005**, *389*, 325–332.

(17) Pokorna, M.; Cioci, G.; Perret, S.; Rebuffet, E.; Kostlanova, N.; Adam, J.; Gilboa-Garber, N.; Mitchell, E. P.; Imberty, A.; Wimmerova, M. Unusual entropy-driven affinity of Chromobacterium violaceum lectin CV-IIL toward fucose and mannose. *Biochemistry* **2006**, *45* (24), 7501–10.

(18) Sabin, C.; Mitchell, E. P.; Pokorna, M.; Gautier, C.; Utille, J. P.; Wimmerova, M.; Imberty, A. Binding of different monosaccharides by lectin PA-IIL from Pseudomonas aeruginosa: Thermodynamics data correlated with X-ray structures. *FEBS Lett.* **2006**, *580* (3), 982–987.

(19) Sudakevitz, D.; Kostlanova, N.; Blatman-Jan, G.; Mitchell, E. P.; Lerrer, B.; Wimmerova, M.; Katcof, f. D. J.; Imberty, A.; Gilboa-Garber, N. A new Ralstonia solanacearum high affinity mannose-binding lectin RS-IIL structurally resembling the Pseudomonas aeruginosa fucose-specific lectin PA-IIL. *Mol. Microbiol.* **2004**, *52*, 691–700.

(20) Adam, J.; Pokorna, M.; Sabin, C.; Mitchell, E. P.; Imberty, A.; Wimmerova, M. Engineering of PA-IIL lectin from Pseudomonas aeruginosa - Unravelling the role of the specificity loop for sugar preference. *BMC Struct. Biol.* **2007**, 7.

(21) Coutinho, P. M.; Dowd, M. K.; Reilly, P. J. Automated docking of glucosyl disaccharides in the glucoamylase active site. *Proteins: Struct., Funct., Genet.* **1997**, *28* (2), 162–173.

(22) Laederach, A.; Coutinho, P. M.; Reilly, P. J. Flexible docking of carbohydrates to proteins: Development and application of specific empirical free energy models. *Abstr. Pap. Am. Chem. Soc.* **2002**, *223*, U497−U497.

(23) Laederach, A.; Reilly, P. J. Modeling protein recognition of carbohydrates. *Proteins: Struct., Funct., Bioinf.* **2005**, *60* (4), 591–597.

(24) 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* (14), 1639–1662.

(25) Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. Development and validation of a modular, extensible docking program: DOCK 5. *J. Comput.-Aided Mol. Des.* **2006**, *20* (10−11), 601–619.

(26) Solis, F. J.; Wets, R. J. B. Minimization by Random Search Techniques. *Math. Oper. Res.* **1981**, *6* (1), 19–30.

(27) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, *18* (9), 1175–1189.

(28) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91* (1−3), 1–41.

(29) Nelder, J. A.; Mead, R. A Simplex-Method for Function Minimization. *Comput. J.* **1965**, *7* (4), 308–313.

(30) Mitchell, E.; Houles, C.; Sudakevitz, D.; Wimmerova, M.; Gautier, C.; Pérez, S.; Wu, A. M.; Gilboa-Garber, N.; Imberty, A. Structural basis for oligosaccharide-mediated adhesion of *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients. *Nat. Struct. Biol.* **2002**, *9*, 918–921.

(31) Imberty, A.; Perez, S. Structure, conformation, and dynamics of bioactive oligosaccharides: Theoretical approaches and experimental validations. *Chem. Rev.* **2000**, *100* (12), 4567–4588.

(32) Perez, S.; Imberty, A.; Engelsen, S. B.; Gruza, J.; Mazeau, K.; Jimenez-Barbero, J.; Poveda, A.; Espinosa, J. F.; van Eyck, B. P.; Johnson, G.; French, A. D.; Louise, M.; Kouwijzer, C. E.; Grootenuis, P. D. J.; Bernardi, A.; Raimondi, L.; Senderowitz, H.; Durier, V.; Vergoten, G.; Rasmussen, K. A comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates. *Carbohydr. Res.* **1998**, *314* (3−4), 141–155.

(33) Rivet, A.; Mazeau, A.; Bettler, E.; Imberty, A.; Perez, S. Glyco 3D - 3D Monosaccharide database. http://www.cermav.cnrs.fr/cgi-bin/monos/monos.cgi (accessed 2008).

(34) Woods, R. J. *Biomolecule Builder, GLYCAM Web*; Complex Carbohydrate Research Center: University of Georgia, 2008.

(35) Vriend, G. What If - a Molecular Modeling and Drug Design Program. *J. Mol. Graphics* **1990**, *8* (1), 52–56.

(36) Sali, A.; Blundell, T. L. Comparative Protein Modeling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234* (3), 779–815.

(37) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E. R. M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y. ; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02*; Gaussian, Inc.: Wallingford, CT, 2004.

(38) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–1688.

(39) Prokop, M.; Adam, J.; Kriz, Z.; Wimmerova, M.; Koca, J. TRITON: a graphical tool for ligand-binding protein engineering. *Bioinformatics* **2008**, *24* (17), 1955–1956.

(40) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106* (3), 765–784.

(41) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.

(42) Lang, P. T.; Moustakas, D.; Brozell, S.; Carrascal, N.; Mukherjee, S.; Pegg, S.; Raha, K. D. S.; Rizzo, R.; Case, D.; Shoichet, B.; Kuntz, I. D. *DOCK, 6.1*; University of California: San Francisco, CA, 2007.

(43) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified

generalized born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55* (2), 383–394.
(44) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.
(45) Nurisso, A.; Kozmon, S.; Imberty, A. Comparison of docking methods for carbohydrate binding in calcium-dependent lectins and prediction of the carbohydrate binding mode to sea cucumber lectin CEL-III. *Mol. Simul.* **2008**, *34* (4), 469–479.
(46) Spackova, N.; Cheatham, T. E.; Ryjacek, F.; Lankas, F.; van Meervelt, L.; Hobza, P.; Sponer, J. Molecular dynamics simulations and thermodynamics analysis of DNA-drug complexes. Minor groove binding between 4′,6-diamidino-2-phenylindole and DNA duplexes in solution. *J. Am. Chem. Soc.* **2003**, *125* (7), 1759–1769.

CI8002107