Glycobiology, 2018, 1–13
doi: 10.1093/glycob/cwy102
Glyco-Informatics

Computational Biology

An efficient use of X-ray information, homology modeling, molecular dynamics and knowledge-based docking techniques to predict protein–monosaccharide complexes

Juan I Blanco Capurro[2,3,†], Matias Di Paola[2,3,†], Marcelo Daniel Gamarra[2,3], Marcelo A Martí[2,3,1], and Carlos P Modenutti[2,3,1]

[2]Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Intendente Guiraldes 2160, C1428EGA, Ciudad Autónoma de Buenos Aires, Argentina and [3]Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), CONICET, Ciudad Universitaria, Intendente Guiraldes 2160, C1428EGA, Ciudad Autónoma de Buenos Aires, Argentina

[1]To whom correspondence should be addressed: cmodenutti@qb.fcen.uba.ar; marti.marcelo@gmail.com

[†]Both authors contributed equally to this work.

Received 12 May 2018; Revised 19 October 2018; Editorial decision 30 October 2018; Accepted 6 November 2018

Abstract

Unraveling the structure of lectin–carbohydrate complexes is vital for understanding key biological recognition processes and development of glycomimetic drugs. Molecular Docking application to predict them is challenging due to their low affinity, hydrophilic nature and ligand conformational diversity. In the last decade several strategies, such as the inclusion of glycan conformation specific scoring functions or our developed solvent-site biased method, have improved carbohydrate docking performance but significant challenges remain, in particular, those related to receptor conformational diversity. In the present work we have analyzed conventional and solvent-site biased autodock4 performance concerning receptor conformational diversity as derived from different crystal structures (apo and holo), Molecular Dynamics snapshots and Homology-based models, for 14 different lectin–monosaccharide complexes. Our results show that both conventional and biased docking yield accurate lectin–monosaccharide complexes, starting from either apo or homology-based structures, even when only moderate (45%) sequence identity templates are available. An essential element for success is a proper combination of a middle-sized (10–100 structures) conformational ensemble, derived either from Molecular dynamics or multiple homology model building. Consistent with our previous works, results show that solvent-site biased methods improve overall performance, but that results are still highly system dependent. Finally, our results also show that docking can select the correct receptor structure within the ensemble, underscoring the relevance of joint evaluation of both ligand pose and receptor conformation.

Key words: carbohydrates, docking, homology-modeling, lectin, molecular dynamics

Introduction

Lectins are sugar-binding proteins, characterized by the presence of a carbohydrate recognition domain which harbors the carbohydrate binding site (CBS). They are ubiquitous, involved in many important biological activities, and have been proposed and validated as therapeutic targets (Orozco et al. 2018). Understanding and predicting lectin–carbohydrate interactions is therefore of enormous relevance for Glycobiology.

Molecular Docking is the most extensively used strategy for predicting protein–ligand structures (Forli et al. 2016), and there are several packages available in the community (Morris et al. 1998; Verdonk et al. 2003; Friesner et al. 2004; Trott and Olson 2010; Ruiz-Carmona et al. 2014). Docking methods are usually composed of two main elements: (i) a conformational search algorithm, and (ii) a scoring function. The conformational search algorithm moves the relative ligand position with respect to the protein receptor, the ligand internal conformation and sometimes also the receptor (or part of it). Since the complexity of the problem scales exponentially with the number of conformational variables, usually the receptor is kept fixed and the ligand's internal conformation is significantly restricted. This can be a severe problem when the proper conformation of the receptor and/or ligand are unknown, and thus efficiently incorporating receptor flexibility in docking is an active research area (Nivedha et al. 2014). Concerning the scoring function, it is usually a parameterized additive force-field like an equation, which estimates the ligand binding energy (Morris et al. 1996; Friesner et al. 2004; Ruiz-Carmona et al. 2014).

Docking of carbohydrates to their lectin receptors is particularly challenging since they display low affinities (in the micromolar range) and their binding sites are shallow and highly hydrophilic (Kadirvelraj et al. 2008). Moreover, ligand conformational space is quite ample beyond the monosaccharide level and most important, scoring functions are usually trained with more hydrophobic drugs containing aromatic rings, and thus perform poorly for carbohydrates (Kerzmann et al. 2008). Previous works from our group showed that solvent structure in the CBS mimics the carbohydrate hydroxyl positions in the lectin–sugar complexes, and thus characterization of this structure by means of "Water Sites" allows improvement of docking performance, both in terms of accuracy and its capacity to detect the correct pose among all possibilities (Gauto et al. 2009; 2013; 2011; Guardia et al. 2011; López et al. 2015; Modenutti et al. 2015).

Water Sites (also called Hydration Sites or Solvent Sites) are defined as space regions adjacent to the protein surface, where the probability of finding a water molecule is significantly higher than in the bulk solvent. They can be derived from explicit water MD simulations and thermodynamically characterized using the "Inhomogeneous Fluid Solvation Theory" (Abel et al. 2008; Gauto et al. 2013). Alternatively, a simpler approach is to determine them by looking in apo-protein structures for the presence of crystallographic waters, which results in what we call crystallographic or X-ray derived Water Sites (Modenutti et al. 2015). Generally, x-ray and MD-derived WS coincide, but usually more sites are detected with MD simulations, plus they can be better characterized (Saraboji et al. 2012). Recently, we also showed that WS could be used to improve docking of common ligands and together with other probes (such as ethanol) allow improvement in the prediction of ligand binding free energies (Arcon et al. 2017).

Despite the mentioned improvements, several challenges remain for carbohydrate docking, particularly related to receptor and ligand flexibility, as well as the use of comparative or homology-based models when no crystal structure of the receptor is available. In the present work, we have analyzed Conventional Autodock Docking Method (CADM) and Solvent Site Biased Docking Method (SSBDM) performance when receptor flexibility is considered, and evaluated whether docking strategy can detect the correct receptor structure among an ensemble of structures derived from either MD and/or comparative models. Our results show that docking methods can select the correct lectin receptor structure within a conformational ensemble derived either from Molecular Dynamics or comparative modeling, underscoring the relevance of joint evaluation of pose, ligand and receptor conformation.

Results

The results are organized as follows: we first analyze the CADM and SSBDM performance for docking of monosaccharides on holo and apo x-ray structures on a set of fourteen monosaccharide-binding lectins. Secondly, we explore the performance of docking methods when no structural information is available, and receptors have to be built by comparative modeling, using templates of different global identity percentage (% GI). Finally, we analyze the impact of using a Molecular Dynamics or a comparative modeling based conformational ensemble.

Docking onto X-ray receptor structures

To start studying the lectin receptor conformation effect on the docking performance, we analyzed the results of the Conventional Autodock Docking Method (CADM) as well as Solvent Site Biased Docking Method (SSBDM), with either X-ray Solvent Sites (X-SSBDM) or MD derived Solvent Sites (MD-SSBDM). The receptor structures used were either the one derived from the corresponding Lectin–monosaccharide complex (i.e., the holo structure), or from another ligand-free structure (i.e., an apo structure). A detailed representative example of the results is presented in Table I for Jacalin (see Tables S1–S13 for other lectins). For each receptor structure type and docking method, we present both the most accurate result (regarding RMSD against the X-ray obtained pose) and highest ranked result. Since we perform several docking runs (100) for each receptor structure, we rank them using a combined energy and population Z-score, called "2D-score" (see Methods for details). In all experiments, we considered as the predicted ligand binding mode as "correct" whenever a ligand pose has a RMSD < 1.5 Å with respect to the holo x-ray pose. RMSD is measured considering just the pyranose ring heavy atoms.

The results for the 14 studied lectins in Table II show that when comparing holo vs. apo receptor structures and the different docking methods that performance is considerably system dependent, but in general and as expected, docking to the holo structure performs better than apo, a fact that might be attributed to the degree of "induced-fit" of each particular system. It is interesting to note that for most studied systems, at least one of the three docking methods can find the correct pose within the top three ranks. Challenging cases include lectins holo sna-II and the apo forms of aia, conA, hpa, jac and sp-D (see Table V in Methods section for full name and PDB id of all lectins), for which CADM fails to find the correct pose, although this situation is sometimes successfully reverted when employing either X-SSBDM or MD-SSBDM.

Concerning the comparison between conventional and biased methods, and as shown in our previous works (Gauto et al. 2013; Modenutti et al. 2015; Arcon et al. 2017), performance is highly

Table I. Docking method comparison for different Jacalin structures

Jacalin - (beta-D-galactose)			Docking Method	First Rank RMSD	Best RMSD[c]	First Rank 2D-Score	Best RMSD 2D-Score[a]	Best RMSD Rank[b]
Jacalin(Jac)	Holo (1ugw-A)		CADM	0.328	0.328	2.214	2.214	1
			X-SSBDM	0.28	0.28	2.77	2.77	1
			MD-SSBDM	0.253	0.253	2.994	2.994	1
	Apo (1ku8-A)		CADM	2.172	2.172	3.791	3.791	1
			X-SSBDM	2.148	0.786	3.268	0.872	3
			MD-SSBDM	2.163	1.099	2.033	0.96	3
	Template 79% GI (1jot-A)	Best DOPE score model (#052)	CADM	0.596	0.596	5.258	5.258	1
			X-SSBDM	0.621	0.621	2.974	2.974	1
			MD-SSBDM	0.562	0.562	5.524	5.524	1
	Template 59% GI (1xxq-A)	Best DOPE score model (#029)	CADM	0.929	0.929	1.81	1.81	1
			X-SSBDM	0.852	0.852	3.806	3.806	1
			MD-SSBDM	1.117	1.117	4.598	4.598	1
	Template 34% GI (2jz4-A)	Best DOPE score model (#033)	CADM	3.344	2.5	2.963	1.36	2
			X-SSBDM	3.388	3.388	1.269	1.269	1
			MD-SSBDM	3.319	2.509	2.992	1.245	2

Comparison of the different docking methods performance for holo and apo X-ray structures of Jacalin, and for comparative models using templates of different global identity percentage (% GI).

[a]2D-score is the combined Population and Binding Energy normalized score (see Methods). [b]Rank is the cluster ranking (ordered by 2D-score), and [c]RMSD is the sugar heavy atoms RMSD against the reference structure, after receptor alignment. [d]Docking performance on an MD generated receptor conformational ensemble.

Table II. Overall comparison of the single structure docking performance

		Structure			
Correct result (rmsd < 1.5)	Method	Holo	Apo	Model > 45% GI	Model < 45% GI
First Ranked	CADM	F17a-G, aia, bark, conA, fbm, hpa, jac (**7/14**)	F17a-G, bark, fbm, wga1 (**4/14**)	jac, conA, bark, sp-D (**4/12**)	fbm (**1/9**)
	X-SSBDM	F17a-G, aia, bark, fbm, hpa, jac, psl, sna-II, wga1 (**9/14**)	F17a-G, bark, fbm, aia, hpa, sna-II, wga1 (**7/14**)	jac, bark, fbm, sna-II (**4/12**)	– (**0/9**)
	MD-SSBDM	F17a-G, aia, bark, conA, ctb, fbm, hpa, jac, psl, sna-II, wga1 (**11/14**)	F17a-G, bark, fbm, gal7, hpa (**5/14**)	ctb, jac, conA, bark, sna-II (**5/12**)	– (**0/9**)
Within Top 3	CADM	ctb, hsi, psl, wga1 (**4/14**)	ctb (**1/14**)	ctb, fbm (**2/12**)	hsi (**1/9**)
	X-SSBDM	conA, ctb, hsi, sp-D (**4/14**)	conA, ctb, gal7, hsi, jac, psl (**6/14**)	F17a-G, aia, conA, ctb, sp-D (**5/12**)	hsi, fbm (**2/9**)
	MD-SSBDM	hsi, sp-D, gal7 (**2/14**)	conA, jac, psl, sna-II, spd, wga1 (**6/14**)	F17a-G, sp-D (**2/12**)	gal7, hsi (**2/9**)
Outside Top 3	CADM	gal7, sp-D (**2/14**)	gal7, hsi, psl, sna-II (**4/14**)	F17a-G, psl, sna-II, (**3/12**)	sp-D (**1/9**)
	X-SSBDM	gal7 (**1/14**)	sp-D (**1/14**)	psl (**1/12**)	cona, sp-D (**2/9**)
	MD-SSBDM	-(**0/14**)	ctb, hsi (**2/14**)	psl (**1/12**)	sp-D (**1/9**)
Not found	CADM	sna-II (**1/14**)	aia, conA, hpa, jac, sp-D (**5/14**)	wga1, aia, hsi (**3/12**)	conA, gal7, psl, bark,hpa, jac (**6/9**)
	X-SSBDM	– (**0/14**)	– (**0/14**)	wga1, hsi (**2/12**)	gal7, psl, bark, hpa, jac (**5/9**)
	MD-SSBDM	– (**0/14**)	aia(**1/14**)	wga1, aia, fbm, hsi (**4/12**)	conA, psl, bark, fbm,hpa, jac (**6/9**)

Comparison for each docking method for both X-ray structures and comparative Models for all 14 studied systems, showing for each method the name of successful cases for which the correct pose (rmsd < 1.5) was found for each rank category. Number in brackets indicate the fraction of successful cases. Models are further split into two categories according to the template Global Identity percentage (% GI).

a)2D-score is the combined Population and Binding Energy normalized score (see methods). b)Rank is the cluster ranking (ordered by 2D-score), and c)RMSD is the sugar heavy atoms RMSD against the reference structure, after receptor alignment. d)Docking performance on an MD generated receptor conformational ensemble.

system dependent, and although both biased methods can improve the results, still not always the correct pose comes ranked first. An example of improvement is shown in Figure 1 for holo sna-II, the only case in which CADM fails to find the correct pose, while X-SSBDM finds it (RMSD 0.4 Å) and ranks it in the first place. It is noteworthy that the best result CADM is able to achieve has an RMSD of 1.6 Å and corresponds to a different pose than that of the holo X-ray, indicating that the RMSD cut-off chosen (1.5) is adequate.

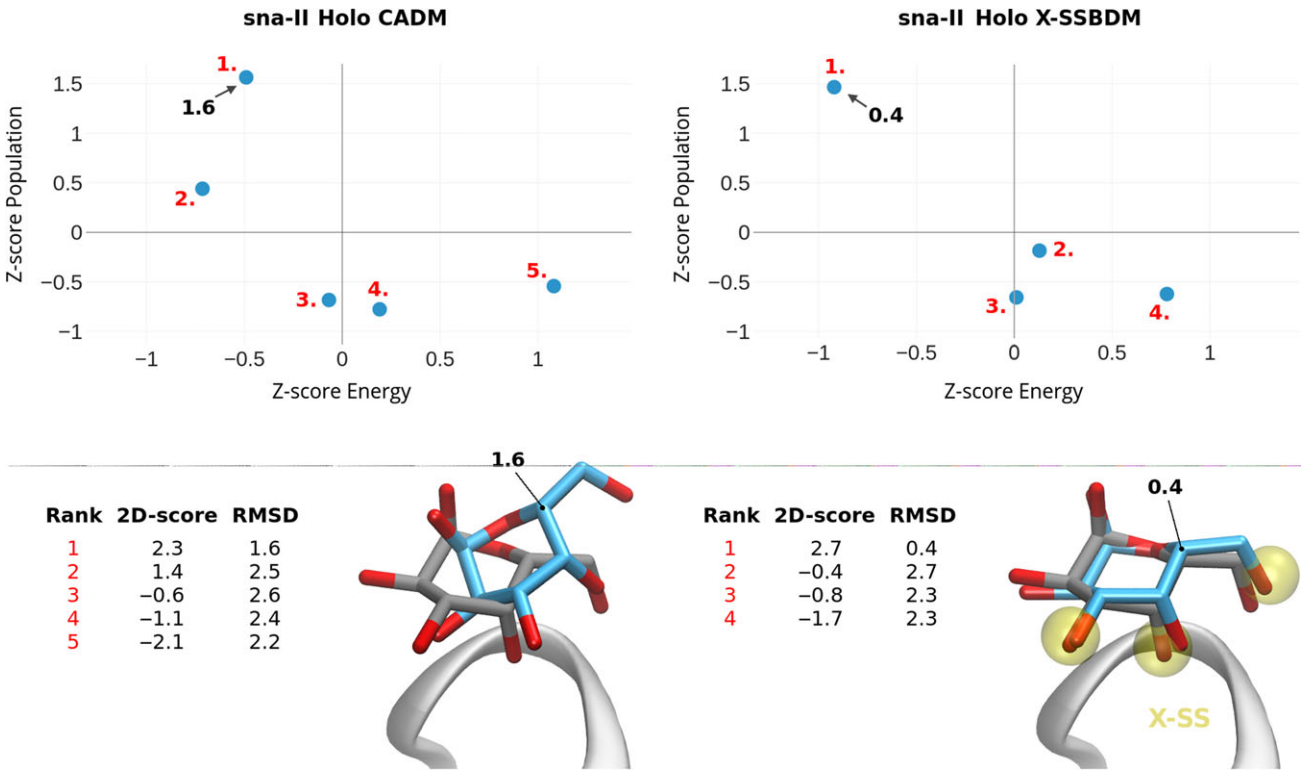

Fig. 1. "Z-score Population vs. Z-score binding energy" plots for the docking of Galactose to sna-II holo structure. *Left panel* shows CADM results, and *right panel* shows X-SSBDM results. All docking poses are ranked according to "2D-score" (see Methods), and mapped onto the plots with red numbers. Best RMSD poses are indicated with arrows, and illustrated with cyan sticks for carbon atoms. Reference holo structure ligand pose is drawn with gray sticks. Crystallographic solvent-sites (X-SS) are drawn as yellow spheres.

Apo docking is generally more difficult than holo (also called "cognate") docking, since in certain complexes sometimes ligand and receptor are conformationally adapted to each other to optimize their interactions. An example of this situation can be observed for Jacalin (Figure 2), where two binding-site tyrosine residues change their spatial orientation upon ligand binding and complex formation. As a result, CADM docking on apo Jacalin is unable to find the correct pose, but X-SSBDM rectifies this situation finding a 0.8 RMSD pose (although not ranked first).

Docking using comparative-modeling generated receptor structures

Given the considerable breadth between the number of known protein sequences (and their correct inclusion in protein families or domains) and the number of known structures, it is relatively common to use comparative (or homology) based models as receptor structures in protein–ligand docking studies. Even though it is possible to build a good lectin model based on moderate sequence identity template (Guardia et al. 2011), the details of the Binding site conformation are expected to vary significantly for different templates, even for different models built from the same template. Model quality can rely on different properties of the template, such as the structure resolution, the presence/absence of a similar ligand in the active site, and/or the global identity (GI). In order to assess the impact of this latter factor on docking performance, we built comparative models using different GI % templates. Briefly, softwares like Modeller allow to build several models from a single template, and model quality can be then compared according to an internal parameter called DOPE-score; The strategy we employed was to build 100 different models (using each template), and then select those with best DOPE-score to perform the docking. Results are shown for Jacalin in Table I, and for all systems in Table II. For clarity, in Table II models are grouped in two categories according to their template GI. The data clearly shows that high and moderate GI models (GI > 45%) perform almost as well as apo structures, while low GI models (GI < 45%) have a poorer performance, failing in most cases. Challenging systems are wga1, aia, fbm and hsi, this last two particularly, since for the low GI model the correct pose was indeed found while for the high GI it was not.

Receptor flexibility

We now turn our attention to receptor flexibility achieved using Molecular Dynamics (MD) simulations and analyze whether the docking technique can "fish a holo-like" receptor structure from within a conformational ensemble obtained by MD. Figure 3 (upper panel) and Table III (first three rows) present the results obtained for the docking of Beta D-Galactose to 100 different snapshots obtained from a 60 ns long MD simulation of Jacalin using the "apo form" as starting structure. The data shows that all three docking methods can detect the correct ligand pose among all MD generated receptor structures and within the first ranks, as evidenced by the 0.7 Å RMSD outlier cluster in the plot. Moreover, the obtained 2D-scores (with values in the 7.7–7.9 range, Table III) for the combined results of a hundred receptor conformations are significantly higher than those obtained for a single structure (2.9–5.2, Table I). This suggests that intrinsic normalization of the population and score (z-

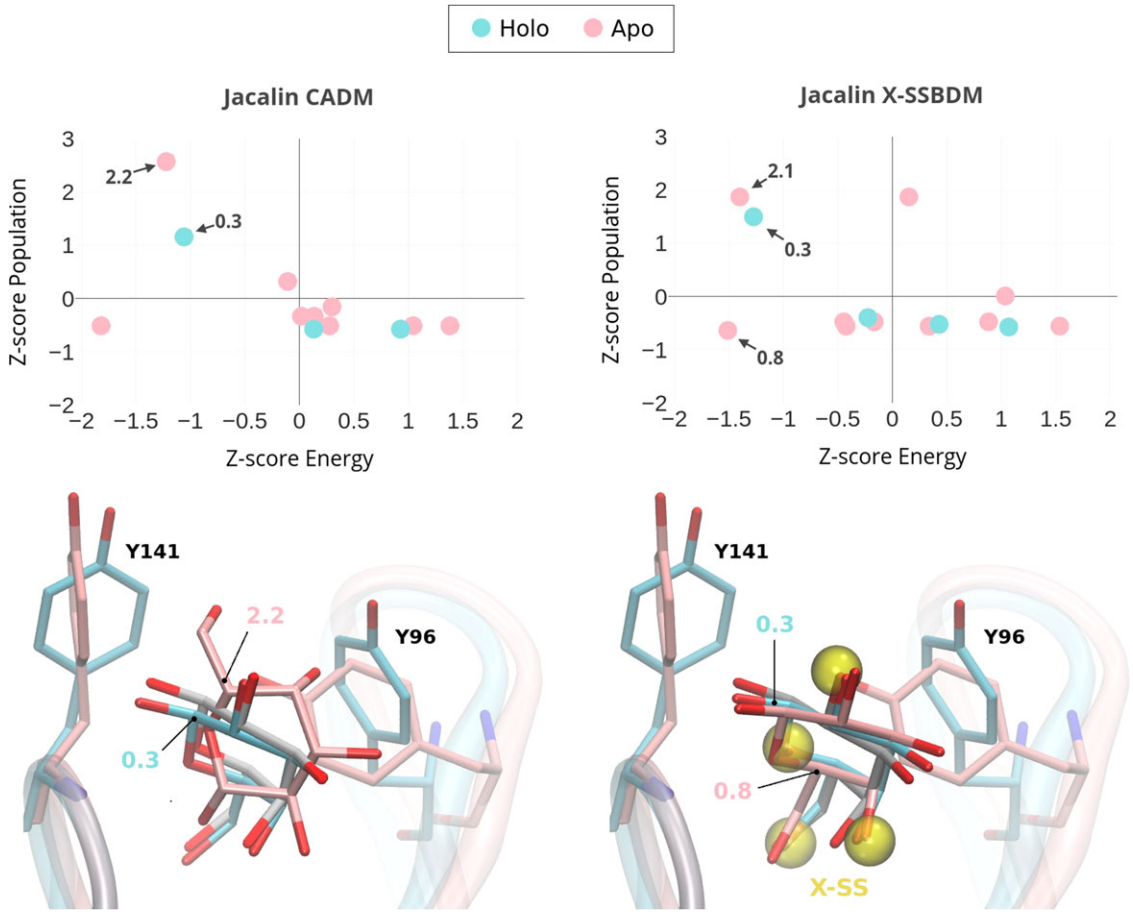

Fig. 2. "Z-score Population vs. Z-score binding energy" plots for the docking of Beta D-Galactose to Jacalin holo (cyan) and apo (pink) superimposed structures. *Left panel* shows CADM results, and *right panel* shows X-SSBDM results. Best RMSD poses are indicated with arrows, and illustrated with cyan sticks (holo poses) or pink sticks (apo poses). Relevant binding-site Tyrosine residues that exhibit different rotamers in holo and apo structures are also drawn with the same color code. Reference holo structure ligand pose is drawn with gray sticks. Crystallographic solvent-sites (X-SS) are drawn as yellow spheres.

function) using a conformational ensemble has the potential to increase the method sensitivity. Also important, analysis of the receptor structure of the best ligand pose in comparison with the holo reference structure (Figure 3, upper panel) shows that CADM was able to "fish" the right conformation of the binding site, as evidenced by a corresponding binding site RMSD showing values of 1.02 Å (for all binding site heavy atoms), and the "holo-like" conformation of the two key Tyrosine residues. These results underscore the MD ability to sample "holo-like" conformational structures, even when starting from a conformationally different state, such as the Jacalin apo form in this example.

Performance on other lectins is depicted in Table IV. Remarkably, when using an 100 Apo-MD snapshot ensemble the correct pose is found for all 14 cases, even when using just CADM (in contrast to X-ray apo structures); Nevertheless, correct poses (ring RMSD < 1.5) are half the time ranked outside the top 5, indicating that the use of an ensemble might sometimes decrease the docking's specificity.

Another valid strategy for ensemble building can be also comparative modeling, since it is fast and computationally cheap. In this context, we performed docking not only on the best-DOPE score models but on the whole 100 models previously built. The results are presented in Table IV. The results show that previously challenging/confusing cases (wga1, hsi, fbm) are now correctly predicted, presumably because the best DOPE-score single models were not conformationally optimal. As for Jacalin, in Figure 3 Z-score plot (lower panel) an 0.6 RMSD outlier corresponding to the correct ligand pose is present. Superimposition of the model structure "fished" by this pose with the holo X-ray shows a binding site conformational structure strikingly similar to that of the X-ray complex (binding site heavy atom RMSD of 0.9 Å), again highlighting the docking methods' capacity to select the correct receptor conformation within a large ensemble.

We now analyzed the docking performance as a function of the ensemble template's GI, using the combined data from all systems. The results presented in Figure 4 show that when GI is above 45%, the correct pose is found for all cases. For GI in the 30–45% range, the results are system and method dependent, but in several cases the correct pose can be detected. Also, as expected, biased methods perform slightly better than CADM.

Summary of monosaccharide docking

A summary of all the previously described results is presented for comparative and general analysis purpose in Figure 5. The figure shows the performance of all three docking methods using receptor structures from the holo, apo, MD frames, comparative models (the latter separated in either high or low template identity, using 45%

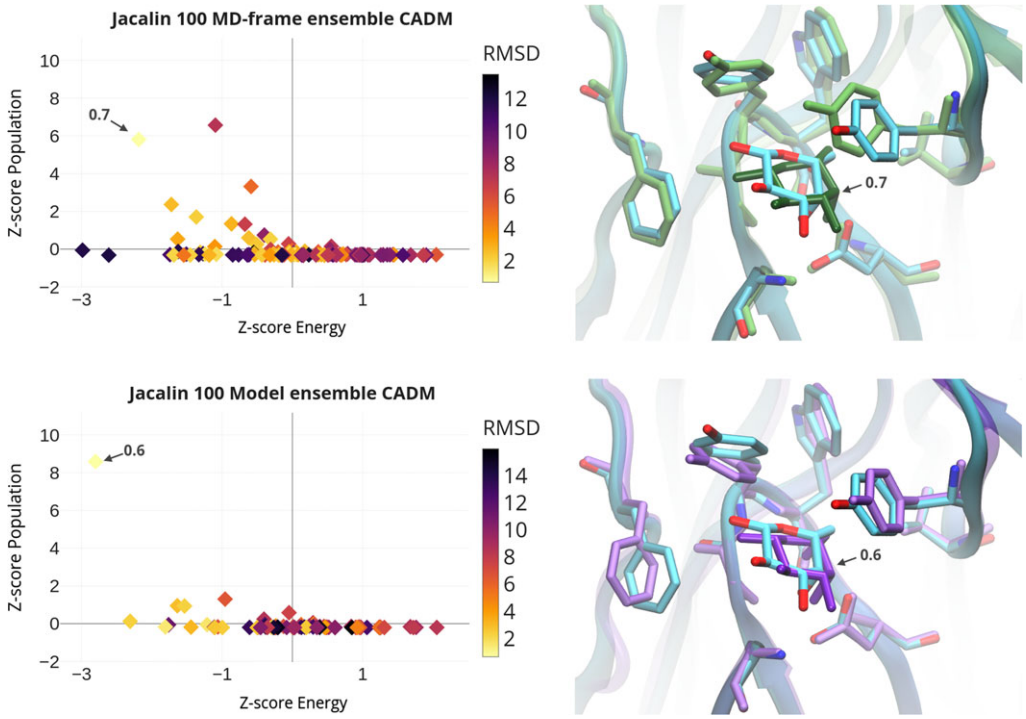

Fig. 3. "Z-score Population vs. Z-score binding energy" plot for the CADM docking of Beta D-Galactose to an ensemble of 100 different receptor structures of Jacalin. *Upper panel*: ensemble of 100 MD simulation snapshots. *Lower panel*: ensemble of 100 comparative models built using a template of 79% GI. X-ray Holo structure ligand pose and binding-site rotamers are always drawn in cyan; Outlier MD snapshot docking pose and binding-site residues are colored in green, outlier comparative model docking pose and binding-site residues are colored in purple.

Table III. Docking method comparison for different Jacalin ensembles

Jacalin - (beta-D-galactose)			Docking Method	First Rank RMSD	Best RMSD[c]	First Rank 2D-Score	Best RMSD 2D-Score[a]	Best RMSD Rank[b]
Jacalin (Jac)	Apo-MD 100 frame ensemble (starting strc. 1ku8-A)		CADM	0.69	0.69	7.994	7.994	1
			X-SSBDM	2.465	0.62	7.861	7.752	2
			MD-SSBDM	0.624	0.624	7.942	7.942	1
	Template 79% GI (1jot-A)	100 model ensemble	CADM	0.656	0.656	11.387	11.387	1
			X-SSBDM	0.599	0.599	8.287	8.287	1
			MD-SSBDM	0.64	0.64	11.184	11.184	1
	Template 59% GI (1xxq-A)	100 model ensemble	CADM	2.448	1.053	12.91	1.63	14
			X-SSBDM	2.549	1.067	9.916	5.8	2
			MD-SSBDM	2.439	0.979	13.743	3.953	2
	Template 34% GI (2jz4-A)	100 model ensemble	CADM	4.415	2.098	13.037	0.291	120
			X-SSBDM	4.142	1.925	12.174	0.343	69
			MD-SSBDM	4.436	1.696	12.808	-0.196	179

Comparison of the different docking methods performance on different receptor structure ensembles of lectin Jacalin.
[a]2D-score is the combined Population and Binding Energy normalized score (see methods). [b]Rank is the cluster ranking (ordered by 2D-score), and [c]RMSD is the sugar heavy atoms RMSD against the reference structure, after receptor alignment. [d]Docking performance on an MD generated receptor conformational ensemble.

GI as cut-off). Performance is compared first, by looking for each case at the pose displaying the lowest RMSD against the reference complex, and also the first ranked pose RMSD. Here we corroborate that in most cases correct ligand pose is found by docking, and as expected, most accurate results are found for the Holo receptor structures. MD frames and high GI models both perform similarly for Apo structures, underscoring the advantage of a multiple receptor structure strategy. Comparing the method specificity – understood as its capacity to detect the correct pose among the top-ranked ones – Figure 5 highlights the better performance of high GI model ensembles over single models and other types of ensembles, and also a slight improvement of the biased methods over plain Autodock.

Metal treatment in modeling and docking

Divalent cations often play major roles in the structure and function of biomolecules, that go from stabilizing structural motifs to even being directly involved in ligand binding (Gabius et al. 2011). For

Table IV. Overall comparison of ensemble docking performance

Correct result (rmsd < 1.5)	Method	Ensemble		
		100 Apo-MD snapshots	100 Models (>45% GI)	100 Models < 45% GI
First Ranked	CADM	F17a-G, bark, conA, jac (**4/14**)	jac, conA, bark, fbm (**4/12**)	fbm (**1/9**)
	X-SSBDM	F17a-G, bark (**2/14**)	jac, bark, fbm, sna-II (**4/12**)	fbm (**1/9**)
	MD-SSBDM	bark, jac, sna-II (**3/14**)	jac, conA, fbm (**3/12**)	-(**0/9**)
Within Top 5	CADM	psl, sna-II (**2/14**)	F17a-G, ctb, sp-D, hsi (**4/12**)	hsi, sp-D, bark (**3/9**)
	X-SSBDM	fbm, hpa, jac, psl, sna-II (**5/14**)	F17a-G, wga1, conA, ctb, psl (**5/12**)	hsi, hpa, psl (**3/9**)
	MD-SSBDM	F17a-G, conA, psl, sp-D, wga1 (**5/14**)	F17a-G, ctb, sna-II, sp-D, hsi (**5/12**)	fbm, gal7, hsi, sp-D, hpa, bark (**6/9**)
Outside Top 5	CADM	aia, ctb, fbm, gal7, hpa, hsi, sp-D, wga1 (**8/14**)	wga1, psl, sna-II (**3/12**)	gal7, psl, conA (**3/9**)
	X-SSBDM	aia, ctb, conA, gal7, hsi, sp-D, wga1 (**7/14**)	aia, sp-D, hsi (**3/12**)	gal7, sp-D, bark, conA (**4/9**)
	MD-SSBDM	aia, ctb, fbm, gal7, hpa, hsi (**6/14**)	aia, wga1, bark, psl (**4/12**)	psl, conA (**2/9**)
Not found	CADM	-(**0/9**)	aia (**1/12**)	hpa, jac (**2/9**)
	X-SSBDM	-(**0/9**)	-(**0/9**)	jac (**1/9**)
	MD-SSBDM	-(**0/9**)	-(**0/9**)	jac (**1/9**)

Comparison for each docking method for different structural ensembles for all 14 studied systems, showing for each method the name of successful cases in which the correct pose (rmsd < 1.5) was found for each rank category. Number in brackets indicate the fraction of successful cases. Models are further split into two categories according to the template Global Identity percentage (% GI).

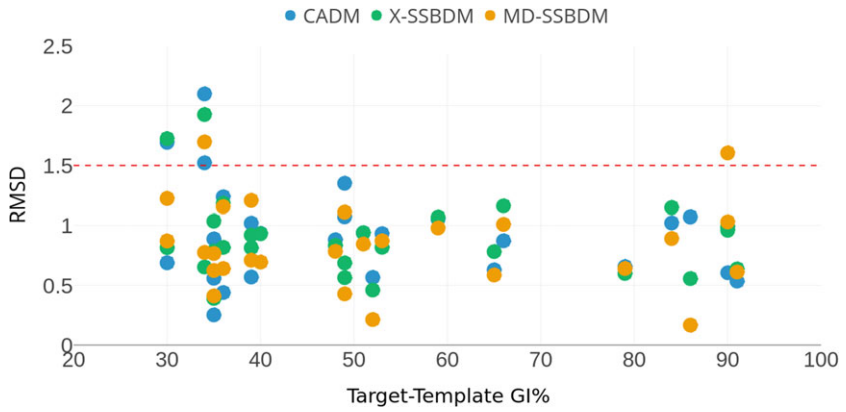

Fig. 4. RMSD of the best obtained pose vs. target-template % GI plot.

example, for lectins that adopt the so called "legume fold", it is known that they bind Ca2+ and Mn2+ in two well identified coordination sites whose proximity to the CBS make them of tremendous importance since they scaffold an important part of the structure and therefore, if not modeled correctly, could result in a poor quality structure for docking and/or modeling purposes. This is the case of *Griffonia simplicifolia* I lectin, a legume family lectin whose structure has been determined both in presence and absence of divalent cations. As can be seen in Figure 6 upper panel, both structures differ quite notably in the conformation of their binding site loop. In the metal-free form, residues involved in metal coordination (Asn 134 and Glu 139) become disordered, and Trp residue (W132) that in the holo structure is located very close to the ligand and helps to shape the CBS is now buried into the structure occupying the spot where the Calcium should be. An identical situation occurs with residue Y12 in another legume lectin, Concanavalin A (not shown), as evidenced comparing any holo structure with the metal-free structure PDB id: 1APN.

Another major concern arises when the cation participates directly in ligand binding through metal–oxygen interactions. This is the case of "C-type lectins", like the presently studied Surfactant Protein D (SP-D). Typically, this proteins bind the metal in a rather shallow site, leaving 2 valences of its coordination sphere free to interact with 2 water molecules (in apo-structures) or with the 2 adjacent hydroxyl groups of the carbohydrate ligand (in holo-structures), as depicted in Figure 6 bottom panel. In this context, a key consideration for successful Docking predictions is to have proper metal parameters to accurately model this type of interactions. For instance, Autodock4 does not assign any gasteiger electrostatic charge to ions by default, although experimented users can adjust this parameter. Nevertheless, since the key oxygen positions can be easily mapped by simply looking at apo structures water molecules and/or determining them from MD simulations, SSBDM performs quite well without the need of manipulating the cation charge. As shown in Table S7, the docking of glucose onto sp-D receptors containing metal with a zero charge successfully predicts the correct ligand pose across the variety of receptor types and ensembles.

Discussion

Appling of Molecular Docking methods for the prediction of lectin–carbohydrate complexes is challenging due to the hydrophilic and

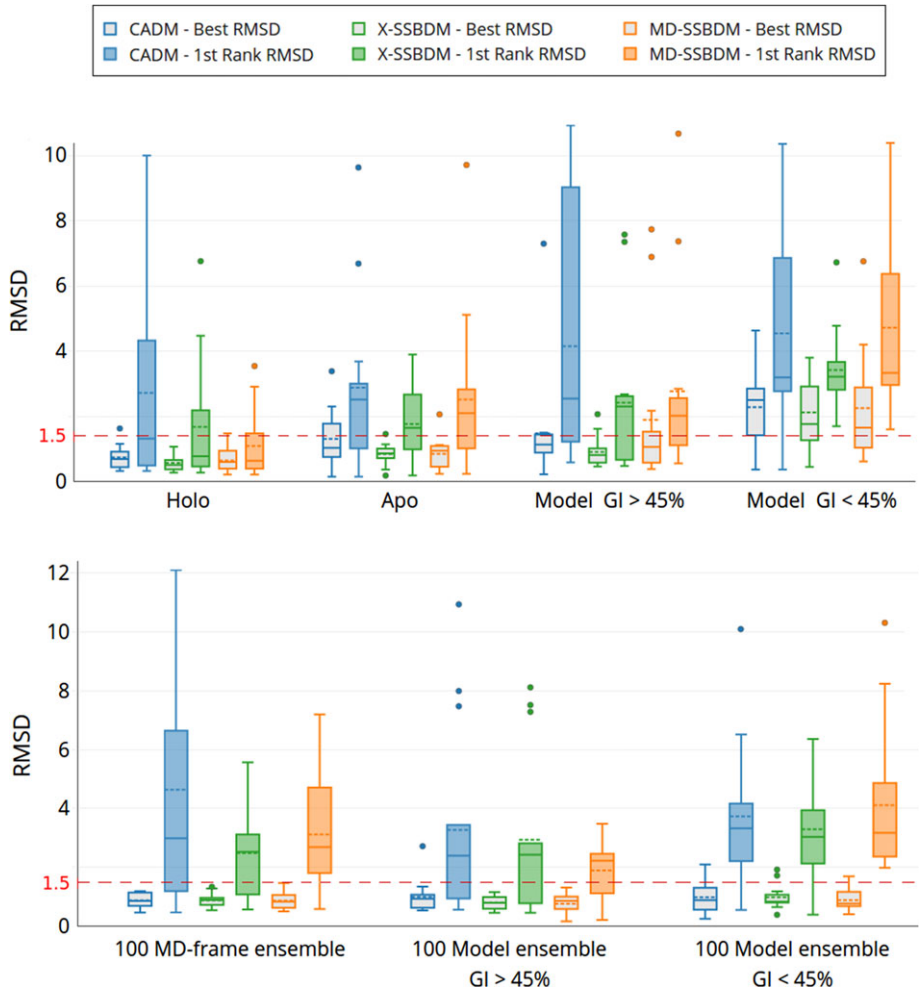

Fig. 5. Overall performance of the docking for the different types of receptor structures (*upper panel*) and receptor ensembles (*Lower panel*), shown as an RMSD Boxplot. "Best RMSD" (empty bars) refers to the pose that has the minimum RMSD value against the reference holo structure, whereas "first Rank RMSD" (full bar) refers to the pose with the highest 2D-score.

shallow nature of the binding site and the ligand conformational diversity. Previous works from our group (Gauto et al. 2013; Modenutti et al. 2015; Arcon et al. 2017), and others showed that inclusion of solvent structure information to bias (or modify) scoring functions to favor those ligand poses where carbohydrate –OH groups replace water sites can significantly improve the results. In the present work, we have extended the analysis of conventional and solvent site biased docking methods performance in the context of several ways that deal with receptor conformational diversity, focusing at the monosaccharide level as an initial approach.

The present results show that both conventional and solvent-site biased docking method allow obtaining accurate lectin–monosaccharide complexes starting from either apo or comparative modeling based receptor structures, being able to correctly deal with the receptor conformational flexibility to find the correct ligand-bound conformation within a large structural ensemble. When starting from apo structures, the receptor conformational ensemble can be built using relatively short MD simulations, such as those used to obtain the solvent sites. When starting just from the sequence, about a hundred models should be made, ideally from high % GI (>45) templates. Special care must be taken when modeling proteins that bind metals. Whether they are metals that do interact directly with the ligand as in C-type lectins, or metals that are important for the binding-site fold architecture as in legume lectins, only proper inclusion of these metals in the modeled receptor structures will provide accurate receptor structures and thus successful docking results.

When comparing the different ensembles regarding to their capability of producing "holo-like" structures, Figure 7 shows the binding-site heavy atom RMSD of each single structure with respect to the holo x-ray for Jacalin (other lectins are shown in Figure S1). As expected, the "holo-like" character of the structures diminishes as the % GI of the template goes down, and this is also reflected in the docking performance (see Table I). Also noteworthy, is that an ensemble of models built from high % GI templates can sometimes produce as good "holo-like" structures as those obtained from an MD ensemble built from an apo X-ray.

For the 14 cases studied in this work, a combination of conventional and biased docking always yields the correct complex among the top scored results, although not always the first. Moreover, performing multiple docking experiments to an ensemble of receptor conformations can be a clever strategy when no structure neither holo nor apo is available, although sometimes might result in a hard

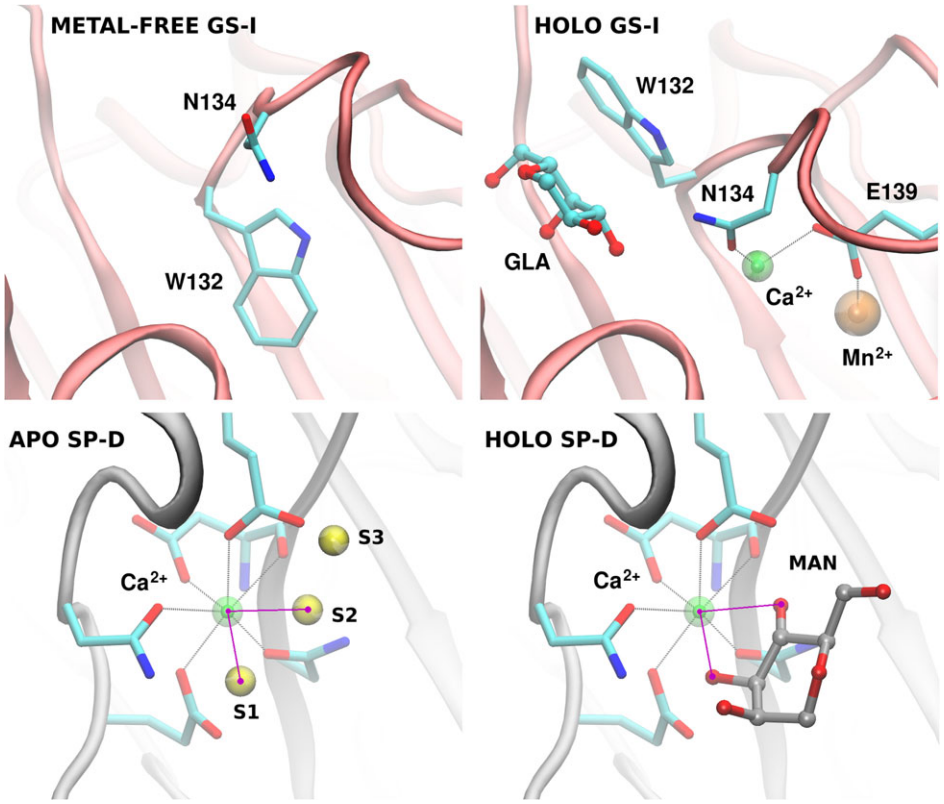

Fig. 6. *Upper panel*: comparison of metal-free (PDB id: 1gnz) and holo metal-bound (PDB id: 1hql) structures of *Griffonia simplicifolia* I lectin. *Lower panel*: SP-D Apo structure (PDB id: 1pw9) showing the X-ray water sites as yellow spheres, and Holo SP-D structure (PDB id: 3g81) showing the pose of ligand Mannose. Protein–metal interactions are drawn with black dashed lines, and water/ligand–metal interactions are drawn as magenta lines.

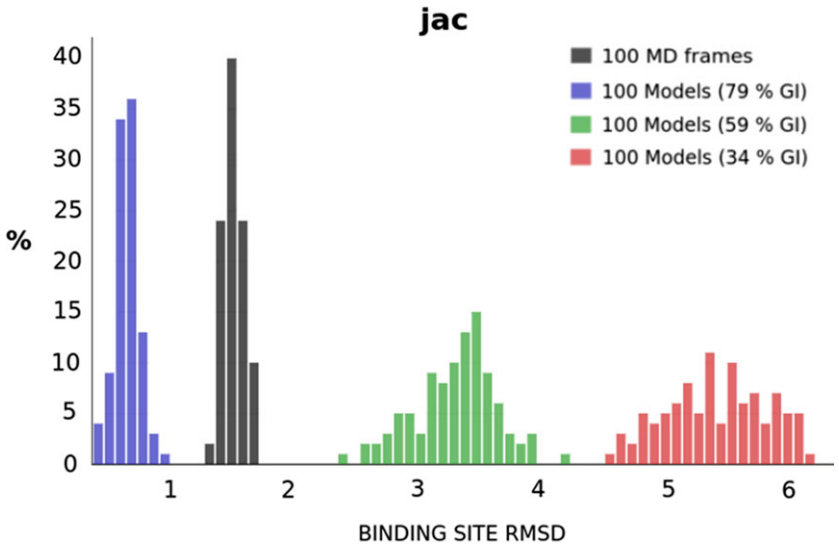

Fig. 7. Jacalin CBS RMSD distribution of the various 100-receptor ensembles against the reference Holo structure (PDB id: 1ugw). RMSD is computed just on the Binding site residues heavy atoms.

identification of the correct ligand-pose among top ranked false positives. This implies that, while the search algorithm combined with some MD or model building to allow receptor conformational diversity seems a fairly adequate approach, in order to achieve a more accurate ranking of the obtained poses better scoring functions are needed. The need for a scoring function improvement underscores our previous observation: in general, solvent-site biased docking improves performance, particularly concerning the method specificity. An example of a critical aspect of scoring function improvement was recently highlighted for carbohydrate ligands by

Table V. Lectin data set

Fold	Protein (acronym)	Holo struc. (ligand in 3 letter code)	Apo struc. selected for docking	Apo struc. used for X-SS determination	Templates for c. modeling (% GI/% BSI)
adhesin	Fimbrial Adhesin (F17a-G)	1o9w-A (NAG)	1o9z-A	1o9z-A	4k0o-A (91/100)
legume	Bark (bark)	1fnz-A (A2G)	1fny-A	1fny-A	4u36-A (65/57) 1gzc-A (49/43) 5t5l-A (36/43)
legume	Concanavalin A (conA)	5cna-B (MMA)	1jbc-A	1nls, 1jbc, 1scs, 1enr, 1qny, 2ctv, 1con, 1scr	3a0k-A (86/100) 1qmo-A (66/57) 1fx5-A (36/43)
ricin type beta trefoil	Fucose Binding Module (fbm)	2j1s-A (FUL)	2j1r-A	2j1r (chain A,B)	2j22-A (52/86) 3cqo-A (35/57)
jacalin-like	Jacalin (jac)	1ugw-A (GAL)	1ku8-A	1ku8 (chain A,B,C,G) 3p8s (chain A,B)	1jot-A (79/100) 1xxq-A (59/43) 2jz4-A (34/14)
legume	Pisum sativum lectin (psl)	1rin-AB (MAN)	2ltn-AB	2ltn (chain A,C)	4u36-A (53/71) 5t5l-A (39/57)
Galectin	Galectin 7 (gal7)	2gal-B (GAL)	3zxf-B	3zxf (chain A,B) 1bkz (chain A,B)	4lbj-A (35/100) 5duu-A (40/100) 2jj6-A (30/29)
ricin type beta trefoil	Toxin HA33/C (hsi)	3aj5-A (NAG)	1qxm-A	1qxm-A	5b2h-A (48/38) 2vse-A (30/38) 4lo0-A (39/13)
C-type	Surfactant Protein D (sp-D)	1pwb-B (GLC)	1pw9-A	1pw9 (chain A,B,C) 1b08 (chain A,B,C) 3dbz (chain A,B,C)	3pak-A (49/71) 4ymd-A (35/71)
ricin type beta trefoil	*S. nigra* agglutinin II - N term (sna-II)	3ca1-A (GAL)	3c9z-A	3c9z-A	2aai-B (51/71)
Jacalin-like	Artocarpin (aia)	1j4u-A (MMA)	1j4t-A	1j4s (chain A,B,C,D) 1j4t (chain A,B,C,D,E,F,G,H)	5krp-C (90/100)
H.L. enterotoxin	Cholera toxin (ctb)	1s5e-D (GAL)	1s5e-E	1s5e (chain E,F,G,H)	1b44-H (84/100)
H-type lectin	roman snail agglutinin (hpa)	2ccv-A (A2G)	2ce6-A	2ce6-A	3wmq-A (34/75)
chitin-recognition	Wheat Germ Agglutinin 1 (wga1)	2uvo-A (NAG)	7wga-A	1wgc (chain A,B) 2cwg (chain A,B) 7wga (chain A,B)	1k7t-A (90/80)

Data set of the 14 systems (lectin–monosaccharide complexes) used in this study, with all their respective PDB id codes. "GI" stands for "Global Identity" and "BSI %" stands for "Binding Site Identity".

Nivedha et al., who showed that re-scoring obtained poses with accurate ligand internal energy functions is crucial for successful docking beyond monosaccharides. This strategy has been successfully integrated into Autodock Vina scoring function and has proven to be a major improvement for carbohydrate docking, yet with still some remaining challenges to overcome (Nivedha et al. 2014; 2016). It would be interesting to explore how this method, called "Vina Carb", performs for oligosaccharides when combined with solvent structure biasing. A final remark on strategies for specificity improvement and false positive elimination is, as recently shown by Makeneni et. al. for a set of antibody-carbohydrate complexes, the use of a computational protocol consisting of "pose clustering" to reduce the number of unlikely poses combined with "postdocking MD simulations" for optimizing the ligand orientation and further eliminating incorrect poses (Makeneni et al. 2018).

Finally, it is interesting to note that the docking methods can select "holo-like" receptor structures, thus filtering bad quality receptor structures. Analysis of docking results for a receptor conformational ensemble – built without any knowledge of the ligand – allows selecting receptor structures which are conformationally adapted to the ligand and thus accurately resemble holo structures. Given the relatively small computational cost required to build these receptor conformational ensembles, this strategy looks promising for its broader application.

Computational methods

Lectin–carbohydrate complexes structural data set

The first analyzed set comprises the following 14 lectin–monosaccharide complexes structures, selected according to the following criteria: (1) they had to have at least one apo structure reported (resolution $\leq$ 2.5 Å), in order to be able to calculate X-ray derived Water Sites (see below); (2) they had to have at least one similar structure of an homologous protein reported, which could be used as a template for comparative modeling. Table V shows all the PDB id codes.

Molecular dynamics (MD) simulations

Starting from the crystal structure and/or homology based model, each system was first optimized using a conjugate gradient algorithm

for 5000 steps, followed by 150 ps long constant volume MD equilibration, in which the first 100 ps were used to gradually raise the temperature of the system from 0 to 300 K (integration step = 0.0005 ps/step). The heating was followed by a 250 ps long constant temperature and constant pressure MD simulation to equilibrate the system density (integration step = 0.001 ps/step). During these temperature and density equilibration processes, the protein alpha-carbon atoms were constrained by 5 kcal/mol/Å force constant using a harmonic potential centered at each atom starting position. Next, a second equilibration MD of 500 ps. was performed, in which the integration step was increased to 2 fs and the force constant for restrained alpha-carbons was decreased to 2 kcal/mol/Å. Finally, a 1 ns. long MD simulation was carried out with no constraints and the "Hydrogen Mass Repartition" technique (Hopkins et al. 2015), which allows an integration step of 4 fs, and this conditions were kept for all the subsequent Production 20 ns long MD runs. This protocol was replicated 3 times, starting always from the same initial structure but assigning different random initial velocities to the atoms, thus generating three independent 20 ns trajectories (with a sample rate of 500 frames per nanosecond) for each case.

All simulations were performed with the Amber package of programs (Case et al. 2014) using the ff14SB force field (Maier et al. 2015) for all amino acid residues. Pressure and temperature were kept constant using the Monte-Carlo barostat and Langevin thermostat, respectively, using the default coupling parameters. All simulations were performed with a 10 Å cut-off for non-bonded interactions, and periodic boundary conditions using the Particle Mesh Ewald summation method for long-range electrostatic interactions. The SHAKE algorithm was applied to all hydrogen-containing bonds in all simulations with an integration step equal or higher than 2 fs. Trajectory processing was performed with the Cpptraj module of the AMBER package (Roe and Cheatham 2013).

Determination of MD and X-ray derived Water Sites

The Water Sites calculation is based on the method developed and thoroughly tested in our previous works (Gauto et al. 2009; Modenutti et al. 2015). Briefly, Molecular Dynamics derived Water Sites (MD-WS) are derived – as their name describe – from explicit water MD simulations (Gauto et al. 2009). The position of the oxygen atoms of water molecules in successive snapshots are clustered together, when they are closer than a user defined cut-off (usually 1.4 Å), and their center of mass defined the corresponding WS coordinates. Subsequently, the probability of finding a water molecule inside a volume of 1 Å centered on the WS coordinates is computed with respect to the bulk solvent. When this value is larger than 2, the WS is kept for the biased docking.

X-ray derived Water Sites (X-WS) are defined simply by the presence of crystallographic waters in the available apo structure/s of the corresponding receptor. To define an X-WS across several structures, if any crystallographic waters from different structures are closer than 1.4 Å, they are combined and then, X-WS position is then defined by the center of mass of all resulting oxygen atoms that form it (Modenutti et al. 2015).

Comparative modeling

All homology models were generated using the Modeller v.9.19 software (Eswar et al. 2008). Template search was done using the HHpred tool of the Max Planck Institute Bioinformatics Toolkit (Alva et al. 2016). Whenever possible, three templates with three different levels of global identity percentage (% GI) were selected for each system (a high, an intermediate and a low % GI). Finally, in such cases where % GI was the same for more than one possible template, the template was selected randomly.

Docking protocol

The "Conventional Docking Method" consists in using Autodock 4.2.6 software, with all its parameters kept default. Briefly, Autodock works by computing energy grids for each ligand atom type, based on the receptor structure. The grid size and position were chosen so that they include the whole CBS. This was achieved by placing the grid center in the geometric center of the CBS, and extending its size 20 Å (for the mono- and disaccharide-binding proteins) in each direction, and using a grid spacing of 0.375 Å. The genetic algorithm parameters for each conformational search run were kept at their default values (150 for initial population size, 2.5×10^6 as the maximum number of energy evaluations and 2.7×10^4 as the maximum number of generations). For each calculation, always 100 different docking runs were performed, and the resulting 100 poses were clustered according to the pyranose ring heavy atom RMSD using a cut-off of 1.5 Å, using the quality threshold algorithm implemented in VMD software.

The Water Site Biased Docking Method (WSBDM) is based on the method developed and thoroughly tested in our previous works (Gauto et al. 2013; Modenutti et al. 2015). For both variants of the method, (MD/X-SSBMD) the strategy is conceptually the same: they take advantage of the fact that carbohydrate –OH groups tend to occupy or replace the positions of the WS, and that there is a positive correlation between the solvent site's probability of being replaced, and a negative correlation with its dispersion. The AutoDock4 scoring function is thereby modified, adding an additional energy term for each ligand oxygen (atom type "OA") to the original function, whose deepness and size is proportional to its probability and dispersion, respectively.

All ligands coordinates were retrieved from the 'holo structure complexes, and randomized before each docking calculation. All ligand torsions were kept active.

Data analysis

To analyze the performance of the different docking strategies we considered two issues: first, the method's accuracy, which is measured as the pyranose ring heavy atoms (C1, C2, C3, C4, C5 and O5) RMSD of each predicted complex with respect to the position of the same group of atoms in the corresponding complex crystal structure. Except for the re-docking experiments (i.e using the holo receptor structure), for all other cases the receptor conformations were structurally aligned to the reference complex structure considering only the receptor binding site heavy atoms. In summary, accuracy measures the docking method's capacity to place the ligand in the correct place and with the correct orientation. Secondly, we analyzed the method's specificity, understood as its capacity to distinguish the right complex from all other wrong predictions. Each Autodock4 run yields the estimated ligand binding energy for the corresponding pose (ΔE). Then the population, which is the percentage of individual docking poses that resulted in a similar binding mode according to a RMSD criteria, can be easily computed after pose clustering.

Previous works showed that both parameters should be taken into account in order to have a reliable prediction. This is done by plotting population vs. binding energy plots for all obtained poses

and looking for outliers in the upper left corner (those with high population and negative ΔGB) (change in free energy upon binding, see results in previous works for examples). Since in the present work we are comparing results across a large number of receptor conformations, for each receptor we combined all docking results. In order to better compare different docking methods, we used a receptor-dependent normalization for both parameters, which is simply the corresponding Z-score. Therefore in the present work we present the results as Z-score(ΔE) vs. Z-score(Pop) plots.

Finally, to rank all obtained poses combining both parameters we designed a "combined Z-score", which we call 2D-score (scheme I) and is defined as follows:

$$2D\text{−score} = -1 * [Z\text{−score}(\Delta E)] + Z\text{−score}(Pop)$$

In other words, top ranking poses correspond to those having negative ΔE and high population (upper left quadrant).

Supplementary data

Supplementary data is available at *GLYCOBIOLOGY* online.

Funding

This work was supported by grant PIP 11220130100469CO awarded to M. A.M. The authors would like to thank Centro de Cómputos de Alto Rendimiento (CeCAR), FCEN-UBA for granting use of computational resources which allowed us to perform the experiments included in this work. M.D., J.I.B.C. and M.G. are CONICET doctoral fellows. C.M. is CONICET postdoctoral fellow. M.A.M. is member of the CONICET.

Abbreviations

CADM, Conventional Autodock Docking Method; GI, Global Identity; MD, Molecular Dynamics; RMSD, Root Mean Square Deviation; SSBDM, Solvent site Biased Docking Method; WS, Water Site

References

Abel R, Young T, Farid R, Berne BJ, Friesner RA. 2008. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J Am Chem Soc*. 130(9):2817–2831.

Alva V, Nam S-Z, Söding J, Lupas AN. 2016. The MPI bioinformatics toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res*. 44(W1):W410–W415.

Arcon JP, Defelipe LA, Modenutti CP, López ED, Alvarez-Garcia D, Barril X, Turjanski AG, Martí MA. 2017. Molecular dynamics in mixed solvents reveals protein–ligand interactions, improves docking, and allows accurate binding free energy predictions. *J Chem Inf Model*. 57(4):846–863, American Chemical Society.

Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, Cheatham TE, III, Darden TA, Duke RE, Gohlke H et al. 2014. AMBER 14, University of California, San Francisco.

Eswar N, Eramian D, Webb B, Shen M-Y, Sali A. 2008. Protein structure modeling with MODELLER. *Methods Mol Biol*. 426:145–159.

Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. 2016. Computational protein-ligand docking and virtual drug screening with the AutoDock Suite. *Nat Protoc*. 11(5):905–919.

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK et al. 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*. 47(7):1739–1749.

Gabius H-J, Andre S, Jimenez-Barbero J, Romero A, Solís D. 2011. From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem Sci*. 36(6):298–313.

Gauto DF, Di Lella S, Estrin DA, Monaco HL, Martí MA. 2011. Structural basis for ligand recognition in a mushroom lectin: solvent structure as specificity predictor. *Carbohydr Res*. 346(7):939–948.

Gauto DF, Di Lella S, Guardia CMA, Estrin DA, Martí MA. 2009. Carbohydrate-binding proteins: dissecting ligand structures through solvent environment occupancy. *J Phys Chem B*. 113(25): 8717–8724.

Gauto DF, Petruk AA, Modenutti CP, Blanco JI, Di Lella S, Martí MA. 2013. Solvent structure improves docking prediction in lectin-carbohydrate complexes. *Glycobiology*. 23(2):241–258.

Grant OC, Tessier MB, Meche L, Mahal LK, Foley BL, Woods RJ. 2016. Combining 3D structure with glycan array data provides insight into the origin of glycan specificity. *Glycobiology*. 26(7):772–783.

Guardia CMA, Gauto DF, Di Lella S, Rabinovich GA, Martí MA, Estrin DA. 2011. An integrated computational analysis of the structure, dynamics, and ligand binding interactions of the human galectin network. *J Chem Inf Model*. 51(8):1918–1930.

Hopkins CW, Le Grand S, Walker RC, Roitberg AE. 2015. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J Chem Theory Comput*. 11(4):1864–1874, American Chemical Society.

Kadirvelraj R, Foley BL, Dyekjaer JD, Woods RJ. 2008. Involvement of water in carbohydrate-protein binding: Concanavalin a revisited. *J Am Chem Soc*. 130(50):16933–16942.

Kerzmann A, Fuhrmann J, Kohlbacher O, Neumann D. 2008. BALLDock/SLICK: a new method for protein-carbohydrate docking. *J Chem Inf Model*. 48(8):1616–1625.

López ED, Arcon JP, Gauto DF, Petruk AA, Modenutti CP, Dumas VG, Marti MA, Turjanski AG. 2015. WATCLUST: a tool for improving the design of drugs based on protein-water interactions. *Bioinformatics*. 31 (22):3697–3699.

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. 2015. ff14SB: improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*. 11(8): 3696–3713.

Makeneni S, Thieker DF, Woods RJ. 2018. Applying pose clustering and MD simulations to eliminate false positives in molecular docking. *J Chem Inf Model*. 58:605–614.

Modenutti C, Gauto D, Radusky L, Blanco J, Turjanski A, Hajos S, Marti M. 2015. Using crystallographic water properties for the analysis and prediction of lectin-carbohydrate complex structures. *Glycobiology*. 25(2): 181–196.

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ, Others. 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem*. 19 (14):1639–1662.

Morris GM, Goodsell DS, Huey R, Olson AJ. 1996. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J Comput Aided Mol Des*. 10(4):293–304.

Nivedha AK, Makeneni S, Foley BL, Tessier MB, Woods RJ. 2014. Importance of ligand conformational energies in carbohydrate docking: sorting the wheat from the chaff. *J Comput Chem*. 35(7):526–539.

Nivedha AK, Thieker DF, Makeneni S, Hu H, Woods RJ. 2016. Vina-Carb: improving glycosidic angles during carbohydrate docking. *J Chem Theory Comput*. 12(2):892–901.

Orozco CA, Martinez-Bosch N, Guerrero PE, Vinaixa J, Dalotto-Moreno T, Iglesias M, Moreno M, Djurec M, Poirier F, Gabius HJ et al. 2018. Targeting Galectin-1 inhibits pancreatic cancer progression by modulating tumor–stroma crosstalk. *Proc Natl Acad Sci USA*. 115(16):E3769–E3778, http://www.pnas.org/content/early/2018/04/02/1722434115.short.

Roe DR, Cheatham TE 3rd. 2013. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput*. 9(7):3084–3095.

Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, Barril X, Hubbard RE, David Morley S. 2014. rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput Biol*. 10(4):e1003571.

Saraboji K, Håkansson M, Genheden S, Diehl C, Qvist J, Weininger U, Nilsson UJ, Leffler H, Ryde U, Akke M et al. 2012. The carbohydrate-binding site in Galectin-3 is preorganized to recognize a sugarlike framework of oxygens: ultra-high-resolution structures and water dynamics. *Biochemistry*. 51(1):296–306.

Trott O, Olson AJ. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 31(2):455–461.

Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. 2003. Improved protein-ligand docking using GOLD. *Proteins*. 52(4): 609–623.