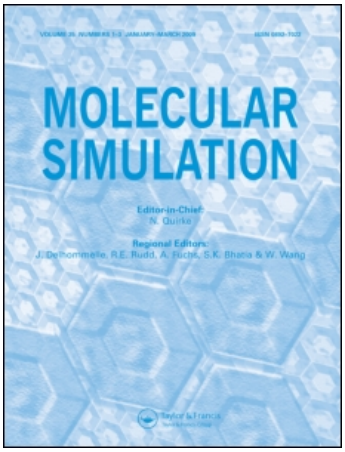

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Comparison of docking methods for carbohydrate binding in calcium-dependent lectins and prediction of the carbohydrate binding mode to sea cucumber lectin CEL-III

A. Nurisso[a]; S. Kozmon[b]; A. Imberty[a]
[a] Centre de Recherches sur les Macromolécules Végétales (CERMAV-CNRS), Grenoble, France [b] Institute of Chemistry, Slovak Academy of Sciences, Bratislava, Slovak Republic

To cite this Article Nurisso, A. , Kozmon, S. and Imberty, A.(2008) 'Comparison of docking methods for carbohydrate binding in calcium-dependent lectins and prediction of the carbohydrate binding mode to sea cucumber lectin CEL-III', Molecular Simulation, 34: 4, 469 — 479
To link to this Article: DOI: 10.1080/08927020701697709
URL: http://dx.doi.org/10.1080/08927020701697709

Molecular Simulation
Vol. 34, No. 4, April 2008, 469–479

Taylor & Francis
Taylor & Francis Group

Comparison of docking methods for carbohydrate binding in calcium-dependent lectins and prediction of the carbohydrate binding mode to sea cucumber lectin CEL-III

A. Nurisso[a1], S. Kozmon[b] and A. Imberty[a1]*

[a]*Centre de Recherches sur les Macromolécules Végétales (CERMAV-CNRS), Grenoble, France;* [b]*Institute of Chemistry, Slovak Academy of Sciences, Bratislava, Slovak Republic*

(*Received 13 August 2007; final version received 23 September 2007*)

Lectins display a variety of strategies for specific recognition of carbohydrates. In several lectin families from different origin, one or two calcium ions are involved in the carbohydrate binding site with direct coordination of the sugar hydroxyl groups. Our work implied a molecular docking study involving a set of bacterial and animal calcium-dependant lectins in order to compare the ability of three docking programs to reproduce key carbohydrate-metal interactions. Flexible docking was performed using AutoDock, DOCK and Grid-based Ligand Docking with Energetics (GLIDE) softwares. All docking packages were almost able to predict the carbohydrate binding orientations but not in every instance the result was obvious to evaluate. DOCK showed good results according to crystallographic information but not in all tested cases the lowest energy conformation identified the experimental data. GLIDE presented the same difficulty in result analysis but the lowest energy pose was always a satisfactory solution, able to mimic the real carbohydrate orientation. AutoDock showed a reasonable accuracy in sugar orientation prediction based on docking cluster number ranking and most accurate distances between calcium and sugar hydroxyl groups. The latest program and GLIDE were used to predict the Gal and GalNAc binding mode in sea cucumber CEL-III, a new calcium-dependent lectin, that displays haemolytic and cytotoxic properties.

Keywords: C-type lectins; carbohydrates; autoDock; DOCK; GLIDE

1. Introduction

Lectins are ubiquitous proteins that recognise specifically carbohydrates but are devoided of catalytic activities and are not immunoglobulins [1]. This large family of proteins encompasses many biological functions that involve the deciphering of the sugar code [2], for example in intracellular glycoprotein trafficking, cell–cell signalisation or host recognition in pathogen infection. Lectins display a variety of carbohydrate binding site architectures, built in a large number of possible protein folds [3]. Due to the aliphatic character of monosaccharides, the carbohydrate binding involves a balance of hydrogen bonding and stacking of aromatic amino acids with CH of carbohydrates [4].

One particular carbohydrate binding mode involves the presence of a bridging cation ion in the binding site. Calcium-dependent lectins were first characterised in animal kingdom. The so-called C-type lectins cover a wide range of extracellular and membrane-bound proteins that contain one or several conserved carbohydrate recognition domains (CRDs) [5]. In mammals, they have multiple roles, including cell adhesion (selectins), glycoprotein clearance (asialoglycoprotein receptor), innate immunity (collectins), while in invertebrates they often play a role in non-self recognition processes involved in innate immunity and establishment of symbiosis [6,7]. Structural variations are observed in loops and disulfide bridges, but the amino acids of the sugar binding site that are involved in calcium binding are conserved. Upon carbohydrate binding, two of the sugar hydroxyl groups are involved in the coordination of calcium and also establish hydrogen bonds with the neighbouring amino acids [8]. Depending on the surrounding amino acids, C-type lectin CRD are specific for mannose and GlcNAc (and fucose) or for galactose.

More recently calcium-dependent lectins have been purified from opportunistic bacteria such as *Pseudomonas aeruginosa* [9]. These soluble lectins may play a role in host tissue recognition and/or in biofilm formation [10]. PA-IL is a tetrameric lectin specific for galactose and the involvement of O3 and O4 of galactose in the coordination of calcium is reminiscent to what is observed in animal C-type lectins [11]. The fucose/mannose specific PA-IIL exhibits a new motif of sugar binding with involvement of two close calcium ions (3.7 Å) that requires the participation of three hydroxyl groups in binding [12]; it is characterised by unusually strong (micromole) affinity for monosaccharides [13]. Whereas PA-IL has only been identified in *P. aeruginosa*, PA-IIL-like lectins were characterised from other opportunistic bacteria such as *Ralstonia solanacearum* and *Chromobacterium violaceum*

*Corresponding author. Email: imberty@cermav.cnrs.fr

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927020701697709
http://www.informaworld.com

and the preference for mannose or fucose is correlated to the amino acid of a neighbouring loop [14,15].

Very recently, the calcium-dependent family of lectin has been completed by a rather surprising member. The sea cucumber lectin CEL-III adopts a double β-trefoil fold [16], also referred as R-type lectin, that has been previously characterised in many organisms and present three galactose binding sites with no calcium. However, in the Gal/GalNAc specific CEL-III lectin, five of the six sites contain a calcium ion in proper place for sugar binding and the lectin can therefore be considered as the first member of calcium-dependent β-trefoil lectin, although crystal structure of the complex has not been yet reported.

Molecular modelling of protein–carbohydrate interaction has proven to be a useful tool to rationalise specificity or to design glycomimetics that could be of therapeutical interest. Energy parameters that are suitable for energy minimisation and/or molecular dynamics of protein carbohydrate complexes are available for different force fields [17,18]. For predicting the carbohydrate orientation in binding sites, flexible docking methods have to be used in order to account for the possible orientations of pendent groups (i.e. hydrogen bond network directed by hydroxyl/hydroxymethyl group orientation) and also the conformational behaviour of the glycosidic linkage for oligosaccharides. Nevertheless docking methods have been used with a high rate of success for example using AutoDock program [19,20]. However, the presence of calcium ions directly involved in the binding site represents a special case. In some computer programs, the parameters for calcium have to be added to classical parameterisation. We propose here to compare three of the classically used flexible docking algorithms AutoDock [21], Dock [22] and Grid-based Ligand Docking with Energetics (GLIDE) [23] for docking monosaccharides and disaccharides in the different types of calcium dependent lectins that are displayed in Figure 1. The comparison with crystal structures will point out the strengths and weaknesses of each program. In addition, we used AutoDock 3 and GLIDE for predicting the docking of galactose in CEL-III, a new type of calcium dependent lectin for which no binding mode is yet structurally determined.

2. Flexible docking algorithms

In all of the three programs used in the study, the receptor is considered as a rigid body while the ligand is free to rotate, translate and change conformation during the docking application. Docking programs consists of two key parts: a search algorithm and a scoring function.

AutoDock is based on a hybrid search method that applied a Lamarckian genetic algorithm [21]. This exploration of binding site is based on a global searching that uses a genetic algorithm followed by an adaptive local search method derived from an optimisation of Solis and Wets algorithm [24] which has the advantage to not requiring gradient computation while it performs torsional space search. In the implementation of the genetic algorithm, the chromosome is composed of a string of real valued genes which describe the ligand translation, using the three Cartesian coordinates, the ligand orientation, involving four variables for the quaternion and the ligand conformation, defined by one real value for each torsion. The genetic algorithm begins by generating a random population of individuals which uniformly explore the grid space, followed by a specified number of generation cycles, each one consisting of a mapping and fitness evaluation, a selection, a crossover, a mutation and an elitist selection. After this step, each generation cycle is followed by the local search. At the end of every docking, AutoDock reports the docked energy, the state variables, the coordinates of docked conformation and the estimated free energy of binding [21].

DOCK [22] employs matching methods for the automated docking. In order to guide the search for ligand orientations, a negative image of the active site volume is created: spheres are only located in the receptor surface area that can interact with solvent molecules. This method allows for the limitation of possible ligand orientations to the most relevant region on the surface of the receptor. The incremental construction algorithm, called anchor-and-grow, separates the ligand flexibility into two steps: first, the largest rigid substructure of the ligand (anchor) is identified, rigidly oriented in the binding site [25]. The possible orientations are evaluated and optimised using the scoring function based on AMBER molecular mechanics force-field [26] and the energy minimiser based on the original Nelder and Mead algorithm [27]. The orientations are then collected according to their score, spatially clustered and prioritised. The remaining flexible portion of the ligand is built onto the best anchor orientations within the context of the receptor (grow). Only the interactions between ligand and protein are considered, leaving only intermolecular van der Waals and electrostatic components in the function; in addition, the receptor potential energy contribution can be pre-calculated and stored on a grid.

GLIDE program [23,28] uses a hierarchical series of filters to search for possible locations of the ligand in the active-site region of the receptor. The shape and properties of the receptor are represented on a grid by several sets of fields that provide progressively more accurate scoring of the ligand poses. Conformational flexibility is handled in GLIDE by an extensive conformational search, augmented by a heuristic screen. The scoring is carried out using Schrödinger's discretised version of the ChemScore empirical scoring function. Much as for ChemScore itself,

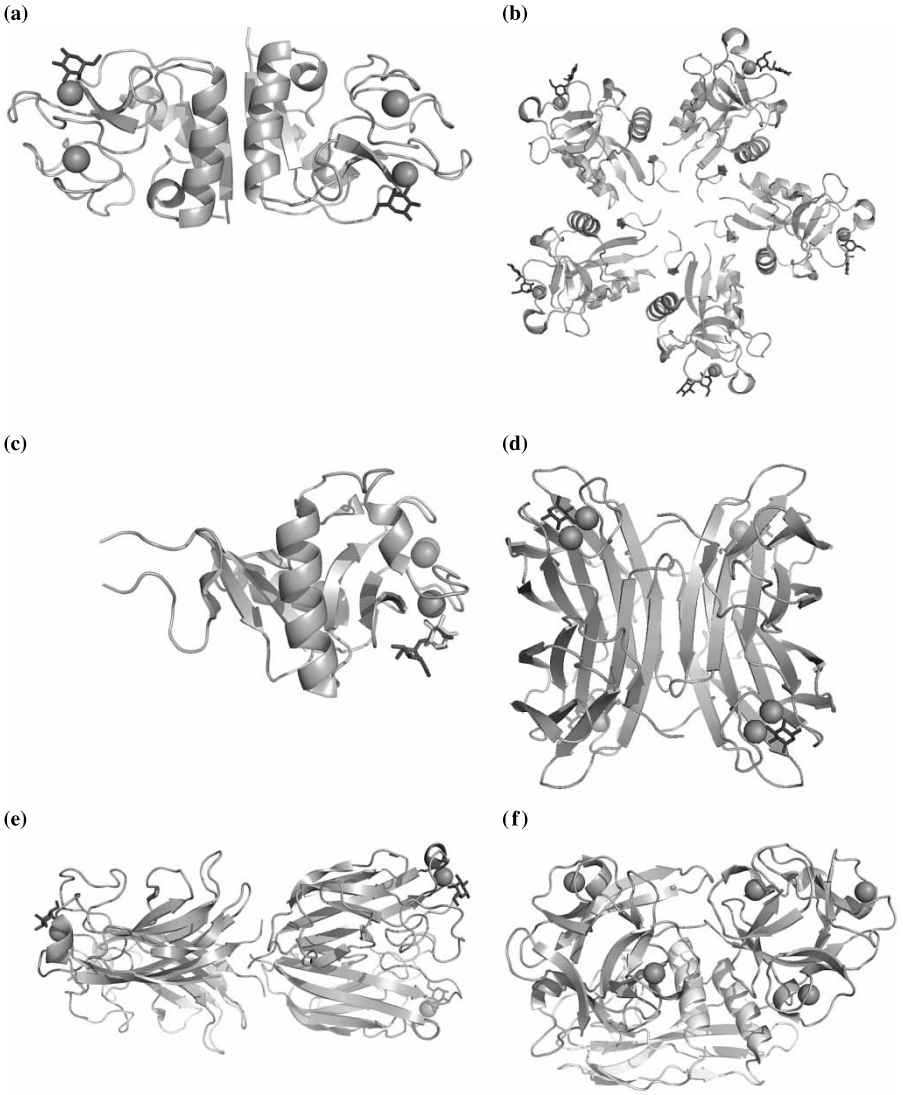

Figure 1. Graphical representation of the different calcium-dependent lectins that have been used in the present study. Peptide chains are represented by ribbon, sugar ligands by sticks and calcium atoms by spheres. (a): Tunicate lectin (C-type) complexed with galactoses (PDB code 1TLG) [38], (b): rattlesnake lectin (C-type) complexed with lactose (1JZN) [39], (c): CRD of human DC-SIGN (C-type lectin) complexed with mannose and mannobiose (1IT6) [40], (d): PA-IIL from *P. aeruginosa* complexed with fucose (1GZT) [12]. (e): PA-IL *P. aeruginosa* complexed with galactose (1OKO) [11]. (f). Sea cucumber CEL-III (1VCL) [16].

this algorithm recognises favorable hydrophobic, hydrogen-bonding, metal–ligand interactions and penalises steric clashes.

3. Materials and methods

3.1 Choice of lectin structures

All bacterial and animal calcium-dependent lectins chosen for this flexible docking study were selected from the lectin data base (http://www.cermav.cnrs.fr/lectines/). Coordinates were taken from the Protein Data Bank [29]. Lectins used in the present work are listed in Table 1 and the different architectures and quaternary arrangements are displayed in Figure 1.

3.2 AutoDock files and parameters

Receptor input files were generated using SYBYL 7.3 [30]. Monomers (or dimers when necessary) of each lectin were isolated; ligands, solvent and heteroatoms were removed except for the calcium ions located in the active sites. Hydrogen atoms were added and partial charges assigned using all-atoms charges of the AMBER force field [31]. Hydrogen atom positions were optimised

Table 1. List of calcium-dependent lectins used in docking calculations.

PDB	Organism	Protein	Ligand	Reference
C-type animal lectin				
1TLG	*Polyandrocarpa misakiensis*	Tunicate lectin	Galactose	[38]
1JZN	*Crotalus atrox*	Rattlesnake venom lectin	Lactose (βGal14βGlc)	[39]
			Dual binding:Mannobiose	
2IT6	*Homo sapiens*	DC-SIGN (CRD)	(αMan1–2αMan) Mannose	[40]
One calcium β-sandwich bacterial lectin				
1OKO	*Pseudomonas aeruginosa*	PA-IL	Galactose	[11]
Two calcium β-sandwich bacterial lectin				
1GZT	*Pseudomonas aeruginosa*	PA-IIL	Fucose	[12]
1UQX	*Ralstonia solanacearum*	RS-IIL	α-Me-mannoside	[15]
β-trefoil lectin fold				
1VCL	*Cucumaria echinata*	CEL-III	Galactose [a]	[16]
			N-acetylgalactosamine	

[a] no crystal structure of complexes available.

through energy minimisation using TRIPOS force field [32]. The programs *cnvmol2topdbq* and *addsol* included in AutoDock v.3.0.5 [21] were used to obtain the receptor atom coordinates, the partial charges, the atomic solvation parameters and fragmental atom volumes. Polar hydrogens were differentiated from non-polar hydrogens using 12-10 hydrogen bonding Lennard-Jones parameters and 12-6 hydrogen bonding Lennard-Jones parameters respectively. Additional atom type was defined for calcium ions using Lennard-Jones parameters previously proposed by Åqvist [33].

Ligands were extracted from CERMAV_3D monosaccharide and disaccharide databases (http://www.cermav.cnrs.fr/glyco3d/) and partial charges were taken according from PIM parameters for the TRIPOS force field [34]. All possible torsions in ligand molecules, including hydroxyl group rotations, were defined using the *deftors* module of AutoDock. The grid maps for van der Waals and electrostatic energies were prepared using *AutoGrid*: grid spacing was set to 0.375 Å, with 60 grid points placed on the centre of the ligand from the corresponding crystal structure complexes. Electrostatic interactions were evaluated using a distance-dependent dielectric constant to model solvent effects. Each single docking experiment consisted of 20 runs, employing a Lamarckian genetic algorithm and a rms deviation tolerance for cluster analysis of 1 Å. Applied parameters for the genetic algorithm are the default ones except for the number of energy evaluations that was set to 1×10^6. Clustering histograms created by AutoDock were analysed through the *get-docked* function to evaluate docking results.

3.3 DOCK files and parameters

The Chimera program [35] was used for preparation of ligand and receptor input files as suggested by Dock 6.1 users manual. Lectin structures were prepared using *Dock Prep* tool of Chimera with removal of ligand and solvent atoms. After addition of hydrogen atoms, partial charges were computed using the Antechamber package [36]. For calcium ions, formal charge was first assigned manually to obtain later the calculation of AM1-BCC atomic partial charge. Ligands were isolated from the original PDB lectins file and saved in MOL2 format after adding hydrogens and computing charges with Antechamber program. The active site for DOCK application was prepared using *dms* and *sphgen* programs. The maximum sphere radius was set to 4 Å while the minimum was set to 1.4 Å. In each case, spheres were selected using a root-mean-square deviation (RMSD) cut-off distance of 10.0 Å from all atom of the ligand crystal structure employing the accessory program *sphere_selector*. The interactive program *showbox* was used to visualise and define the location and size of the receptor box that defines the space for the docking conformational search taking into account the sphere set and enclosing an extra-margin of 5 Å in all the directions.

Grid maps were created using the accessory program *grid*. The *grid.in* file specified the parameters: the distance between grid points along each axis was 0.3 Å, the maximum distance between two atoms for their contribution to the energy score to be computed was set to 9999, the van der Waals energy potential parameters were taken from AMBER99 based on Lennard-Jones values using 6 and 12 as exponents of attractive and repulsive terms respectively. The van der Waals allowed overlap was set to 0.75 was and the dielectric factor coefficient to 4. An all atom model approach was chosen. All ligands were allowed to be flexible during the docking process driven by DOCK version 6.1 [10]. The input file *anchor_and_grow.in* was generated to set the default parameters for the anchor-and-grow algorithm. The number of scored conformers written was set to 20. Docking evaluations were performed with *ViewDock* utility of Chimera [35] executing the structure file *flex_scored.mol2* which contains a summary of best poses generated during the simulation.

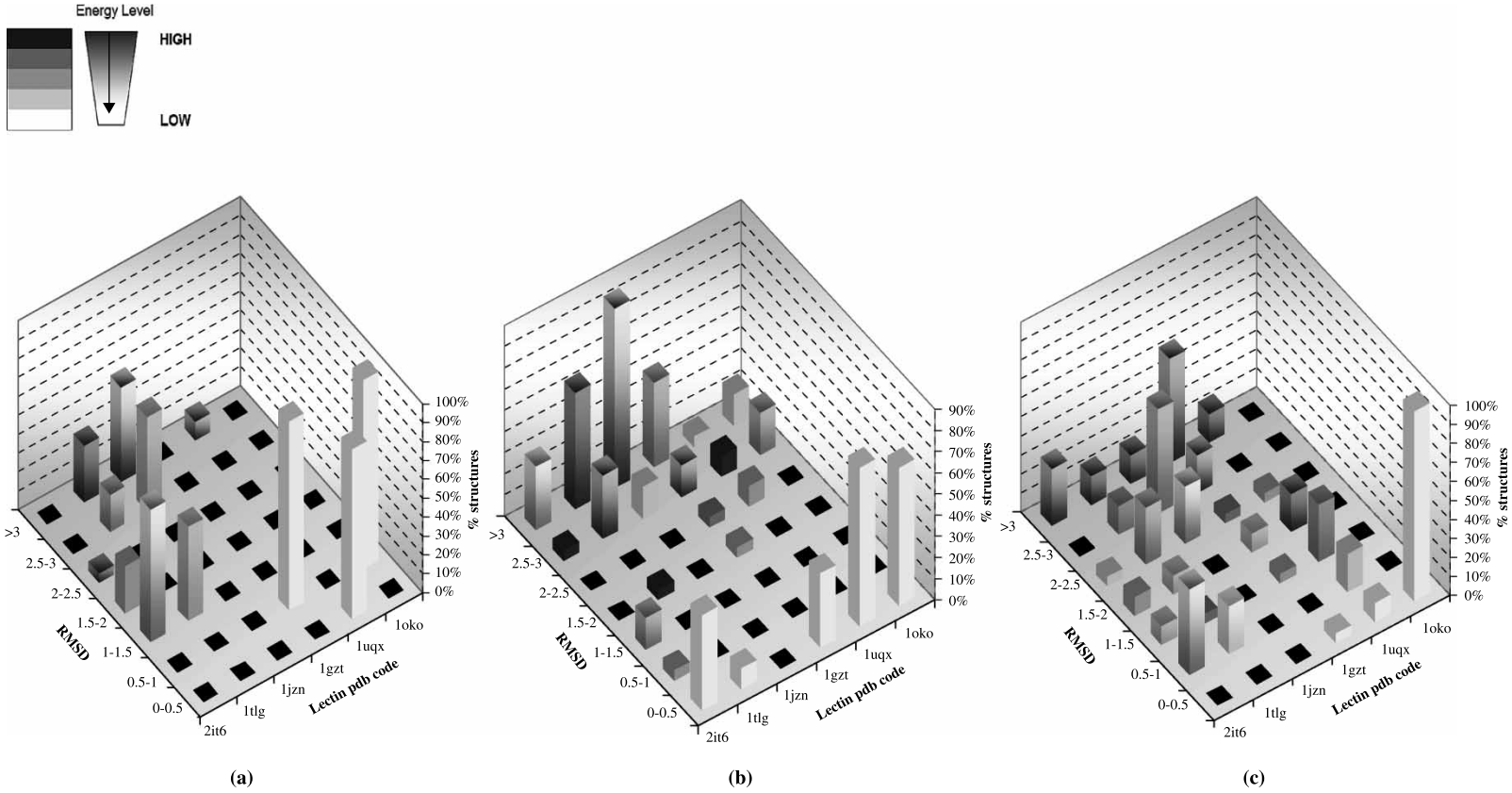

Figure 2. Graphical visualisation of docking simulation results: prediction accuracy of three different programs, AutoDock (a), DOCK (b) and GLIDE (c).

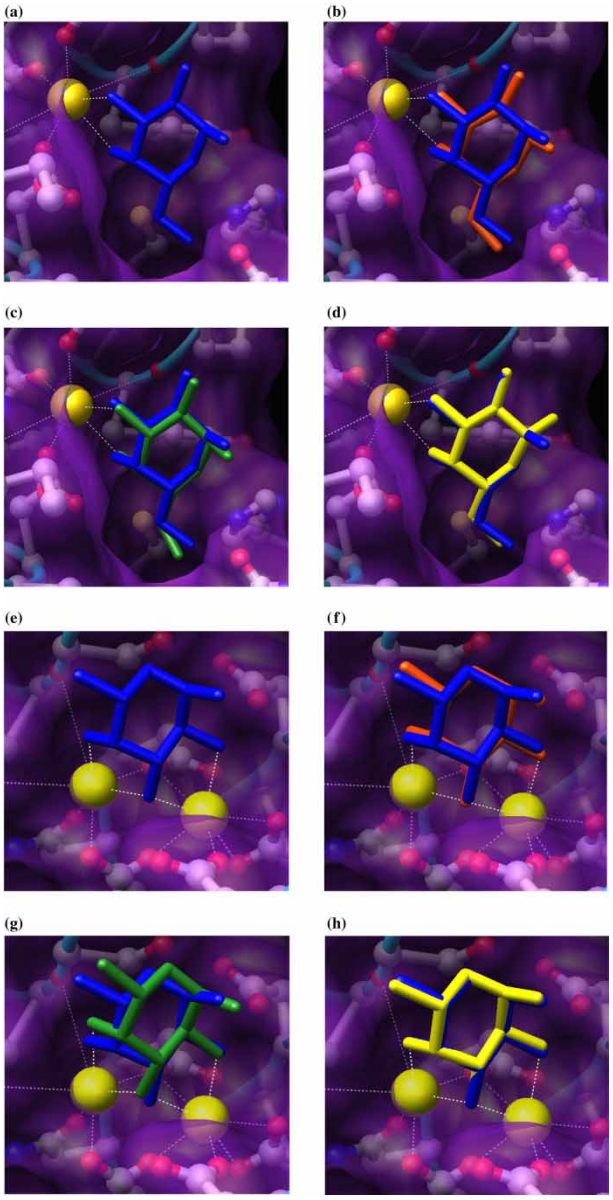

Figure 3. (a)–(d): Docking of α-D-galactose in PA-IL binding site in presence of one calcium ion; crystal structure 1OKO, ((a), blue) and comparison with lowest RMSD poses identified by AutoDock ((b), orange), DOCK ((c), green) and GLIDE ((d), yellow). (e)–(f): Docking of α-L-fucose in PA-IIL binding site in presence of two calcium ions; crystal structure 1GZT ((e), blue) comparison with lowest RMSD poses identified by AutoDock ((f), orange), DOCK ((g), green) and GLIDE ((h), yellow).

3.4 *GLIDE files and parameters*

The GLIDE program [23] was used in the First Discovery 2.7 package from Schrödinger, Inc. Missing hydrogens in the protein structure were added by Protein preparation routine implemented in First Discovery. Binding site was defined on the base of the available lectin structures with the bonded ligand. The box size was set to 14 Å in all three dimensions centred on the ligand. During the grid generation the parameter for Van der Waals radii scaling was scaled by 1.00 for atoms with partial atomic charge less then 0.25 and no constrains were defined. During the docking procedure the OPLS2001 partial atomic charges where assigned to the ligands. The parameters for Van der Waals radii scaling in docking was scaled by 0.80 for atoms with partial atomic charge less then 0.15. Ligand's poses were clustered within RMSD less then 0.5 Å and within maximum atomic displacement less then 1.3 Å. After the docking procedure 20 poses with the best score have been saved and used for the analyses. Docking results were analysed by GLIDE pose viewer included in MAESTRO (Schrödinger, Inc., NY).

3.5 *Calculation of RMSD*

We measured the RMSD between crystal and docked structures considering only the heavy atoms. All the values were determined using MAESTRO (Schrödinger, Inc., NY) and its *Superpose* command.

4. Results and discussion

4.1 *Comparison of predicted binding modes and crystal structures*

Since the aim of the study is to analyse how AutoDock, DOCK and GLIDE dock carbohydrates into the active sites characterised by calcium ions (i.e. the key heteroatoms implicated in sugar recognition), we first evaluated the orientation of the docked structures by comparing with crystal structures of complexes. Therefore the RMSD between crystal structure and lowest energy docking mode (and eventually most populated clusters) have been calculated and reported in Figure 2.

For AutoDock results (Figure 2(a)), histograms of cluster demonstrate that the binding mode of monosaccharides in bacterial lectins with one or two calcium (1oko, 1uqx and 1gzt) were very well predicted, within a window of 0.5 and 1 Å with respect to experimental data (Figure 3 (b),(f)). The best results were obtained for docking the monosaccharide α-methyl-mannoside (αMeMan) in *R. solanacearum* lectin (1uqx) binding site: 90% of carbohydrate conformations showed a RMSD which did not exceed 0.5 Å from the corresponding crystal structure while only 10% exceeded a rms value of 3 Å. Agreement was not so good for C-type lectins since for the rattlesnake venom lectin (1jzn), docking simulation gave 50% of structures with a RMSD between 2.5 and 3 Å and 50% with a RMSD of more than 3 Å. The tunicate lectin (1tlg) presented 50% of poses derived from the fusion of two clusters with different energy value and population number with a RMSD between 1 and 1.5 Å, followed by a percentage of 20 (RMSD 2.5–3 Å) and 30 (RMSD $>$ 3 Å) related to the other possible solutions.

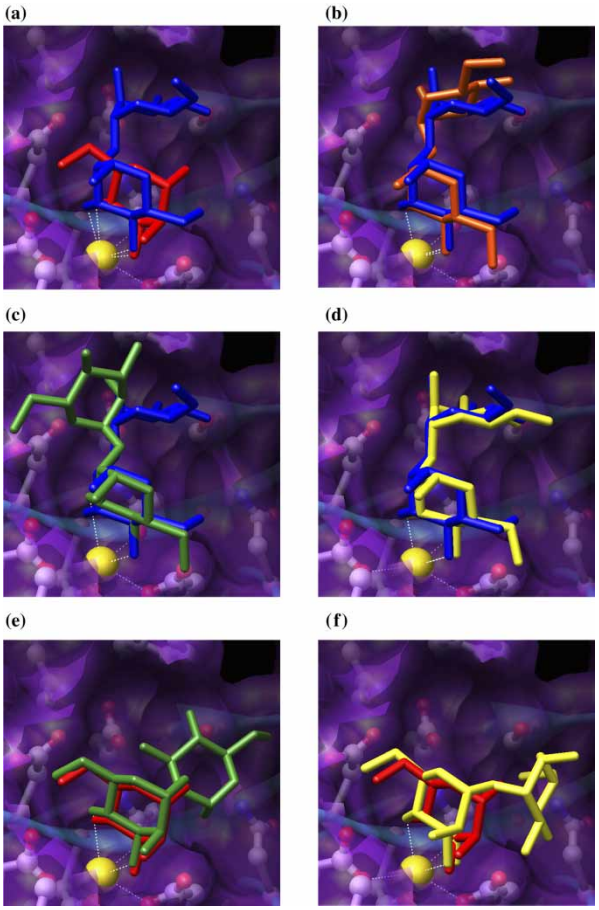

Figure 4. Docking of mannobiose in DC-SIGN binding site in presence of one calcium; crystal structure 2IT6 ((a), blue and red for the two binding modes observed in the crystal) and comparison with lowest RMSD poses identified by AutoDock ((b), orange), DOCK ((c), green), GLIDE ((d), yellow structure). Only DOCK and GLIDE (e, f) are able to reproduce the minor orientation observed for mannose in the crystal ((a), red).

However, for the human DC-SIGN (2it6), docking results always adopted the main conformation of mannobiose (Figure 4(b)) with a RMSD between 1 and 1.5 in 70% of cases. Each best pose (i.e. lower RMSD) derived from AutoDock was always the pose which had the highest number of cluster population in cluster histograms.

For DOCK results (Figure 2(b)), the best results were also obtained with bacterial lectins since for 1oko and 1uqx, the highest number of structures (65, 75%, respectively) were grouped in a single cluster with a RMSD of less then 0.5 Å. The high quality of the prediction is depicted in Figure 3(c), (d), for one-calcium and two-calcium bacterial lectins, respectively. Both major and minor orientation of the carbohydrate ligand in complex with the lectin DC-SIGN (2it6) were predicted with 35% of results fitting with the major binding mode in crystal (Figure 4(c)), 45% according with the minor binding mode (Figure 4(e)) while the rest of solutions had

Table 2. Experimental data and best docked structures: comparison in terms of root-mean-square deviation (RMSD).

Lectin PDB code	Autodock	DOCK	GLIDE
1oko	0.74	0.34	0.67
1uqx	0.43	0.41	0.29
1gzt	0.53	0.32	0.18
1jzn	2.92	2.62	2.11
1tlg	1.42	0.29	0.72
2it6 (a)	1.01	2.61	0.73
2it6 (b)[a]	–	0.48	0.55
RMSD average	1.18	1.01	0.75

[a] This row identifies the values derived from the second possible orientation of the carbohydrate in the binding site.

RMSD larger than 3 Å. Only a small cluster represented the right lactose orientation in snake lectin (1jzn) binding site with a RMSD between 2.5 and 3 Å. This group did not have the lowest energy conformation value. The same situation was noticed in docking tunicate lectin (1tlg): only 10% of poses had a RMSD which did not exceed 0.5 Å with respect of the known galactose binding mode.

In GLIDE docking analysis (Figure 2(c)), most results for the lectin PA-IL (1oko) displayed a RMSD less then 0.5 Å (Figure 3(d)). The carbohydrate orientation in RS-IIL and PA-IIL (Figure 3(h)) was also well recognised in 10% and 5% of cases, exhibiting the lowest energy conformation in each situation. The orientation of the disaccharide lactose in rattlesnake venom lectin (1jzn) binding site was reproduced in 30% of docking solutions with a RMSD between 2 and 2.5 Å. Most of results gave a RMSD between 2.5 and 3 Å (55%) followed by 15% of poses with a RMSD over 3 Å. For the tunicate lectin (1tlg), GLIDE showed 25% of results with a RMSD between 0.5 and 1 Å and favourable energy conformation. As for DC-SIGN, 43% of structures predicted the minor sugar orientation of mannobiose in 2it6 binding site, related to the major one by a 180° rotation, with a RMSD between 0.5 and 1 Å (Figure 4(d),(f)). In general, in each docking simulation performed by GLIDE, the lowest energy pose matched the conformation with the lowest RMSD, except few exceptions (Table 3) in which lowest energy poses did not differ so much with respect to crystallographic information.

Quantification of geometrical differences between the crystal structure and the closer docking mode is given in Table 2. Averaging over the six crystal structures analysed here indicated good performance of the three programs with rather limited variations among them. GLIDE displayed the lower mean RMSD (0.75 Å) with the two other ones giving values slightly higher than 1 Å. It should be noted that in the present state, the absolute values of the docking energy reported by each program are not comparable.

Table 3. Distances in Å between oxygen atoms of best docked carbohydrates and calcium ions in lectin binding site. Comparison with crystallography information.

		Crystal structure	Auto Dock	DOCK	GLIDE
1oko	Ca–O3 Gal	2.46	2.30	2.66	2.62
	Ca–O4 Gal	2.50	2.75	2.57	2.34
1uqx	Ca 1–O2 αMeman	2.51	2.32	3.07	2.92
	Ca 1–O3 αMeman	2.50	2.30	2.49	2.41
	Ca 2–O3 αMeman	2.52	2.43	2.83	2.79
	Ca 2–O4 αMeman	2.55	2.41	2.75	2.48
1gzt	Ca 1–O2 Fuc	2.54	2.24	2.52	2.77
	Ca 1–O3 Fuc	2.47	2.40	2.94	2.45
	Ca 2–O3 Fuc	2.48	2.32	2.53	2.49
	Ca 2–O4 Fuc	2.47	2.28	3.32	2.50
1jzn	Ca 1–O3 Gal	2.58	3.14	3.48	2.90
	Ca 1–O4 Gal	2.71	2.39	2.67	3.98
1tlg	Ca 1–O3 Gal	2.46	2.32	2.54	2.59
	Ca 1–O4 Gal	2.48	3.17	2.59	3.26
2it6	Ca 1–O3 Man	2.54	2.37	3.53	2.57
	Ca 1–O4 Man	2.48	2.27	2.56	2.81
2it6(b)	Ca 1–O3 Man	2.34	–	2.51	2.46
	Ca 1–O4 Man	2.62		3.13	2.65
Average relative error (%)			9.5	12.4	9.7

4.2 Comparison of calcium coordination

Since the direct involvement of calcium ion in carbohydrate binding is the characteristic feature of the lectins studied here, we analysed the ability of docking programs to predict co-ordination bonds with calcium ions. For the best docked solution, distances between the oxygen atoms from carbohydrate and the calcium ions are listed in Table 3. While the hydroxyl group oxygen atoms of carbohydrate displayed an average value of 2.5 Å to calcium ions in the crystalline state (2.49 Å ± 0.03), the docking program AutoDock had a tendency to shorten these distances (2.31 Å ± 0.07). Both DOCK (2.82 Å ± 0.38), and GLIDE (2.55 Å ± 0.15), would elongate them, and the results for both program also showed a larger distribution of values, indicating that this type of bond is maybe not sufficiently constrained in the parameterisation of these two programs. When calculating the average relative error in oxygen–calcium bond lengths, GLIDE, AutoDock and DOCK yielded values of 9.51, 12.39 and 9.70%, respectively. Thus AutoDock was the most accurate in keeping carbohydrate-calcium ion coordination distance immediately followed by GLIDE while DOCK accuracy was not so high.

It should be noted that the geometrical comparison performed above has been done using the solution with lowest RMSD compared to crystal structure. However, this "correct" solution, was not always the lowest energy one, and the programs have also been compared for their

Table 4. Heavy atom RMSDs and energies of best docked carbohydrate conformations. Italic entries identify structures in which best RMSD and lowest energy poses do not correspond.

		AutoDock RMSD/energy	DOCK RMSD/energy	GLIDE RMSD/energy
1oko	Lowest RMSD (Å)	0.74/−7.53	0.34/−35.78	0.67/−64.2
	Lowest energy (kcal/mol)			
1uqx	Lowest RMSD (Å)	0.43/−8.04	0.41/−42.82	0.29/−88.4
	Lowest energy (kcal/mol)			
1gzt	Lowest RMSD (Å)	0.53/−6.89	0.32/−36.91	0.18/−91.2
	Lowest energy (kcal/mol)			
1jzn	Lowest RMSD (Å)	*2.92/−8.30*	*2.62/−45.06*	*2.11/−62.0*
	Lowest energy (kcal/mol)	*> 4.00/−8.61*	*3.27/−46.42*	*2.35/−80.4*
1tlg	Lowest RMSD (Å)	*1.42/−6.45*	*0.29/−29.67*	*0.72/−36.5*
	Lowest energy (kcal/mol)	*> 4.00/−6.58*	*> 4.00/−31.20*	*0.79/−48.0*
2it6	Lowest RMSD (Å)	1.01/−7.75	2.61/−40.06	0.55/−55.9
	Lowest energy (kcal/mol)		0.48/−43.30[a]	0.73/−57.7[a]

[a] The asterisk corresponds to the second possible orientation of carbohydrate in the binding site.

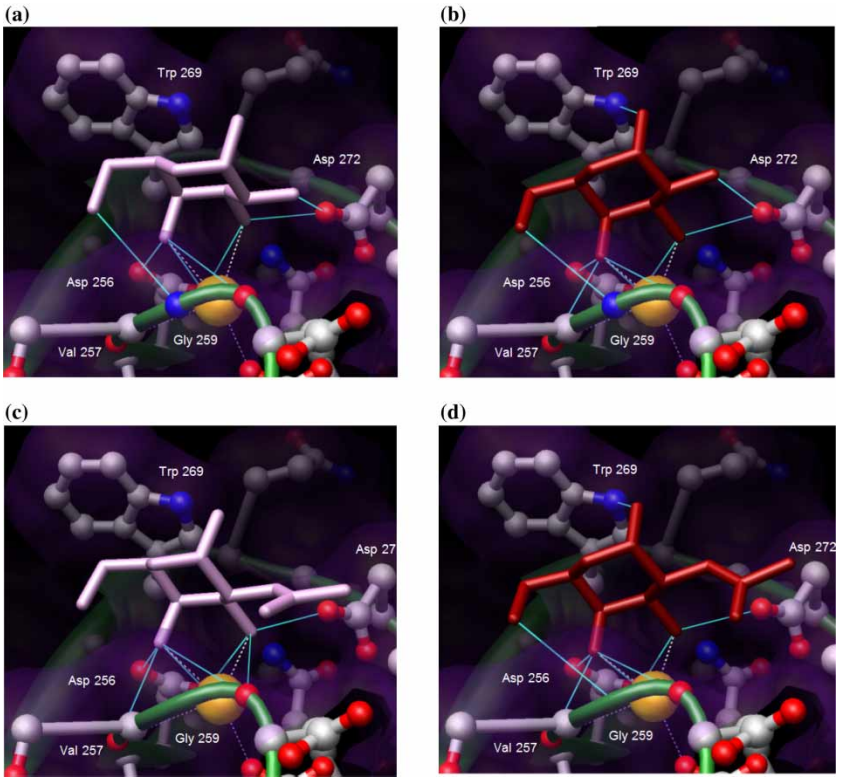

Figure 5. Modelling of the interation of galactose and *N*-acetylgalactosamine in CEL-III sub-domain 2γ according to GLIDE (plum, (a) for Gal and C for GalNAc) and AutoDock (red, (b) for Gal and (d) for GalNAc) prediction. Hydrogen bond network is coloured *cyan* while coordination bonds with calcium ion are showed in *white*.

ability to score the "correct" solution. Comparisons with best RMSD result and lowest energy one are listed in Table 4. All programs performed perfectly well for the bacterial lectins, but all of them failed in predicting the correct lowest energy conformation for C-type lectin from tunicate and snake. As seen above in Figure 2, no program could predict correctly the binding mode of lectin in tunicate lectin (1jzn). For the snake lectin (1tlg), DOCK and GLIDE could predict the correct binding with a reasonable energy cost. Alternatively, AutoDock had the advantage to predict correctly the docking mode when looking at the most populated cluster, instead of looking at the energy level.

4.3 Prediction of monosaccharide binding in sea cucumber CEL-III lectin

The sea cucumber (*Cucumaria echinata*) lectin CEL-III is a member of a new family of calcium-dependent lectin that have not yet been crystallised as complex with the ligand. The lectin adopts a double β-trefoil fold [16], also referred as R-type lectin and it is specific for galactose and *N*-acetylgalactosamine. Due to tandem repeats in the structure, several binding sites are predicted to occur and, indeed, five calcium ions have been located in putative binding sites.

According to the results described above, we used GLIDE and AutoDock for the prediction of Gal and GalNAc orientation in CEL-III. We focused our docking conformational search on the 2 γ-subdomain (named by reference to the ricin structure), concentrating on binding modes that present the lowest energy conformation for GLIDE software and the highest population number of its respective cluster in AutoDock. The resulting models are displayed in Figure 5 and the resulting hydrogen bond network and coordination geometries are listed in Tables 5 and 6. When docking galactose in CEL-III with either GLIDE or AutoDock, the resulting models indicated that O3 and O4 galactose oxygen atoms are involved directly in the coordination of calcium ion. From hydrogen bonds analysis, we found that the residue Asp 256 play a key role in the binding site forming hydrogen bonds with O3 and O4 hydroxyl groups and participating in calcium co-ordination. The residue Asp272 establishes contacts with O2 and O3 of galactose and coordinates calcium ion whereas Trp269 is involved in aromatic stacking interactions with carbohydrate ring. The large interaction area created by this interaction seems to strongly influence the sugar orientation in CEL-III binding site. The docking

Table 5. Hydrogen bonds and calcium coordination for α-D-galactose docked in sub-domain 2γ of CEL-III using GLIDE and AutoDock programs. The interatomic distances are reported in Å, weak hydrogen bonds (>3.1 Å) are indicated in italics.

Atom 1	Atom 2	GLIDE	AutoDock
Hydrogen bonds			
Gal.O1	Trp 269.NE1	–	2.74
Gal.O2	Asp 272.OD2	2.72	2.78
Gal.O3	Asp 256.OD1	2.92	2.48
Gal.O3	Asp 272.OD2	2.57	2.73
Gal.O4	Asp 256.OD1	*3.25*	3.02
Gal.O4	Asp 256.OD2	2.85	2.70
Gal.O4	Val 257.O	–	2.78
Gal.O4	Gly 259.O	*3.41*	3.07
Gal.O6	Gly 259.N	*3.18*	2.99
Coordination bonds			
O3.Gal	Ca	2.81	2.35
O4.Gal	Ca	2.55	2.17

Table 6. Hydrogen bonds and calcium coordination for α-D-GalNAc docked in sub-domain 2γ of CEL-III using GLIDE and AutoDock programs. The interatomic distances are reported in Å, weak hydrogen bonds (> 3.1 Å) are indicated in italics.

Atom 1	Atom 2	GLIDE	AutoDock
Hydrogen bonds			
Gal.O1	Trp 269.NE1	–	2.77
Gal.O3	Asp 272.OD2	2.66	2.80
Gal.O3	Asp 256.OD1	*3.15*	2.60
Gal.O4	Asp 256.OD1	*3.28*	*3.23*
Gal.O4	Asp 256.OD2	2.76	2.81
Gal.O4	Val 257.O	*3.12*	2.81
Gal.O4	Gly 259.O	*3.35*	3.08
Gal.O6	Gly 259.N	–	3.07
Coordination bonds			
O3.Gal	Ca	2.29	2.28
O4.Gal	Ca	2.56	2.35

of GalNAc did not present large variations as compared to galactose. Same contacts were maintained with the exception of the hydrogen bond between O2 of galactose and Asp272 that could not occur for GalNAc.

5. Conclusions

We evaluated and compared the reliability of three different softwares in performing flexible docking using calcium-dependent lectins as receptors and carbohydrates as ligands. This study revealed that all docking programs demonstrated the ability to accomplish docking work. The commercial package GLIDE slightly outperformed the other academic available docking engines in terms of RMSD from experimental structures but the best result did not always correspond to the lowest energy conformation. This conclusion is in agreement with recent evaluation study for docking carbohydrate derivative to Zn-containing enzyme [37]. Apart from showing a lower docking accuracy than GLIDE in terms of RMSD, DOCK program presented the same results as those noticed in GLIDE. The academic program AutoDock gave good results in tested simulations and clustering histograms created by the software always estimated the right carbohydrate position according to experimental information. Further improvements could be performed by optimising the docking parameters. Also results would be more statistically reliable by significantly increasing the number of runs. In the future, it will be of interest to also test the ability of these programs for docking large flexible oligosaccharides to calcium-dependent lectins and to evaluate the agreement between computed energies and known affinities.

Acknowledgements

The PhD thesis of A.N. and the stay of S.K. in Grenoble are financed by EEC Marie-Curie training program MEST-CT-2004-503322. This work was supported by the European Community's Human Potential Programme as a Research Training Network (Contract N° MRTN-CT-2006-035546[NODPERCEPTION]). The authors are thankful to Dr. Serge Pérez for his careful reading of the manuscript.

Note

1. Affiliated with the Joseph Fourier of Grenoble and Member of the Institut de Chimie Moléculaire de Grenoble.

References

[1] H. Lis and N. Sharon, *Lectins: carbohydrate-specific proteins that mediate cellular recognition*, Chem. Rev. 98 (1998), p. 637.
[2] H.J. Gabius, S. Andre, H. Kaltner, and H.C. Siebert, *The sugar code: functional lectinomics*, Biochim. Biophys. Acta 1572 (2002), p. 165.
[3] R. Loris, *Principles of structures of animal and plant lectins*, Biochim. Biophys. Acta 1572 (2002), p. 198.
[4] J.M. Rini, *Lectin structure*, Annu. Rev. Biophys. Biomol. Struct. 24 (1995), p. 551.
[5] K. Drickamer and M.E. Taylor, *Biology of animal lectins*, Annu. Rev. Cell Biol. 9 (1993), p. 237.
[6] A. Cambi, M. Koopman, and C.G. Figdor, *How C-type lectins detect pathogens*, Cell Microbiol. 7 (2005), p. 481.
[7] K. Drickamer and A.J. Fadden, *Genomic analysis of C-type lectins*, Biochem. Soc. Symp. (2002), p. 59.
[8] K. Drickamer, *Ca(2 +)-dependent sugar recognition by animal lectins*, Biochem. Soc. Trans. 24 (1996), p. 146.
[9] N. Gilboa-Garber, Pseudomonas aeruginosa *lectins*, Methods Enzymol. 83 (1982), p. 378.
[10] A. Imberty, M. Wimmerova, C. Sabin, and E.P. Mitchell, *Structures and roles of* Pseudomonas aeruginosa *lectins*, in *Protein–Carbohydrate Interactions in Infectious Disease*, C. Bewley ed., The Royal Society of Chemistry, Cambridge, 2006, p. 30.
[11] G. Cioci, E.P. Mitchell, C. Gautier, M. Wimmerova, D. Sudakevitz, S. Pérez, N. Gilboa-Garber, and A. Imberty, *Structural basis of calcium and galactose recognition by the lectin PA-IL of* Pseudomonas aeruginosa, FEBS Lett. 555 (2003), p. 297.
[12] E. Mitchell, C. Houles, D. Sudakevitz, M. Wimmerova, C. Gautier, S. Pérez, A.M. Wu, N. Gilboa-Garber, and A. Imberty, *Structural basis for oligosaccharide-mediated adhesion of* Pseudomonas aeruginosa *in the lungs of cystic fibrosis patients*, Nat. Struct. Biol. 9 (2002), p. 918.

[13] E.P. Mitchell, C. Sabin, L. Šnajdrová, M. Pokorná, S. Perret, C. Gautier, C. Hofr, N. Gilboa-Garber, J. Koča, M. Wimmerová, and A. Imberty, *High affinity fucose binding of* Pseudomonas aeruginosa *lectin PA-IIL: 1.0 Å resolution crystal structure of the complex combined with thermodynamics and computational chemistry approaches*, Proteins: Struct. Funct. Bioinform. 58 (2005), p. 735.

[14] M. Pokorná, G. Cioci, S. Perret, E. Rebuffet, N. Kostlánová, J. Adam, N. Gilboa-Garber, E.P. Mitchell, A. Imberty, and M. Wimmerová, *Unusual entropy driven affinity of* Chromobacterium violaceum *lectin CV-IIL towards fucose and mannose*, Biochemistry 45 (2006), p. 7501.

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

[16] T. Uchida, T. Yamasaki, S. Eto, H. Sugawara, G. Kurisu, A. Nakagawa, M. Kusunoki, and T. Hatakeyama, *Crystal structure of the hemolytic lectin CEL-III isolated from the marine invertebrate* Cucumaria echinata*: implications of domain structure for its membrane pore-formation mechanism*, J. Biol. Chem. 279 (2004), p. 37133.

[17] A. Imberty and S. Pérez, *Structure, conformation and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations*, Chem. Rev. 100 (2000), p. 4567.

[18] S. Pérez, A. Imberty, S.B. Engelsen, J. Gruza, K. Mazeau, J. Jiménez-Barbero, A. Poveda, J.F. Espinosa, B.P. van Eyck, G. Johnson et al., *A comparison and chemometric analysis of several molecular mechanics force fields and parameters sets applied to carbohydrates*, Carbohydr. Res. 314 (1998), p. 141.

[19] P.M. Coutinho, M.K. Dowd, and P.J. Reilly, *Automated docking of glucosyl disaccharides in the glucoamylase active site*, Proteins 28 (1997), p. 162.

[20] A. Laederach and P.J. Reilly, *Modeling protein recognition of carbohydrates*, Proteins 60 (2005), p. 591.

[21] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, and A.J. Olson, *Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function*, J. Comp. Chem. 19 (1998), p. 1639.

[22] D.T. Moustakas, P.T. Lang, S. Pegg, E. Pettersen, I.D. Kuntz, N. Brooijmans, and R.C. Rizzo, *Development and validation of a modular, extensible docking program: DOCK5*, J. Comput. Aided Mol. Des. 20 (2006), p. 601.

[23] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry et al., *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy*, J. Med. Chem. 47 (2004), p. 1739.

[24] F.J. Solis and R.J.-B. Wets, *Minimization by random search techniques*, Math. Oper. Res. 6 (1981), p. 19.

[25] T.J.A. Ewing and I. D. Kuntz, *Critical evaluation of search algorithms for automated molecular docking and database screening*, J. Comp. Chem. 18 (1997), p. 1175.

[26] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case, *Development and testing of a general Amber force field*, J. Comput. Chem. 25 (2004), p. 1157.

[27] J.A. Nelder and R. Mead, *A Simplex-method for function minimization*, Comput. J. 7 (1964), p. 308.

[28] T.A. Halgren, R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard, and J.L. Banks, *GLIDE: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening*, J. Med. Chem. 47 (2004), p. 1750.

[29] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The protein data bank*, Nucleic Acids Res. 28 (2000), p. 235.

[30] SYBYL, Tripos Associates, 1699 S. Hanley Road, Suite 303, St Louis, MO 63144 USA.

[31] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M.J. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules*, J. Am. Chem. Soc. 117 (1995), p. 5179.

[32] M. Clark, R.D.I. Cramer, and N. van den Opdenbosch, *Validation of the general purpose Tripos 5.2 force field*, J. Comput. Chem. 10 (1989), p. 982.

[33] J. Aqvist, *Ion–water interaction potentials derived from free energy perturbation simulations*, J. Phys. Chem. 94 (1990), p. 8021.

[34] A. Imberty, E. Bettler, M. Karababa, K. Mazeau, P. Petrova, and S. Pérez, *Building sugars: the sweet part of structural biology*, in *Perspectives in Structural Biology*, M. Vijayan, N. Yathindra & A.S. Kolaskar, eds., Indian Academy of Sciences and Universities Press, Hyderabad, 1999, p. 392.

[35] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin, *UCSF Chimera—a visualization system for exploratory research and analysis*, J. Comput. Chem. 25 (2004), p. 1605.

[36] J. Wang, W. Wang, P.A. Kollman, and D.A. Case, *Automatic atom type and bond type perception in molecular mechanical calculations*, J. Mol. Graph. Model. 25 (2006), p. 247.

[37] P. Englebienne, H. Fiaux, D.A. Kuntz, C.R. Corbeil, S. Gerber-Lemaire, D.R. Rose, and N. Moitessier, *Evaluation of docking programs for predicting binding of Golgi alpha-mannosidase II inhibitors: a comparison with crystallography*, Proteins, 69 (2007), p. 160.

[38] S.F. Poget, G.B. Legge, M.R. Proctor, P.J. Butler, M. Bycroft, and R.L. Williams, *The structure of a tunicate C-type lectin from* Polyandrocarpa misakiensis *complexed with* D*-galactose*, J. Mol. Biol. 290 (1999), p. 867.

[39] J.R. Walker, B. Nagar, N.M. Young, T. Hirama, and J.M. Rini, *X-ray crystal structure of a galactose-specific C-type lectin possessing a novel decameric quaternary structure*, Biochemistry 43 (2004), p. 3783.

[40] H. Feinberg, R. Castelli, K. Drickamer, P.H. Seeberger, and W.I. Weis, *Multiple modes of binding enhance the affinity of DC-SIGN for high mannose* N*-linked glycans found on viral glycoproteins*, J. Biol. Chem. 282 (2007), p. 4202.