



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Evaluation of Selected Classical Force Fields for Alchemical Binding Free Energy Calculations of Protein-Carbohydrate Complexes

**Citation for published version:**

Mishra, SK, Calabró, G, Loeffler, HH, Michel, J & Koča, J 2015, 'Evaluation of Selected Classical Force Fields for Alchemical Binding Free Energy Calculations of Protein-Carbohydrate Complexes', *Journal of Chemical Theory and Computation*, vol. 11, no. 7, pp. 3333-45. <https://doi.org/10.1021/acs.jctc.5b00159>

**Digital Object Identifier (DOI):**

[10.1021/acs.jctc.5b00159](https://doi.org/10.1021/acs.jctc.5b00159)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Journal of Chemical Theory and Computation

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

## Evaluation of Selected Classical Force Fields for Alchemical Binding Free Energy Calculations of Protein-Carbohydrate complexes

Journal:	<i>Journal of Chemical Theory and Computation</i>
Manuscript ID:	ct-2015-00159u.R2
Manuscript Type:	Article
Date Submitted by the Author:	01-Jun-2015
Complete List of Authors:	Mishra, Sushil; Central European Institute of Technology (CEITEC), Masaryk University Calabró, Gaetano; University of Edinburgh, EastCHEM School of Chemistry Loeffler, Hannes; STFC Daresbury, Scientific Computing Department Michel, Julien; University of Edinburgh, EastCHEM School of Chemistry Koča, Jaroslav; Central European Institute of Technology (CEITEC), Masaryk University

SCHOLARONE™  
Manuscripts

# Evaluation of Selected Classical Force Fields for Alchemical Binding Free Energy Calculations of Protein-Carbohydrate complexes

*Sushil K. Mishra,<sup>†</sup> Gaetano Calabró,<sup>§</sup> Hannes H. Loeffler,<sup>‡</sup> Julien Michel,<sup>§\*</sup> Jaroslav Koča*

*<sup>†\*</sup>*

<sup>†</sup> Central European Institute of Technology (CEITEC), and National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice-5, 625 00 Brno, Czech Republic. <sup>§</sup> EaStCHEM School of Chemistry, Joseph Black Building, King's Buildings, Edinburgh EH9 3JJ, UK. <sup>‡</sup> Scientific Computing Department, STFC Daresbury, Warrington, WA4 4AD, UK.

**Keywords:** protein-carbohydrate, lectins, free-energy, thermodynamics, molecular recognition, GLYCAM, thermodynamic integration.

**Abstract:** Protein-carbohydrate recognition is crucial in many vital biological processes including host-pathogen recognition, cell-signaling, and catalysis. Accordingly, computational prediction of protein-carbohydrate binding free-energies is of enormous interest for drug design. However, the accuracy of current force fields (FFs) for predicting binding free energies of protein-carbohydrate complexes is not well understood owing to technical challenges such as the highly polar nature of the complexes, anomerization and conformational flexibility of carbohydrates. The present study evaluated the performance of

1  
2  
3 alchemical predictions of binding free-energies with the GAFF1.7/AM1-BCC and  
4  
5 GLYCAM06j force fields for modelling protein-carbohydrate complexes. Mean unsigned  
6  
7 errors of  $1.1\pm 0.06$  (GLYCAM06j) and  $2.6\pm 0.08$  (GAFF1.7/AM1-BCC)  $\text{kcal}\cdot\text{mol}^{-1}$  are  
8  
9 achieved for a large dataset of monosaccharide ligands for *Ralstonia solanacearum* lectin  
10  
11 (RSL). The level of accuracy provided by GLYCAM06j is sufficient to discriminate potent,  
12  
13 moderate and weak binders, a goal that has been difficult to achieve through other scoring  
14  
15 approaches. Accordingly the protocols presented here could find useful applications in  
16  
17 carbohydrate-based drug and vaccine developments.  
18  
19

## 20 21 22 **1. Introduction**

23  
24 The problem of computing the binding free-energy of a ligand for a receptor is a long-  
25  
26 standing challenge for computational chemistry.<sup>1-3</sup> Ever since the very first alchemical free-  
27  
28 energy (AFE) calculations were reported for ligand binding processes,<sup>4</sup> numerous studies  
29  
30 have focused on the binding energetics of organic molecules to proteins.<sup>3,5,6</sup> Despite  
31  
32 successes in guiding the design of organic molecules as protein ligands<sup>1,7-9</sup> applications to  
33  
34 other classes of biomolecular interactions such as protein-DNA, protein-lipids and protein-  
35  
36 carbohydrate complexes have been less explored.<sup>2,10,11</sup> This work is concerned with the  
37  
38 validation of parameter sets for accurate modelling of protein-carbohydrate recognition with  
39  
40 the aid of alchemical free-energy methods.  
41  
42  
43  
44

45 Protein-carbohydrate complexes pose specific challenges for molecular modeling due to the  
46  
47 large number of hydroxyl groups in the ligands, weak binding affinities, anomerization, ring  
48  
49 flexibility, CH... $\pi$  interactions, and frequent role of water<sup>12</sup> and/or ions<sup>13</sup> in receptor binding  
50  
51 sites. Progress is necessary owing to the significant role of protein-carbohydrate interactions  
52  
53 in biology. A few notable examples includes biological processes like, cell adhesion,  
54  
55 differentiation, and metastasis.<sup>14,15</sup> Protein-carbohydrate interactions are also important in  
56  
57  
58  
59  
60

1  
2  
3 medical sciences, e.g. alterations of cell surface glycosylation pattern is linked to the  
4  
5 development and progression of specific diseases like cancer.<sup>16</sup> Among protein-carbohydrate  
6  
7 complexes, lectin-carbohydrate complexes are of immense interest because lectins have the  
8  
9 ability to distinguish between minuscule differences in sugar structures and can be used to  
10  
11 detect specific carbohydrate patterns.<sup>17</sup> Thus, understanding the structural and energetic  
12  
13 aspects of the lectin-carbohydrate complexes is essential for the elucidation of carbohydrate  
14  
15 recognition principles, which should ultimately aid the design of carbohydrate-based  
16  
17 pharmaceuticals.<sup>14,18</sup>  
18  
19

20  
21 Current docking programs and empirical scoring functions do not generally provide an  
22  
23 accurate description of protein-carbohydrate binding energetics.<sup>19-23</sup> Efforts to tackle the  
24  
25 problem with end-point free-energy methods such as Molecular Mechanics Poisson-  
26  
27 Boltzmann Surface Area (MM-PB/SA), Molecular Mechanics Generalized-Born Surface  
28  
29 Area (MM-GB/SA) or Linear Interaction Energy (LIE) have also been reported with mixed  
30  
31 success.<sup>24-27</sup> Mishra et al. have parameterized the LIE approach directly on carbohydrates but  
32  
33 found significant overestimations of the calculated binding free-energies for low-affinity  
34  
35 binders and non-binders.<sup>27</sup> Topin et al. have shown that MM-GB/SA method yields a poor  
36  
37 correlation between the predicted and experimentally determined free-energies for lectins  
38  
39 LecB ( $r^2=0.22$ ) and BambL ( $r^2=0.02$ ).<sup>28</sup> Moreover, outliers are frequently seen in MM-  
40  
41 PB/GBSA calculation studies of protein-carbohydrate complexes.<sup>24,28</sup>  
42  
43  
44

45 AFE calculation protocols (e.g. free energy perturbation (FEP), thermodynamic integration  
46  
47 (TI)) are attractive alternatives owing to a more rigorous statistical thermodynamics  
48  
49 framework.<sup>3,29</sup> However, the reliability of current force fields (FFs) for AFE calculations of  
50  
51 protein-carbohydrate complexes is still unclear. The carbohydrate FFs that can be used for  
52  
53 simulation in the biomolecular context are mainly CHARMM<sup>30</sup>, GROMOS-45A4<sup>31</sup>, OPLS-  
54  
55 AA-SIE<sup>32</sup> and GLYCAM<sup>33</sup>. Among them GLYCAM is steadily growing and the most cited  
56  
57  
58  
59  
60

1  
2  
3 FF due to their generalized parameterization scheme that can be readily extended to  
4  
5 oligosaccharides.<sup>25</sup> Because the derivation of carbohydrate parameters is laborious, and thus  
6  
7 parameters cannot be easily derived for non-natural carbohydrate based ligands. However,  
8  
9 generic force fields such as the General Amber Force Field<sup>34,35</sup> with AM1-BCC charges<sup>36,37</sup>  
10  
11 (GAFF/AM1-BCC) can possibly provide a faster route to carbohydrate simulation.  
12  
13 GAFF/AM1-BCC offers a simple framework for rapid parameterization of small organic  
14  
15 molecules including carbohydrate derived scaffolds. Indeed GAFF has been used  
16  
17 occasionally for modeling of carbohydrate or their derivatives.<sup>38-41</sup> This is important to  
18  
19 support computer-aided design of functionalized carbohydrates, carbohydrate hybrid drugs  
20  
21 and glycomimetic drugs,<sup>42</sup> notable examples include Miglitol (Glyset),<sup>43</sup> Voglibose  
22  
23 (Glustat),<sup>44</sup> or Miglustat (Zavesca).<sup>45</sup> By contrast, specialized force fields such as  
24  
25 GLYCAM<sup>33</sup> focus on accurate carbohydrate modelling, at the expense of a smaller range of  
26  
27 parameter sets. This makes it more difficult to apply GLYCAM to a broad range of  
28  
29 carbohydrate-based ligand design problems. To set the scene for AFE-guided carbohydrate  
30  
31 ligands design it is thus crucial to establish whether the deficiencies of GLYCAM related to  
32  
33 its limited domain of applicability is compensated by an improved accuracy in predictions of  
34  
35 binding energetics. To this end 30 AFE calculations were performed on a dataset of 9  
36  
37 monosaccharides ligands of lectin RSL with both force fields. This dataset is larger than  
38  
39 those used in preceding protein-carbohydrate binding free-energy studies,<sup>12,46,47,41,48</sup> and  
40  
41 includes a wide range of monosaccharides ranging from high-affinity binders, low-affinity  
42  
43 binders to non-binders.  
44  
45  
46  
47  
48

## 49 **2. Theory and Methods**

### 50 **2.1 Preparation of Molecular Models**

51  
52 *Protein setup:* *Ralstonia solanacearum* lectin (RSL<sup>49</sup>) is a protein isolated from the Gram-  
53  
54 negative bacterial pathogen *Ralstonia solanacearum* that causes lethal wilt disease in many  
55  
56  
57  
58  
59  
60

1  
2  
3 agricultural crops all over the world, leading to massive losses in the agricultural industry.<sup>50</sup>  
4  
5 RSL is a six-bladed  $\beta$ - propeller trimeric structure, with 90 amino acid residues in each  
6  
7 monomeric chain. Each RSL monomer unit contains one fucose binding site located between  
8  
9 the two  $\beta$ -sheet blades called the intramonomeric binding site, and the other is formed at the  
10  
11 interface between the neighboring monomers called the intermonomeric binding site. Thus,  
12  
13 there are a total of six symmetrically arranged binding sites reported in the crystal structure.  
14  
15 Isothermal Titration Calorimetry (ITC) has suggested that intermonomeric and  
16  
17 intramonomeric binding sites are indistinguishable.<sup>49</sup> The calculated free-energy of Me- $\alpha$ -L-  
18  
19 Fucoside (MeFuc) in all six binding sites was statistically equivalent in LIE calculations  
20  
21 reported elsewhere.<sup>27</sup> The initial coordinates of RSL bound to MeFuc were obtained from the  
22  
23 X-ray crystal structure (PDB ID: 2BT9).<sup>49</sup> A couple of perturbations (**1** $\leftrightarrow$ **2** and **1** $\leftrightarrow$ **4**) were  
24  
25 performed in all six RSL binding sites (Table S1). Refer to Fig. 1 for number assigned to  
26  
27 each ligand. The other perturbations were performed in the intramonomeric binding site of  
28  
29 chain A (**site S1**) only, with the other five binding sites kept empty.  
30  
31  
32

33  
34 **Protein-Ligand complex setup:** A full range of monosaccharides spanning from binders,  
35  
36 low-affinity binders to non-binders were selected (Fig. 1). The experimental dissociation  
37  
38 constants of these monosaccharides have been previously measured using an SPR assay.<sup>23,49</sup>  
39  
40 The 3D coordinates of the ligands were modeled using the GLYCAM Carbohydrate Builder  
41  
42 webserver.<sup>51</sup> Since there is no evidence that RSL recognizes monosaccharides in furanose  
43  
44 form, pyranose form of all the ligands were selected. The starting structures of the RSL-  
45  
46 saccharide complexes for the other monosaccharides were prepared manually by  
47  
48 superposition with the ring atoms of MeFuc (**1**), keeping the orientation of O2, O3, and O4  
49  
50 hydroxyls unchanged where possible. As the potential energy barrier for ring flips in  
51  
52 pyranoses can be ca. 3-5 kcal $\cdot$ mol<sup>-1</sup> or greater,<sup>52,53</sup> all monosaccharides were kept in their  
53  
54 most favorable chair conformation. Input files for alchemical free-energy calculations were  
55  
56  
57  
58  
59  
60

1  
2  
3 prepared with the software FESetup<sup>54</sup> that uses AmberTools14<sup>55</sup> for ligand parameterization.  
4  
5 All initial structures were solvated in a rectangular box of TIP3P water molecules extending  
6  
7 12 Å away from the edges of the solute(s) using *tLeap*. The total charge of each system was  
8  
9 zero and no ions were required to neutralize any of the systems. The protein was described  
10  
11 with the Amber ff12SB<sup>56</sup> force field, and saccharides with either the GAFF<sup>34,35</sup> force field  
12  
13 (version 1.7) with AM1-BCC<sup>36,37</sup> charges (as computed by antechamber from  
14  
15 AmberTools14) or the GLYCAM<sup>33</sup> force field version 06j (GLYCAM06j). The simulation  
16  
17 systems were equilibrated by firstly performing 3000 steps of energy minimization to relax  
18  
19 unfavorable conformations, followed by a 1 ns 300K NVT simulation with harmonic  
20  
21 positional restraints of force constant 5 kcal•mol<sup>-1</sup>•Å<sup>-2</sup> on all the non-solvent atoms. Finally,  
22  
23 an unrestrained 3 ns NPT simulation was performed to equilibrate solvent density. The final  
24  
25 snapshot was used as input for subsequent free-energy calculations.  
26  
27  
28

## 29 30 2.2 Alchemical free-energy simulations

31  
32 Neglecting contribution from changes in pressure-volume terms<sup>57</sup>, relative binding free-  
33  
34 energies (eq 1) were computed as the difference in the free-energy changes for transforming  
35  
36 monosaccharide X into Y in the RSL binding site ( $\Delta G_P(X \rightarrow Y)$ ) and in aqueous  
37  
38 solution( $\Delta G_W(X \rightarrow Y)$ ):  
39

$$40 \quad \Delta \Delta G_b(X \rightarrow Y) = \Delta G_P(X \rightarrow Y) - \Delta G_W(X \rightarrow Y) \quad (1)$$

41  
42 Where each free-energy change was obtained by thermodynamic integration (TI):  
43

$$44 \quad \Delta G_P(X \rightarrow Y) = \int_{\lambda=0}^{\lambda=1} \frac{\partial G}{\partial \lambda} d\lambda \quad (2)$$

45  
46 Where  $\lambda$  is a coupling parameter that allows smooth transformation of the potential energy  
47  
48 function corresponding to the starting state X ( $\lambda=0$ ) and final state Y ( $\lambda=1$ ). The finite  
49  
50 difference thermodynamic integration approach was firstly used to evaluate free-energy  
51  
52 gradients at several values of  $\lambda$  between 0 and 1.<sup>58</sup> The integral in eq 2 was then numerically  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 approximated by using polynomial regression<sup>59</sup> and setting the polynomial order to seven.  
4  
5 Unless stated separately in the text, the free-energy gradients were calculated at 16 non-  
6  
7 equidistant  $\lambda$  values (0.0, 0.00616, 0.02447, 0.07368, 0.11980, 0.19045, 0.28534, 0.40631,  
8  
9 0.57822, 0.70755, 0.80955, 0.88020, 0.92632, 0.97553, 0.99384, and 1.0).<sup>59</sup> A 4 ns NPT  
10  
11 simulation at each  $\lambda$  value was performed to collect free-energy gradients. To test for  
12  
13 convergence, longer simulations of 10 ns per window were run, or a  $\lambda$  schedule with 27  
14  
15 points was applied in selected cases. Additional points were added near noisy parts of the  
16  
17 gradient when needed, e.g. **1**→**3** and **2**→**3** perturbations. Before collecting free-energy  
18  
19 gradients, the systems were further energy minimized (1000 steps) and then equilibrated for  
20  
21 100 ps at the chosen value of  $\lambda$ . To avoid numerical instabilities, a soft-core<sup>60</sup> potential  
22  
23 energy function similar to Michel et al. was used throughout.<sup>61</sup> Free-energy gradients were  
24  
25 collected every 200 fs. The first 500 ps of every simulation was discarded to allow for re-  
26  
27 equilibration.  
28  
29  
30

31  
32 A velocity-Verlet integrator was used with a timestep of 2 fs. All simulations were  
33  
34 performed in the NPT ensemble. Temperature control was achieved with an Andersen  
35  
36 thermostat and a coupling constant of 10 ps<sup>-1</sup>.<sup>62</sup> Pressure control was achieved by attempting  
37  
38 isotropic box edge scaling Monte Carlo moves every 25 time-steps. Periodic boundary  
39  
40 conditions are used with a 10Å atom-based cutoff distance for the non-bonded interactions.  
41  
42 All simulations were performed using an atom-based cutoff of 10 Å with Barker Watts  
43  
44 reaction field with dielectric constant set to 78.3.<sup>63</sup> The methodology used here to handle long  
45  
46 range electrostatic interactions differs from the parameters used in other studies performed  
47  
48 with the Amber ff12SB force field. However, this was deemed acceptable as Fennel and  
49  
50 Gezelter have reported that atom-based reaction-field treatments yield energetics and  
51  
52 dynamics that reproduce well Particle-Mesh Ewald.<sup>64</sup> Further, in this work the same  
53  
54 methodology was used consistently to compare GAFF and GLYCAM results. Production  
55  
56  
57  
58  
59  
60

molecular dynamics simulations were performed on GPUs (GeForce GTX465 and Tesla M2090/K20 cards) using mixed precision in the Sire/OpenMM simulation framework (rev. 2702 of Sire<sup>65</sup> and OpenMM-6.0<sup>66</sup>). To test convergence and reproducibility, free-energy changes along both ‘direct’ ( $\Delta\Delta G_b(X \rightarrow Y)$ ) and ‘reverse’ paths ( $\Delta\Delta G_b(Y \rightarrow X)$ ) were computed and relative binding free-energies estimated with eq 3:

$$\Delta\Delta G_{b,calc}(X \rightarrow Y) = \frac{1}{2}[\Delta\Delta G_b(X \rightarrow Y) - \Delta\Delta G_b(Y \rightarrow X)] \quad (3)$$

To account for uncertainties due to sampling errors and biases from numerical integration of the free energy profiles, triplicates of each forward and reverse perturbation calculations were performed for each system. Each  $\Delta\Delta G_b(X \rightarrow Y)$  and  $\Delta\Delta G_b(Y \rightarrow X)$  value was taken as the average of the triplicates. Statistical uncertainties in the reported  $\Delta\Delta G_{b,calc}(X \rightarrow Y)$  values were estimated as the standard error of the mean with eq 4:

$$\text{err}(\Delta\Delta G_{b,calc}(X \rightarrow Y)) = \frac{s}{\sqrt{n}} \quad (4)$$

Where  $s$  is the standard deviation of free energy from the  $n=6$  replicas (3 forward and three reverse). For two step pathways, errors were propagated as the sum of errors from each steps.

### 2.3 Experimental Binding Free-energy Calculation:

The experimental RSL binding free-energies of the monosaccharides were calculated from the equilibrium dissociation constants ( $K_d$ )<sup>23,49</sup> measured by Surface Plasmon Resonance (SPR) assay at 298.15 K and standard reference concentration ( $C^0$ ) of 1 mol.L<sup>-1</sup> using eq 5:

$$\Delta G = RT \ln\left(\frac{K_d}{C^0}\right) \quad (5)$$

The experimental relative free-energy of binding of  $Y$  relative to  $X$  has been denoted as  $\Delta\Delta G_{b,exp}(X \rightarrow Y)$ . No experimental uncertainties in  $K_d$  measurement were reported, thus an experimental uncertainty of 0.4 kcal.mol<sup>-1</sup> in  $\Delta\Delta G_{b,exp}(X \rightarrow Y)$  was assumed as done by Brown et al. and Mikulskis et al.<sup>67,68</sup>

Overall agreement between theory and experiment was assessed by comparison of individual relative free-energy changes, and by computation of correlation coefficient ( $R^2$ ), mean unsigned error (MUE) and predictive indices (PI) for the full dataset as proposed by Pearlman and Charifson.<sup>69</sup> As done elsewhere<sup>70</sup>, uncertainties in these metrics were determined by resampling estimated binding free-energies. These were correlated to the experimentally measured binding free energies to produce distributions of  $R^2$ , MUE and PI values. The procedure was repeated 1 million times to yield a distribution of likely  $R^2$ , MUE and PI values for each simulation protocol. Uncertainties in the dataset metrics are quoted as an approximate  $\pm 1\sigma$  interval that covers 68% of the distributions.

### 3. Results

#### 3.1 Relative Free-energies of Methylated Monosaccharides

The mono-carbohydrates discussed here are hemiacetals at *C1* and therefore readily undergo anomerization. Their O1-methylated acetal counterparts, however, are stable and thus display well-defined anomers. Binding affinities of RSL are known for three methylated sugars, Me- $\alpha$ -L-Fucoside (**1**), Me- $\beta$ -D-Arabinoside (**2**) and Me- $\alpha$ -D-Mannoside (**3**), and are -8.6, -6.7 and -3.5 kcal $\cdot$ mol<sup>-1</sup>, respectively. Accordingly, a number of relative binding free-energy calculations for MeFuc $\rightarrow$ MeAra (**1** $\rightarrow$ **2**), MeFuc $\rightarrow$ MeMan (**1** $\rightarrow$ **3**) and MeAra $\rightarrow$ MeMan (**2** $\rightarrow$ **3**) transformations have been performed (Fig. 1).

Figure 2 illustrates the trend of calculated versus experimental binding free-energies for all the perturbations with the GAFF and GLYCAM force fields, and detailed figures are given in the supplementary information (Table S2). For perturbation **1** $\rightarrow$ **2**, the  $\Delta\Delta G_{b,calc}(\mathbf{1} \rightarrow \mathbf{2})$  values from both GAFF (1.9 $\pm$ 0.1 kcal $\cdot$ mol<sup>-1</sup>) and GLYCAM (1.8 $\pm$ 0.1 kcal $\cdot$ mol<sup>-1</sup>) are in an excellent agreement with  $\Delta\Delta G_{b,exp}(\mathbf{1} \rightarrow \mathbf{2})$  (1.9 kcal $\cdot$ mol<sup>-1</sup>). In **1** $\rightarrow$ **2**, the equatorial methyl group at position C5 in **1** is replaced by a hydrogen in **2**. This C6-methyl projects into a

1  
2  
3 hydrophobic region lined by the side chains of the residues Ile59, Ile61 and Trp10 of RSL  
4  
5 (Fig. 3). The change in the binding free-energy is particularly unfavorable in this case  
6  
7 because this scaffold modification results in a loss of hydrophobic interactions with the  
8  
9 protein environment.  
10

11  
12 In the perturbations **1**→**3**, and **2**→**3** larger groups of atoms need to be perturbed. The ring  
13  
14 carbon atoms C1, C2, C4 and C5 in **1** and **2** have been mapped to C5, C4, C3 and C2 in **3**,  
15  
16 respectively, such that the orientation of the O2, O3 and O4 hydroxyls remain unchanged.  
17  
18 However, the axial -OCH<sub>3</sub> group (methoxy) and the equatorial hydrogen of C1 in **1** are  
19  
20 perturbed into a hydrogen and hydroxymethyl group in **3**, respectively. Additionally, the axial  
21  
22 hydrogen of C5 in **1** is perturbed into a methoxy group, and the equatorial methyl at C5 in **1**  
23  
24 is also perturbed into a hydrogen in **3** (Fig. 1). For these two calculations, we found serious  
25  
26 convergence problems while evaluating  $\Delta G_w(\mathbf{1} \rightarrow \mathbf{3})$ ,  $\Delta G_p(\mathbf{1} \rightarrow \mathbf{3})$ ,  $\Delta G_w(\mathbf{2} \rightarrow \mathbf{3})$  and  
27  
28  $\Delta G_p(\mathbf{2} \rightarrow \mathbf{3})$ . Analysis of the free-energy gradients shows a considerable peak of the free-  
29  
30 energy gradients at  $\lambda \sim 0.7$  for **1**→**3** and **2**→**3**. This is mirrored at  $\lambda \sim 0.3$  for a perturbation  
31  
32 done through the reverse paths (**3**→**1** and **3**→**2**). The free-energy gradients have very large  
33  
34 values within these  $\lambda$  regions, and the resulting free-energy profile is noisy (Fig. 4A).  
35  
36 Increasing the length of the simulation or the number of  $\lambda$  points does not improve the  
37  
38 precision of the results (Fig. S1 & S2). A careful investigation was undertaken to diagnose  
39  
40 the problem. The perturbations were broken down into two sequential calculations involving  
41  
42 an intermediate compound **10** so as to minimize the magnitude of the structural changes  
43  
44 attempted in one step. Compound **1** and **2** were thus first perturbed into **10** where the  
45  
46 equatorial hydrogen of C1 in **1** and **2** is perturbed into a methyl. In the second step, this  
47  
48 methyl is then perturbed into the final hydroxymethyl in **3** (Fig. 5). A complication for the  
49  
50 GLYCAM force field is that the intermediate structure **10** does not have parameters, and  
51  
52 force field parameters were thus manually adapted by analogy from those used to describe **3**.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Since **10** is merely a convenient computational intermediate, accurate force field parameters  
4 are not crucial for this particular compound. It is evident from Fig 4 that the convergence is  
5 considerably improved, and the free-energy profile for **1**↔**10** and **10**↔**3** perturbations (Fig.  
6 4B-C & S3-S4) is quite smooth in comparison with the single-step perturbation (Fig. 4A).  
7 The free-energy gradient profiles for the **10**↔**3** perturbation are somewhat less smooth (Fig.  
8 S4) but the calculated free-energies for three independent simulations differ by less than 0.5  
9 kcal•mol<sup>-1</sup> from each other (Table S3 & S4), which is within the range of statistical error.  
10 Thus, it proves more effective to break down this complex perturbation into sequences of  
11 small perturbations that yield readily converged free-energy gradients.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21

22  
23 Extending simulations up to 10 ns for each window, or adding additional intermediate  $\lambda$   
24 values, did not provide any statistically significant difference in several chosen perturbations  
25 (Table S3 & S4). This indicates that the current setup that affords 4 ns per window is a good  
26 compromise between computational resources needed and accuracy of the results. Moreover,  
27 the mean  $\Delta\Delta G_b$  (**1** → **2**) for all six RSL binding sites (Table S1) is comparable to  
28  $\Delta\Delta G_{b,calc}$  (**1** → **2**) (Table S2 and Figure S5). This shows that the differences in  $\Delta\Delta G_b$  (**1** →  
29 **2**) among all the six binding sites are statistically insignificant, and an average  
30  $\Delta\Delta G_{b,calc}$  (**1** → **2**) estimated from three independent simulations of the first binding site (S1)  
31 is sufficient to yield well converged binding free-energy estimates.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

43 Figure 2 shows experimental and calculated change in the free-energy of binding of  
44 methylated monosaccharides **3** relative to **1** and **2** calculated by the two-step perturbation  
45 protocol using GAFF and GLYCAM force fields. The  $\Delta\Delta G_{b,calc}$  (**1** → **3**) values using GAFF  
46 and GLYCAM are 1.1±0.2 and 2.3±0.2 kcal•mol<sup>-1</sup> respectively, and  $\Delta\Delta G_{b,calc}$  (**2** → **3**) values  
47 are 0.2±0.3 and 0.5±0.2 kcal•mol<sup>-1</sup> respectively. While the two force fields follow a similar  
48 trend in values for  $\Delta\Delta G_{b,calc}$  (**1** → **3**) and  $\Delta\Delta G_{b,calc}$  (**2** → **3**), it is clear that the GLYCAM  
49 calculations provide better agreement with  $\Delta\Delta G_{b,exp}$  (**1** → **3**) and  $\Delta\Delta G_{b,exp}$  (**2** → **3**) (5.1 and  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 3.2 kcal•mol<sup>-1</sup> respectively). Relative binding free-energies for **1**→**3** and **2**→**3** are  
4  
5 overestimated by ca. 2.5 kcal•mol<sup>-1</sup>. The thermodynamic cycle closure error for **1**→**10**→**2**→**1**  
6  
7 and **1**→**2**→**10**→**1** perturbations (Fig. S6) from GAFF is -0.3 kcal•mol<sup>-1</sup> and 1.4 kcal•mol<sup>-1</sup>,  
8  
9 respectively. While thermodynamic cycle closure error from GLYCAM is 0.3 kcal•mol<sup>-1</sup> in  
10  
11 both cases. This indicates that the computed relative binding free-energies with this force  
12  
13 field appear well converged and that deviations from experimental data may be attributed to  
14  
15 inaccurate in force field parameters for **3** since the relative binding free-energy for **1**→**2**  
16  
17 matches well with experimental data. Indeed, the methoxy group in **3** is placed below Trp76,  
18  
19 which is expected to disrupt CH... $\pi$  stacking that is observed in **1** and **2** (Fig. 6A-6C). Current  
20  
21 classical force fields are of limited accuracy for the modeling of CH... $\pi$  stacking  
22  
23 interactions.<sup>71</sup> A close inspection of the GLYCAM simulations also shows that the  
24  
25 hydroxymethyl group of C5 in **3** is projected outward from the binding site, and creates a  
26  
27 hydrogen bond with the hydroxyl of Tyr37. The O6 hydroxyl can alter its orientation during  
28  
29 the simulation but this does not affects total free energy of binding (Fig. 6D). A similar  
30  
31 behavior is seen for perturbation **2**→**3**.  
32  
33  
34  
35

36 The resulted cycle closure error for the thermodynamic cycle (shown in Fig. S6) for  
37  
38 perturbation **1**→**2**→**3**→**1** is 0 and 1.0 kcal•mol<sup>-1</sup> for GLYCAM and GAFF simulations,  
39  
40 respectively, indicating confidence in the GLYCAM computed free-energies. By contrast, the  
41  
42 free-energies computed with the GAFF force field are not consistent. Turning now to the  
43  
44 accuracy of the force fields, for carbohydrates, relatively less accurate energies from  
45  
46 GAFF/AM1-BCC calculations are expected for of two reasons. Firstly, GAFF is  
47  
48 parameterized to cover mostly small organic compounds, and has not been optimized for  
49  
50 performance on hexopyranoses or carbohydrate-like structures.<sup>34</sup> Secondly, in asymmetric  
51  
52 molecules, such as sugars, the possibility of hydroxyl and hydroxymethyl group rotation  
53  
54 leads to ambiguity in selecting a single conformation for charge calculations. Thus, the  
55  
56  
57  
58  
59  
60

1  
2  
3 intramolecular interactions or solution properties are often poorly reproduced, unless  
4  
5 conformationally averaged charges are employed.<sup>72</sup>  
6

7  
8 Interestingly, both GAFF and GLYCAM systematically overestimate the binding affinity  
9  
10 of **3** to RSL. Mishra et al. also made a similar observation in a LIE study of this system using  
11  
12 the OPLS-AA 2005 force field, where the predicted absolute binding free-energy for **3** was  
13  
14 overestimated by approximately  $-2.0 \text{ kcal}\cdot\text{mol}^{-1}$ .<sup>27</sup> The possibility that experimental artifacts  
15  
16 have affected SPR measurements of the weak binding of compound **3** ( $K_d \sim 2.5 \text{ mM}$ ) should  
17  
18 not be ruled out.<sup>73</sup>  
19

### 20 21 *3.2 Relative Free-energies of L-Fucose and L-Galactose*

22  
23 Perturbation of MeFuc into L-Fuc (**1**→**4**) is computationally demanding because L-Fuc can  
24  
25 exist in both anomers in solution and the protein bound state. Thus, we decided to transform  
26  
27 MeFuc into both  $\alpha$ -L-Fuc (**4a**) and  $\beta$ -L-Fuc (**4b**). To close an additional thermodynamic cycle  
28  
29 the **4a**→**4b** perturbation was also performed. The calculated and experimental changes in the  
30  
31 binding free-energy of the **1**→**4** perturbations are presented in Fig 2.  
32  
33

34  
35 The  $\Delta\Delta G_{b,calc}(\mathbf{1} \rightarrow \mathbf{4a})$  and  $\Delta\Delta G_{b,calc}(\mathbf{1} \rightarrow \mathbf{4b})$  values using GLYCAM are  $0.3 \pm 0.3$   
36  
37  $\text{kcal}\cdot\text{mol}^{-1}$  and  $-0.2 \pm 0.1 \text{ kcal}\cdot\text{mol}^{-1}$ , respectively, which is in good agreement with the  
38  
39 experiment and well converged as shown by the cycle closure error of the thermodynamic  
40  
41 cycle **1**→**4b**→**4a**→**1** that is close to zero. Lower  $\Delta\Delta G_{b,calc}(\mathbf{1} \rightarrow \mathbf{4b})$  values using GAFF ( $-$   
42  
43  $0.6 \pm 0.1 \text{ kcal}\cdot\text{mol}^{-1}$ ) and GLYCAM ( $-0.2 \pm 0.1 \text{ kcal}\cdot\text{mol}^{-1}$ ) compared to  $\Delta\Delta G_{b,calc}(\mathbf{1} \rightarrow \mathbf{4a})$   
44  
45 from GAFF ( $0.2 \pm 0.1 \text{ kcal}\cdot\text{mol}^{-1}$ ) and GLYCAM ( $0.3 \pm 0.2 \text{ kcal}\cdot\text{mol}^{-1}$ ) suggests that RSL will  
46  
47 prefer to bind **4b**. Structural details from the computed trajectories show that the higher  
48  
49 affinity observed with GAFF may be due to additional electrostatic interactions of the O1  
50  
51 hydroxyl in **4b** with the protein. The O1 hydroxyl in the equatorial position (**4b**) interacts  
52  
53 with Arg17, Cys31 and Try37 (Fig. 7), but in the axial position (**4a**) it interacts largely with  
54  
55 water molecules. Regardless of whether RSL binds preferably **4a** or **4b** and on the basis of  
56  
57  
58  
59  
60

1  
2  
3 the experimental data, neither anomer can be a better binder than its methylated form **1**. Thus,  
4  
5 the GLYCAM calculations are in better agreement with the experiment but the trend for  
6  
7  $\Delta\Delta G_{b,calc}$  (**1**  $\rightarrow$  **4 $\beta$** ) from both GLYCAM and GAFF for **4 $\beta$**  is not correct. However, it must  
8  
9 be emphasized that  $\Delta\Delta G_{b,exp}$  (**1**  $\rightarrow$  **4**) is only 0.7 kcal/mol, thus the magnitude of the  
10  
11 deviation from experiment remains within range of the accuracy typically expected from an  
12  
13 AFE calculation. It can also be stressed here that the difference between calculated binding  
14  
15 free-energies for the **4 $\alpha$**  $\rightarrow$ **4 $\beta$**  perturbation is very small, suggesting that both the anomers may  
16  
17 readily bind to RSL. Table S1 shows the  $\Delta\Delta G_{b,calc}$  (**1**  $\rightarrow$  **4**) on all six binding sites of the  
18  
19 RSL. The  $\Delta\Delta G_{b,calc}$  for **4 $\alpha$**  and **4 $\beta$**  from a single run is slightly different among the binding  
20  
21 sites, also indicating an individual preference in each of the six binding sites. However,  
22  
23 similar to perturbation **1** $\rightarrow$ **2**, the mean  $\Delta\Delta G_{b,calc}$  for all the six binding sites is comparable to  
24  
25 the average from three independent simulations in the binding site S1. Similar free-energy  
26  
27 profiles (Fig. S5) in all the six binding sites of RSL suggests that differences in  $\Delta\Delta G_{b,calc}$  are  
28  
29 merely statistical errors.  
30  
31  
32  
33  
34

35 The  $\Delta\Delta G_{b,calc}$  values for both  $\alpha$ - and  $\beta$  anomers of L-Gal (**5**) using GAFF and GLYCAM  
36  
37 differ considerably from the SPR data (Fig. 1). Structurally fucose is a 6-deoxy galactose, i.e.,  
38  
39 galactose has an additional hydroxy group at C6 (Fig. 1). The C6 methyl in **1** is pointing  
40  
41 towards the hydrophobic patch created by the side chains of Ile59, Ile61 and Trp10. When the  
42  
43 one hydrogen of C6 is perturbed into a hydroxyl, it starts interacting with the water molecules  
44  
45 towards the protein surface (Fig. 8). It was found that the distance between O4 and O6  
46  
47 hydroxyls in **5** is quite stable in a position towards the protein surface that is away from O4,  
48  
49 where weak electrostatic interaction with Arg18 are possible (Fig. S7). All other interactions  
50  
51 are similar to **1** $\rightarrow$ **4** perturbation.  
52  
53  
54

55 On the other hand, unlike in the previous perturbations where the performance of GAFF  
56  
57 was comparable to GLYCAM, especially in modeling changes in hydrophobic interactions,  
58  
59  
60



1  
2  
3 binding free-energies for **5** from GLYCAM are more accurate (error ca.  $\sim 2$  kcal $\cdot$ mol $^{-1}$ ) than  
4  
5 those from GAFF (error ca.  $\sim 3$  kcal $\cdot$ mol $^{-1}$ ). This is attributed to inaccurate energetics for  
6  
7 strong CH $\dots\pi$  stacking interactions between non-polar hydrogens of C6 in **1** and the aromatic  
8  
9 amino acid residues of Trp76.<sup>74</sup> It has been shown by Wimmerova et al. that strong stacking  
10  
11 interactions between carbohydrate and Trp76 in RSL largely contributes to the binding  
12  
13 energy.<sup>75</sup> There are different views on the strength of stacking interactions between amino  
14  
15 acids and saccharides, but these interactions can be stronger than a hydrogen bond and  
16  
17 measured interaction energies are  $\sim 8.0$  kcal $\cdot$ mol $^{-1}$ .<sup>74,75</sup> Current force fields incorporate  
18  
19 stacking (CH $\dots\pi$  interactions) partially in the form of van der Waals interaction, but, there is  
20  
21 no explicit term in the force fields to model CH $\dots\pi$  interactions. The CH $\dots\pi$  interactions are  
22  
23 strongest when a CH bond is pointing directly towards the center of the aromatic ring.<sup>55</sup> In **1**,  
24  
25 C6 methyl is strongly involved in the CH $\dots\pi$  interaction with Trp76, but upon perturbation to  
26  
27 hydroxymethyl in **5**, this group rearranges to interact with water. The CH groups in  
28  
29 hydroxymethyl can still form CH $\dots\pi$  interaction with the aromatic ring of Trp76, but this is  
30  
31 modeled poorly with the present force fields. Hydrogens of C6 in **5** are not perpendicular to  
32  
33 the aromatic ring due to dominating O6 hydroxyl-water interactions. This orientation of the  
34  
35 O6 hydroxyl group is further stabilized by favorable electrostatic interactions with Arg18.  
36  
37 Thus, a plausible explanation for the overestimation of the calculated binding free energies of  
38  
39 **5** by ca. 2, and 3 kcal $\cdot$ mol $^{-1}$  from GLYCAM and GAFF respectively, is the inability of these  
40  
41 force fields to account for CH $\dots\pi$  interactions, consequently promoting additional  
42  
43 electrostatic interactions of O6 with water and Arg18.  
44  
45  
46  
47  
48

### 49 50 3.3 Relative Free-energies of D-Fructose

51  
52 The MeFuc to D-Fru perturbation (**1** $\rightarrow$ **6**) is the largest structural change attempted in this  
53  
54 study. D-Fru (**6**) is a weak binder ( $\Delta G_{b,exp} = -5.5$  kcal $\cdot$ mol $^{-1}$ ). As with the previously  
55  
56 discussed non-methylated monosaccharides, **1** $\rightarrow$ **6** perturbations are performed for both  $\alpha$ -D-  
57  
58  
59  
60

1  
2  
3 Fru (**6 $\alpha$** ) and  $\beta$ -D-Fru (**6 $\beta$** ) anomers. The **1** $\rightarrow$ **6 $\alpha$**  and **1** $\rightarrow$ **6 $\beta$**  perturbations do not alter  
4 interactions of O2, O3 and O4 hydroxyl of **1**, and the changes are limited to C1 and C5. The  
5 axial methoxy and the equatorial hydrogen of C1 in **1** are morphed respectively into  
6 hydroxymethyl and hydroxyl groups in **6 $\alpha$**  and vice versa in **6 $\beta$** . The equatorial methyl of C5  
7 in **1** is perturbed into hydrogen in both **6 $\alpha$**  and **6 $\beta$**  (Fig. 1). The free-energy gradients  
8 accumulated for the **1** $\rightarrow$ **6 $\alpha$**  perturbation are sufficiently converged along the given pathway.  
9 However, the computed free-energy gradients for perturbation **1** $\rightarrow$ **6 $\beta$**  show a similar noisy  
10 profile as seen for the **1** $\rightarrow$ **3** and **2** $\rightarrow$ **3** perturbations discussed previously (Fig. S8).  
11 Performing the perturbation from **1** to **6 $\beta$**  in two steps via the intermediate **10** resolved the  
12 issue (Fig. S9).  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

25 The  $\Delta\Delta G_{b,calc}$  values obtained with GLYCAM for both, **1** $\rightarrow$ **6 $\alpha$**  ( $3.5\pm 0.4$  kcal $\cdot$ mol $^{-1}$ ) and  
26 **1** $\rightarrow$ **6 $\beta$**  ( $1.9\pm 0.2$  kcal $\cdot$ mol $^{-1}$ ) perturbation are in closer agreement with  $\Delta\Delta G_{b,exp}(\mathbf{1} \rightarrow \mathbf{6})$  ( $3.1$   
27 kcal $\cdot$ mol $^{-1}$ ) than the values obtained with GAFF ( $-0.3\pm 0.2$  kcal $\cdot$ mol $^{-1}$  and  $-0.2\pm 0.2$  kcal $\cdot$ mol $^{-1}$ ,  
28 respectively). Anomer **6 $\beta$**  is predicted to bind more favorably than **6 $\alpha$**  with GLYCAM. In  
29 simulations performed with both GAFF and GLYCAM force fields, the equatorial  
30 hydroxymethyl of C2 in **6 $\beta$**  forms a strong hydrogen bond with the Tyr37, which could not be  
31 formed with anomer **6 $\alpha$** . This anomer has a hydroxyl group in equatorial position of C2,  
32 which is too far to establish direct interaction with Tyr37. This explains why **6 $\beta$**  binds  
33 stronger than **6 $\alpha$** . Further analysis of the computed trajectories is useful to establish why the  
34 binding energetics differ between the two force fields (Table S2). A close inspection of the  
35 free-energy gradients and trajectories of end-states does not reveal any noticeable features in  
36 the geometry or structural interactions of the molecules. A plausible explanation for the poor  
37 performance of GAFF here is that AM1-BCC may significantly underestimate the solvation  
38 free-energy of the carbohydrate ligand. Others have reported a systematic underestimation of  
39 hydration free-energies of ca. 1.5 kcal/mol for alcohols.<sup>76</sup> In saccharides, these errors could  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 be larger due to the greater number of hydroxyl groups present in the structures; compound **6**  
4  
5 contains five hydroxyls. The axial group R<sub>1</sub> is always solvent exposed and the equatorial  
6  
7 group R<sub>2</sub> interacts with water in **6a** (R<sub>2</sub>= -OH), but makes a hydrogen bond with Tyr37 in **6b**  
8  
9 (R<sub>2</sub>= -CH<sub>2</sub>OH). This suggests that accurate prediction of hydration energetics for hydroxyls  
10  
11 is critical to model carbohydrate-protein binding.  
12

### 13 14 3.4 Relative Free-energies of D-Rhamnose

15  
16 D-Rhamnose (**7**) is a weak binder ( $\Delta G_{b,exp} = -4.6 \text{ kcal}\cdot\text{mol}^{-1}$ ) of RSL. As for other non-  
17  
18 methylated monosaccharides, the **1**→**7** perturbation was performed separately for both  $\alpha$  (**7a**)  
19  
20 and  $\beta$  (**7b**) anomers. These perturbations do not affect orientations of O2, O3 and O4  
21  
22 hydroxyls in **1**, but changes are made at C1 and C5 in **1** only. The axial methoxy group and  
23  
24 the equatorial hydrogen of C1 in **1** are perturbed into hydrogen and methyl groups  
25  
26 respectively, whereas the axial hydrogen and the equatorial methyl of C5 in **1** are morphed  
27  
28 into hydroxyl and hydrogen for **7a** and vice versa for **7b**. The free-energy gradients collected  
29  
30 for **1**→**7a** and **1**→**7b** perturbations are smooth and well converged along the pathway. The  
31  
32 trends of  $\Delta\Delta G_{b,calc}$  values obtained with GAFF and GLYCAM are in agreement with  
33  
34  $\Delta\Delta G_{b,exp}$  (Table S2). While both GAFF and GLYCAM give similar trends, it is found that  
35  
36 GLYCAM calculations produced better results compared to GAFF, which overestimates  
37  
38 energies by ca. 2.0 kcal/mol.  
39  
40  
41  
42

43  
44 The  $\Delta\Delta G_{b,calc}$  values from GLYCAM for **7a** and **7b** are  $4.3\pm 0.1$  and  $2.6\pm 0.1 \text{ kcal}\cdot\text{mol}^{-1}$ ,  
45  
46 respectively, whereas  $\Delta\Delta G_{b,exp}(\mathbf{1} \rightarrow \mathbf{7})$  is  $4.0 \text{ kcal}\cdot\text{mol}^{-1}$ . Thus, the calculations suggest that  
47  
48 the binding of **7b** is favored over **7a**, and the binding free energy is too favorable by ca. 1.4  
49  
50  $\text{kcal}\cdot\text{mol}^{-1}$ , too. The intermolecular interactions of both anomers during the simulations are  
51  
52 quite similar except for electrostatic interactions between the O1 hydroxyl of C1 in **7a** and  
53  
54 Trp81. The O1 hydroxyl resides inside the hydrophobic pocket in **7b**. Interestingly, **7a** also  
55  
56 samples a second binding mode during the simulations (Fig. S10). In this alternative binding  
57  
58  
59  
60

1  
2  
3 mode, hydrogen bonds formed by O2, O3 and O4 hydroxyl of **7a** are still maintained, but O1  
4 hydroxyl now interacts with water molecules. Over a 10 ns simulation, **7a** interconverts twice  
5  
6 between the primary and secondary binding modes. This suggests that these two binding  
7  
8 modes for **7a** are possible, and **7a** may have similar binding energy in both the binding  
9  
10 modes as transitions are observed on this short timescale. The overall difference in binding  
11  
12 affinities for **7a** and **7b** relative to **1** is mainly attributed to the loss of hydrophobic  
13  
14 interactions mediated by C6 methyl in **1**, and additional contribution from axial  
15  
16 hydroxymethyl to hydrogen and equatorial hydrogen to methyl perturbation at C1 in **1**  
17  
18  
19

### 20 21 *3.5 Relative Free-energies of non-binder D-Galactose and L-Rhamnose*

22  
23 D-Galactose (**8**) and L-Rhamnose (**9**) have not shown any significant binding in SPR  
24  
25 experiments. Thus, both **8** and **9** are either non-binders or their binding was so weak that it  
26  
27 could not be measured. The lowest  $K_d$  measured using SPR measurement is 2.5 mM for **3**  
28  
29 (MeMan). As a consequence, it is assumed that **8** and **9** are either non-binders or their  $K_d$  is  
30  
31 higher than 2.5 mM. Both **8** and **9** are denoted as non-binders in this report. Based on the  $K_d$   
32  
33 values for **3**,  $\Delta\Delta G_{b,exp}$  is expected to be greater than 6.0 kcal•mol<sup>-1</sup> for **1**→**8** and **1**→**9**  
34  
35 perturbations. The **1**→**8** perturbation involves replacing the methoxy and hydrogen groups of  
36  
37 C1 in **1** by hydrogen and hydroxymethyl groups, and methyl and hydrogen of C5 in **1** by  
38  
39 hydroxyl and hydrogen groups (Fig. 1). Additionally, both **1**→**8** and **1**→**9** perturbations  
40  
41 involve hydrogen to hydroxyl and vice-versa transformation at C2 and C4. The free-energy  
42  
43 gradients for the **1**→**8** perturbation were not well converged with the one-step protocol.  
44  
45 Transformation of the equatorial H of C1 in **1** to hydroxymethyl displayed again convergence  
46  
47 issues. Thus, the same intermediate **10** was used to obtain smooth gradients for **1**→**8a** and  
48  
49 **1**→**8b** perturbations. Interestingly **8** and **9** are the only monosaccharides where orientation of  
50  
51 O2 and O4 hydroxyl differs as compared to the high affinity binders **1** to **5**. Because of this,  
52  
53 these non-binders cannot interact with binding site residues Trp76, Glu28 or Arg17 as  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 observed for more potent ligands. Consequently **8** and **9** adopt binding modes that differ from  
4  
5 **1** during the simulations.

6  
7 The  $\Delta\Delta G_{b,calc}$  values for **1**→**8 $\alpha$**  and **1**→**8 $\beta$**  are  $5.2\pm 0.2$  and  $2.7\pm 0.4$  kcal•mol<sup>-1</sup> with GAFF  
8  
9 and  $8.0\pm 0.4$  and  $6.5\pm 0.3$  kcal•mol<sup>-1</sup> with GLYCAM, respectively. While both GAFF and  
10  
11 GLYCAM give results in qualitative agreement with experiment, once again GLYCAM  
12  
13 predictions are quantitatively much closer to the expected change of  $>6.5$  kcal•mol<sup>-1</sup> in  
14  
15 binding free energies (Table 1). For the **1**→**8** perturbation, the GAFF/AM1-BCC energies are  
16  
17 quite different from the GLYCAM energies, showing a similar trend of errors as seen for  
18  
19 **1**→**6 $\beta$** , which might also be ascribed to systematic errors in hydration free-energies of  
20  
21 alcohols, though this cannot be quantified exactly as no experimental hydration free-energies  
22  
23 are available in the literature for these monosaccharides.

24  
25  
26  
27 The MeFuc→L-Rha perturbations (**1**→**9**) for both **9 $\alpha$**  and **9 $\beta$**  anomers converged well in a  
28  
29 single step transformation. The  $\Delta\Delta G_{b,calc}$  (**1** → **9 $\alpha$** ) and  $\Delta\Delta G_{b,calc}$  (**1** → **9 $\beta$** ) values are  
30  
31  $6.2\pm 0.1$  and  $5.0\pm 0.3$  kcal•mol<sup>-1</sup> with GAFF and  $6.7\pm 0.2$  and  $6.9\pm 0.2$  kcal•mol<sup>-1</sup> with  
32  
33 GLYCAM respectively. As expected, the calculated change in the free-energy of binding  
34  
35 using GLYCAM is greater than  $6.0$  kcal•mol<sup>-1</sup> for both anomers. In the simulations, the  $\alpha/\beta$   
36  
37 L-Rha ligands are quite unstable inside the binding site. Figure 9 shows the starting  
38  
39 conformation in green and other conformations sampled at  $\lambda=0$  during the **9**→**1** perturbation.  
40  
41 During the simulation, both **8** and **9** attains conformations where interaction of the saccharide  
42  
43 with protein side-chains differs significantly. Thus, the simulations clearly suggest that both **8**  
44  
45 and **9** are non-binders owing to the lack of a defined binding mode. This could be the reason  
46  
47 why significant hysteresis was observed between forward and reverse runs for the  
48  
49 perturbations from **1** to **8** and **9** (Table S1) as the full range of weakly interacting binding  
50  
51 modes may not have been fully explored over short time scale MD simulations. Other factors  
52  
53 such as biases due the numerical integration of the free energy gradients may not be ruled  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 out. Although this hysteresis has little effect on the ranking of the ligands in this dataset, this  
4  
5 matter warrants further future investigation.  
6

#### 7 **4. Discussion and Conclusion**

8  
9 A procedure that allows an efficient binding free-energy calculation for monosaccharide  
10 ligand substituents in protein-carbohydrate complexes using GAFF/AM1-BCC and  
11 GLYCAM force fields is described. The current results show that accurate relative binding  
12 free-energies of protein-carbohydrate complexes are possible using AFE calculations.  
13  
14  
15  
16

17  
18 Moreover, unlike in previous studies performed with alternative modelling methods,<sup>23,27</sup> AFE  
19 calculations are here able to reliably discriminate non-binders, weak-binders and potent-  
20 binders. On the technical side, setting up AFE calculations for carbohydrates presented a  
21 number of difficulties, and in a number of instances, two steps pathways were shown to  
22 outperform direct perturbations between pairs of monosaccharides. A two step pathway is  
23 recommended when appearing/disappearing hydroxymethyl groups at equatorial position of  
24 C1 or C5 carbon of hexopyranose in their  ${}^1C_4$  or  ${}^4C_1$  chair conformation. Provided the  
25 perturbation pathways are adequately defined, the AFE calculation setup using 16 non-  
26 equally distant  $\lambda$  windows and four ns simulation for each window is a good compromise  
27 between computational resources needed and accuracy of the results.  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

40  
41 The main results of this study are a careful assessment of the accuracy of the GAFF/AM1-  
42 BCC and GLYCAM force fields in computing free-energies of binding of protein-  
43 carbohydrate complexes. To achieve high precision, each relative binding free-energy was  
44 estimated from at least 6 independent calculations. In general, GLYCAM outperformed  
45 GAFF/AM1-BCC. The results suggest that GAFF/AM1-BCC can be as accurate as  
46 GLYCAM when perturbations only affect a change in hydrophobic interaction as for the  
47  $1 \rightarrow 2$  perturbation. GAFF/AM1-BCC proved to be less accurate in modeling the energetics of  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 hydroxyl groups and this is attributed to systematic errors in hydration energetics for this  
4  
5 functional group.  
6

7  
8 By contrast, the GLYCAM force field reproduces experiential changes in the free-energy  
9  
10 of binding in most of the cases. Assuming a relative binding free-energy of  $6.5 \text{ kcal}\cdot\text{mol}^{-1}$  for  
11  
12 all the non-binders, GLYCAM produces a mean unsigned error  $\text{MUE} = 1.1 \pm 0.06 \text{ kcal}\cdot\text{mol}^{-1}$ ,  
13  
14 a correlation coefficient  $R^2 = 0.85 \pm 0.02$  and a predictive index  $\text{PI} = 0.94 \pm 0.01$ . In  
15  
16 comparison GAFF/AM1-BCC achieves on this dataset a statistically significant inferior  
17  
18 performance, with  $\text{MUE} = 2.60 \pm 0.08 \text{ kcal}\cdot\text{mol}^{-1}$ ,  $R^2 = 0.59 \pm 0.03$  and  $\text{PI} = 0.80 \pm 0.01$ .  
19  
20 Thus, while GAFF shows reasonable ranking abilities, the mean errors with GAFF are too  
21  
22 large to reliably discriminate potent binders, weak binders and non-binders which is  
23  
24 important for hit-to-lead and lead optimization purposes. Thus the versatility of the current  
25  
26 version of GAFF and ease of parameterization of carbohydrates derived ligands is more than  
27  
28 offset by significant performance degradation and it is preferable to focus efforts on  
29  
30 extending GLYCAM parameter sets for detailed modelling studies of carbohydrate-based  
31  
32 ligands. The present results can be compared with other free-energy studies of carbohydrates  
33  
34 reported in the literature. Kadirvelraj et al achieved an accuracy of  $0.5 \text{ kcal}\cdot\text{mol}^{-1}$  using  
35  
36 GLYCAM for hydroxyl to methoxy and hydroxyl to hydroxyethyl perturbation<sup>12</sup>, and  $\sim 0.5$   
37  
38  $\text{kcal}\cdot\text{mol}^{-1}$  for 6 mutations in antibody-carbohydrate-antigen complexes<sup>77</sup> using  
39  
40 GLYCAM\_98R. Although a similar level of accuracy is obtained using LIE models  
41  
42 parameterized on protein-carbohydrate complexes, transferability of parameters to other  
43  
44 systems remains a challenge.<sup>27</sup> Errors are higher when using MM-PB/GBSA on protein-  
45  
46 carbohydrate complexes.<sup>24,28</sup>  
47  
48  
49  
50

51  
52 As the same dataset has been studied previously by docking and LIE methodologies, a  
53  
54 systematic comparison of the predictive power of a broad range of methodologies is  
55  
56 presented in Figure 10. AFE calculations using GLYCAM provides the highest correlation  
57  
58  
59  
60

1  
2  
3 (R<sup>2</sup>=0.85±0.02), predictive index (PI=0.94±0.01) and lowest MUE (1.1±0.06 kcal.mol<sup>-1</sup>) with  
4  
5 experiment data (see also Table S7 for details). Among the several docking programs  
6  
7 evaluated only AutoDock3 with RESP did well in terms of correlation (R<sup>2</sup>=0.78) and  
8  
9 predictive index (PI=0.83), but not MUE (3.6 kcal.mol<sup>-1</sup>). Thus, AutoDock3 outperforms  
10  
11 GAFF on this dataset. The performance of the tool is sensitive to the charge model used, and  
12  
13 results obtained with Gasteiger charges are inferior. All other docking programs are much  
14  
15 inferior to GLYCAM and GAFF. The LIE approach (performed with an OPLS-AA 2005  
16  
17 force field) is broadly speaking comparable in accuracy with the present AFE results obtained  
18  
19 with GAFF/AM1-BCC (R<sup>2</sup>=0.60, PI=0.80, MUE 2.6 kcal.mol<sup>-1</sup>). Overall, AFE calculations  
20  
21 with GLYCAM outperform all other methods, but the accuracy of GAFF is comparable to  
22  
23 LIE or the best performing docking protocols.  
24  
25  
26

27  
28 The accuracy achieved with GLYCAM in the present study is thus very encouraging, and  
29  
30 we expect a broad applicability to other protein-carbohydrate systems albeit within the limits  
31  
32 of current force fields. The present study is also the first to include non-binders in the dataset,  
33  
34 thus demonstrating the capability of TI to predict reliably a wide range of binding energetics.  
35  
36 The only noticeable error with GLYCAM is for the relative binding free-energies of MeMan  
37  
38 (**3**) and L-Gal (**5**), which is overestimated using both GAFF/AM1-BCC and GLYCAM. Only  
39  
40 ligands **3** and **5** have a hydroxymethyl and methoxy group at ring carbon C5 and C1, which is  
41  
42 just below Trp76. Thus, both **3** and **5** remains a challenge presumably because of the inability  
43  
44 to quantify CH... $\pi$  interactions between carbohydrates and aromatic amino acid residues  
45  
46 accurately in current force fields. Such errors are also possible for **8** but this interaction was  
47  
48 not apparent owing to a lack of defined binding mode during the simulations. Nevertheless,  
49  
50 predicted energies for L-Gal (**5**) are in much better agreement than those predicted with  
51  
52 MM/PBSA where the absolute binding energy of L-Gal complexed with PA-IIL was  
53  
54 overestimated by  $\sim -20$  kcal/mol.<sup>24</sup> While hexopyranoses assume preferentially a chair  
55  
56  
57  
58  
59  
60



1  
2  
3 conformation ( ${}^1C_4$  L-Galacto or  ${}^4C_1$  D-Manno), ring conformation sampling of the  
4  
5 hexopyranoses has not been addressed in this work.

6  
7 To conclude the present study has identified protocols for reliable computation of the  
8  
9 binding energetics of monosaccharides by AFE calculations, identified classes of  
10  
11 carbohydrate-protein interactions that are well addressed by existing classical force fields,  
12  
13 and other interactions that may need further attention. The procedures outlined here could be  
14  
15 used further to explore protein recognition by oligo/poly saccharides, with immediate  
16  
17 practical applications for carbohydrate-based drug design.

#### 20 21 **Abbreviations:**

22  
23 MeFuc: Me- $\alpha$ -L-Fuc, MeAra: Me- $\beta$ -D-Ara, MeMan: Me- $\alpha$ -D-Man.,  $\alpha$ Fuc:  $\alpha$ -L-Fuc,  $\beta$ Fuc:  
24  
25  $\beta$ -L-Fuc,  $\beta$ Fru:  $\beta$ -D-Fru,  $\alpha$ Fru:  $\alpha$ -D-Fru, W: Water, P: Protein, Cal: Calculated, Exp :  
26  
27 Experimental, methoxy:  $-OCH_3$ , Hydroxymethyl:  $-CH_2OH$

#### 28 29 **ASSOCIATED CONTENT**

30  
31  
32 Absolute binding free energies, plot of free energy gradients, and relative binding energies  
33  
34 from LIE and docking and AFE calculations are provided as supporting information. This  
35  
36 information is available free of charge via the internet at <http://pubs.acs.org>

#### 37 38 39 **AUTHOR INFORMATION**

##### 40 41 **Corresponding Authors**

42  
43 \* Phone: +420-549494947 (JK). E-mail: [jkoca@ceitec.cz](mailto:jkoca@ceitec.cz) (J.K)

44  
45 \* Phone: +44-131-6504797 (JM). E-mail: [mail@julienmichel.net](mailto:mail@julienmichel.net) (JM)

##### 46 47 48 **Authors Contribution**

49  
50  
51 SKM, JM and JK conceived and designed the calculations. SKM performed calculations and  
52  
53 analyzed the data. GC, HL and JM provided GLYCAM support in FESetup and  
54  
55

1  
2  
3 Sire/OpenMM calculations. The manuscript was written through contributions of all authors.  
4  
5 All authors have given approval to the final version of the manuscript.  
6  
7

### 8 **Funding Sources**

9  
10  
11 This work was supported by the Grant Agency of the Czech Republic [13-25401S]. This  
12  
13 work was further supported by the project “CEITEC–Central European Institute of  
14  
15 Technology” [CZ.1.05/1.1.00/02.0068] from the European Regional Development Fund.  
16  
17 Additional support was provided by the project "Employment of Newly Graduated Doctors  
18  
19 of Science for Scientific Excellence" [CZ.1.07/2.3.00/30.0009] to SKM, co-financed from the  
20  
21 European Social Fund and the state budget of the Czech Republic. JM is supported by a  
22  
23 Royal Society University Research Fellowship. The research leading to these results has  
24  
25 received funding from the European Research Council under the European Union’s Seventh  
26  
27 Framework Programme (FP7/2007-2013)/ERC grant agreement no. 336289. HL  
28  
29 acknowledges support from CCP-BioSim for development of the FEsetup software. Access to  
30  
31 the MetaCentrum computing facilities provided under the programme “Large Infrastructure  
32  
33 Projects for Research, Development, and Innovation” [LM2010005]; and CERIT-SC  
34  
35 computing facilities provided by the CERIT Scientific Cloud programme center, part of the  
36  
37 Operational Program Research and Development for Innovation [reg. no.  
38  
39 CZ.1.05/3.2.00/08.0144] are appreciated.  
40  
41  
42  
43  
44

### 45 **Notes**

46  
47  
48 The authors declare no competing financial interests.  
49  
50

### 51 **References:**

- 52  
53  
54 (1) De Ruiter, A.; Oostenbrink, C. Free Energy Calculations of Protein-Ligand Interactions.  
55 *Curr. Opin. Chem. Biol.* **2011**, *15* (4), 547–552.  
56 (2) Steinbrecher, T.; Labahn, A. Towards Accurate Free Energy Calculations in Ligand  
57 Protein-Binding Studies. *Curr. Med. Chem.* **2010**, *17* (8), 767–785.  
58  
59  
60

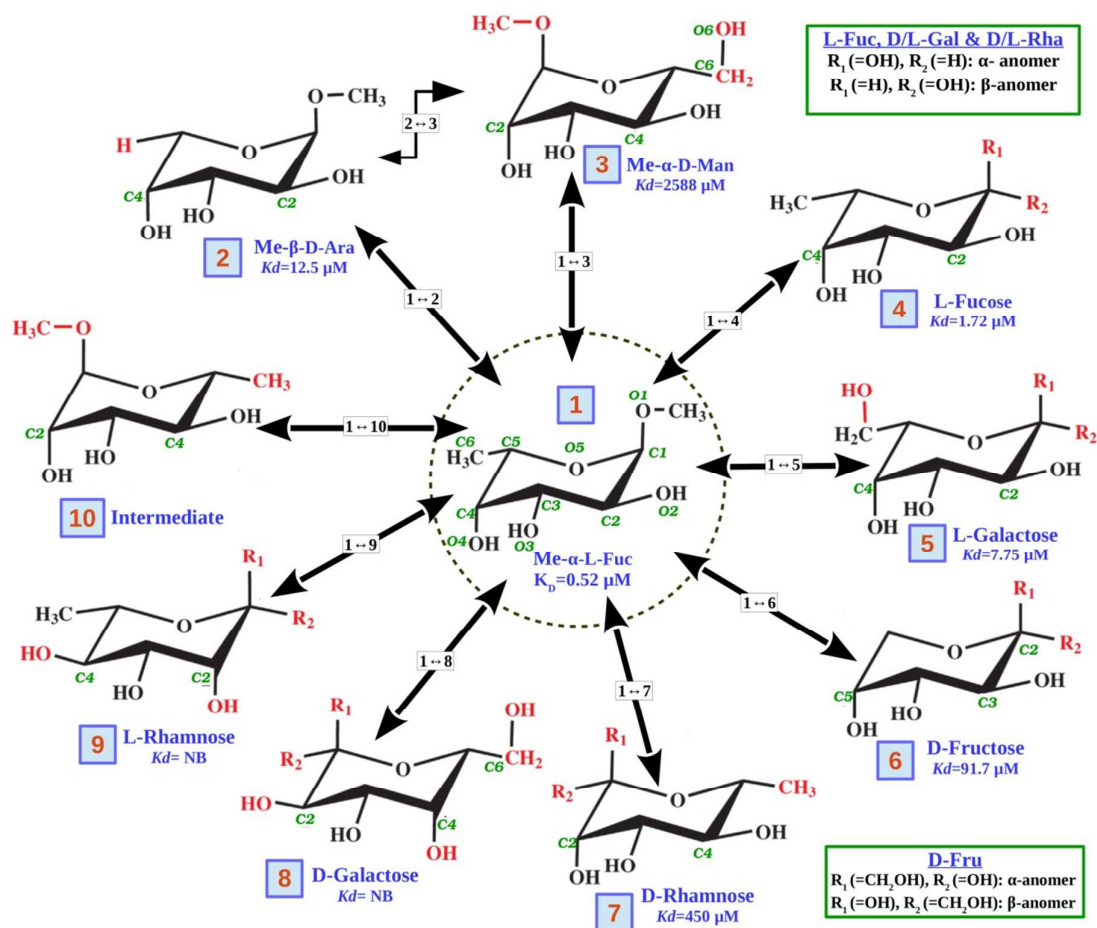
- 1
- 2
- 3 (3) Michel, J.; Foloppe, N.; Essex, J. W. Rigorous Free Energy Calculations in Structure-
- 4 Based Drug Design. *Mol. Inform.* **2010**, *29* (8-9), 570–578.
- 5 (4) Bash, P.; Singh, U.; Brown, F.; Langridge, R.; Kollman, P. Calculation of the Relative
- 6 Change in Binding Free Energy of a Protein-Inhibitor Complex. *Science* **1987**, *235*
- 7 (4788), 574–576.
- 8 (5) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical
- 9 Phenomena. *Chem. Rev.* **1993**, *93* (7), 2395–2417.
- 10 (6) Homeyer, N.; Stoll, F.; Hillisch, A.; Gohlke, H. Binding Free Energy Calculations for
- 11 Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design
- 12 Context. *J. Chem. Theory Comput.* **2014**, *10* (8), 3331–3344.
- 13 (7) Pan, K.; Deem, M. W. Predicting Fixation Tendencies of the H3N2 Influenza Virus by
- 14 Free Energy Calculation. *J. Chem. Theory Comput.* **2011**, *7* (5), 1259–1272.
- 15 (8) Tzoupis, H.; Leonis, G.; Mavromoustakos, T.; Papadopoulos, M. G. A Comparative
- 16 Molecular Dynamics, MM-PBSA and Thermodynamic Integration Study of Saquinavir
- 17 Complexes with Wild-Type HIV-1 PR and L10I, G48V, L63P, A71V, G73S, V82A and
- 18 I84V Single Mutants. *J. Chem. Theory Comput.* **2013**, *9* (3), 1754–1764.
- 19 (9) Michel, J. Current and Emerging Opportunities for Molecular Simulations in Structure-
- 20 Based Drug Design. *Phys. Chem. Chem. Phys.* **2014**, *16* (10), 4465–4477.
- 21 (10) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A
- 22 Review. *J. Chem. Theory Comput.* **2014**, *10* (7), 2632–2647.
- 23 (11) Shirts, M. Best Practices in Free Energy Calculations for Drug Design. *Methods Mol.*
- 24 *Biol. Clifton NJ* **2012**, *819*, 425–467.
- 25 (12) Kadirvelraj, R.; Foley, B. L.; Dyekjær, J. D.; Woods, R. J. Involvement of Water in
- 26 Carbohydrate–Protein Binding: Concanavalin A Revisited. *J. Am. Chem. Soc.* **2008**, *130*
- 27 (50), 16933–16942.
- 28 (13) Abhilash, J.; Dileep, K. V.; Palanimuthu, M.; Geethanandan, K.; Sadasivan, C.; Haridas,
- 29 M. Metal Ions in Sugar Binding, Sugar Specificity and Structural Stability of
- 30 Spatholobus Parviflorus Seed Lectin. *J. Mol. Model.* **2013**, *19* (8), 3271–3278.
- 31 (14) Krivan, H. C.; Plosila, L.; Zhang, L.; Holt, V.; Kyogashima, M. Cell Surface
- 32 Carbohydrates as Adhesion Receptors for Many Pathogenic and Opportunistic
- 33 Microorganisms. In *Microbial Adhesion and Invasion*; Hook, M., Switalski, L., Eds.;
- 34 Springer New York, 1992; pp 1–13.
- 35 (15) Jones, C. Vaccines Based on the Cell Surface Carbohydrates of Pathogenic Bacteria. *An.*
- 36 *Acad. Bras. Ciênc.* **2005**, *77* (2), 293–324.
- 37 (16) Tuccillo, F. M.; de Laurentiis, A.; Palmieri, C.; Fiume, G.; Bonelli, P.; Borrelli, A.;
- 38 Tassone, P.; Scala, I.; Buonaguro, F. M.; Quinto, I.; Scala, G. Aberrant Glycosylation as
- 39 Biomarker for Cancer: Focus on CD43. *BioMed Res. Int.* **2014**, *2014*, e742831.
- 40 (17) Gabius, H.-J.; Siebert, H.-C.; André, S.; Jiménez-Barbero, J.; Rüdiger, H. Chemical
- 41 Biology of the Sugar Code. *Chembiochem Eur. J. Chem. Biol.* **2004**, *5* (6), 740–764.
- 42 (18) Albersheim, P.; Anderson-Prouty, A. J. Carbohydrates, Proteins, Cell Surfaces, and the
- 43 Biochemistry of Pathogenesis. *Annu. Rev. Plant Physiol.* **1975**, *26* (1), 31–52.
- 44 (19) Neumann, D.; Kohlbacher, O.; Lenhof, H.-P.; Lehr, C.-M. Lectin-Sugar Interaction.
- 45 Calculated versus Experimental Binding Energies. *Eur. J. Biochem. FEBS* **2002**, *269*
- 46 (5), 1518–1524.
- 47 (20) Kerzmann, A.; Neumann, D.; Kohlbacher, O. SLICK--Scoring and Energy Functions
- 48 for Protein-Carbohydrate Interactions. *J. Chem. Inf. Model.* **2006**, *46* (4), 1635–1642.
- 49 (21) Kerzmann, A.; Fuhrmann, J.; Kohlbacher, O.; Neumann, D. BALLDock/SLICK: A
- 50 New Method for Protein-Carbohydrate Docking. *J. Chem. Inf. Model.* **2008**, *48* (8),
- 51 1616–1625.
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

- 1  
2  
3 (22) Adam, J.; Kříž, Z.; Prokop, M.; Wimmerová, M.; Koča, J. In Silico Mutagenesis and  
4 Docking Studies of Pseudomonas Aeruginosa PA-III Lectin — Predicting Binding  
5 Modes and Energies. *J. Chem. Inf. Model.* **2008**, *48* (11), 2234–2242.
- 6 (23) Mishra, S. K.; Adam, J.; Wimmerova, M.; Koca, J. In Silico Mutagenesis and Docking  
7 Study of Ralstonia Solanacearum RSL Lectin: Performance of Docking Software To  
8 Predict Saccharide Binding. *J. Chem. Inf. Model.* **2012**, *52* (5), 1250–1261.
- 9 (24) Mishra, N. K.; Kríz, Z.; Wimmerová, M.; Koca, J. Recognition of Selected  
10 Monosaccharides by Pseudomonas Aeruginosa Lectin II Analyzed by Molecular  
11 Dynamics and Free Energy Calculations. *Carbohydr. Res.* **2010**, *345* (10), 1432–1441.
- 12 (25) Fadda, E.; Woods, R. J. Molecular Simulations of Carbohydrates and Protein-  
13 Carbohydrate Interactions: Motivation, Issues and Prospects. *Drug Discovery. Today*  
14 **2010**, *15* (15-16), 596–609.
- 15 (26) Bryce, R. A.; Hillier, I. H.; Naismith, J. H. Carbohydrate-Protein Recognition:  
16 Molecular Dynamics Simulations and Free Energy Analysis of Oligosaccharide Binding  
17 to Concanavalin A. *Biophys. J.* **2001**, *81* (3), 1373–1388.
- 18 (27) Mishra, S. K.; Sund, J.; Åqvist, J.; Koča, J. Computational Prediction of  
19 Monosaccharide Binding Free Energies to Lectins with Linear Interaction Energy  
20 Models. *J. Comput. Chem.* **2012**, *33* (29), 2340–2350.
- 21 (28) Topin, J.; Arnaud, J.; Sarkar, A.; Audfray, A.; Gillon, E.; Perez, S.; Jamet, H.; Varrot,  
22 A.; Imberty, A.; Thomas, A. Deciphering the Glycan Preference of Bacterial Lectins by  
23 Glycan Array and Molecular Docking with Validation by Microcalorimetry and  
24 Crystallography. *PLoS One* **2013**, *8* (8), e71149.
- 25 (29) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.;  
26 Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher,  
27 T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.;  
28 Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.;  
29 Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding  
30 Potency in Prospective Drug Discovery by Way of a Modern Free Energy Calculation  
31 Protocol and Force Field. *J. Am. Chem. Soc.* **2015**.
- 32 (30) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.;  
33 Mackerell, A. D. Additive Empirical Force Field for Hexopyranose Monosaccharides. *J.*  
34 *Comput. Chem.* **2008**, *29* (15), 2543–2564.
- 35 (31) Lins, R. D.; Hünenberger, P. H. A New GROMOS Force Field for Hexopyranose-Based  
36 Carbohydrates. *J. Comput. Chem.* **2005**, *26* (13), 1400–1412.
- 37 (32) Kony, D.; Damm, W.; Stoll, S.; Van Gunsteren, W. F. An Improved OPLS-AA Force  
38 Field for Carbohydrates. *J. Comput. Chem.* **2002**, *23* (15), 1416–1429.
- 39 (33) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C.  
40 R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field.  
41 Carbohydrates. *J. Comput. Chem.* **2008**, *29* (4), 622–655.
- 42 (34) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and  
43 Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.
- 44 (35) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond  
45 Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Modell.* **2006**,  
46 *25* (2), 247–260.
- 47 (36) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-  
48 Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21* (2),  
49 132–146.
- 50 (37) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic  
51 Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.*  
52 **2002**, *23* (16), 1623–1641.
- 53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 (38) Rangarajan, E. S.; Proteau, A.; Cui, Q.; Logan, S. M.; Potetinova, Z.; Whitfield, D.;  
4 Purisima, E. O.; Cygler, M.; Matte, A.; Sulea, T.; Schoenhofen, I. C. Structural and  
5 Functional Analysis of Campylobacter Jejuni PseG: A Udp-Sugar Hydrolase from the  
6 Pseudaminic Acid Biosynthetic Pathway. *J. Biol. Chem.* **2009**, *284* (31), 20989–21000.
- 7 (39) Hendrickx, P. M. S.; Corzana, F.; Depraetere, S.; Tourwé, D. A.; Augustyns, K.;  
8 Martins, J. C. The Use of Time-Averaged 3JHH Restrained Molecular Dynamics (tar-  
9 MD) Simulations for the Conformational Analysis of Five-Membered Ring Systems:  
10 Methodology and Applications. *J. Comput. Chem.* **2010**, *31* (3), 561–572.
- 11 (40) Cruz, L.; Brás, N. F.; Teixeira, N.; Mateus, N.; Ramos, M. J.; Dangles, O.; De Freitas,  
12 V. Vinylcatechin Dimers Are Much Better Copigments for Anthocyanins than Catechin  
13 Dimer Procyanidin B3. *J. Agric. Food Chem.* **2010**, *58* (5), 3159–3166.
- 14 (41) Sommer, R.; Exner, T. E.; Titz, A. A Biophysical Study with Carbohydrate Derivatives  
15 Explains the Molecular Basis of Monosaccharide Selectivity of the Pseudomonas  
16 Aeruginosa Lectin LecB. *PloS One* **2014**, *9* (11), e112822.
- 17 (42) Ernst, B.; Magnani, J. L. From Carbohydrate Leads to Glycomimetic Drugs. *Nat. Rev.*  
18 *Drug Discov.* **2009**, *8* (8), 661–677.
- 19 (43) Campbell, L. K.; Baker, D. E.; Campbell, R. K. Miglitol: Assessment of Its Role in the  
20 Treatment of Patients with Diabetes Mellitus. *Ann. Pharmacother.* **2000**, *34* (11), 1291–  
21 1301.
- 22 (44) Chen, X.; Zheng, Y.; Shen, Y. Voglibose (Basen, AO-128), One of the Most Important  
23 Alpha-Glucosidase Inhibitors. *Curr. Med. Chem.* **2006**, *13* (1), 109–116.
- 24 (45) Weinreb, N. J.; Barranger, J. A.; Charrow, J.; Grabowski, G. A.; Mankin, H. J.; Mistry,  
25 P. Guidance on the Use of Miglustat for Treating Patients with Type 1 Gaucher Disease.  
26 *Am. J. Hematol.* **2005**, *80* (3), 223–229.
- 27 (46) Zacharias, M.; Straatsma, T. P.; McCammon, J. A.; Quioco, F. A. Inversion of  
28 Receptor Binding Preferences by Mutagenesis: Free Energy Thermodynamic Integration  
29 Studies on Sugar Binding to L-Arabinose Binding Proteins. *Biochemistry (Mosc.)* **1993**,  
30 *32* (29), 7428–7434.
- 31 (47) Bucher, D.; Grant, B. J.; McCammon, J. A. Induced Fit or Conformational Selection?  
32 The Role of the Semi-Closed State in the Maltose Binding Protein. *Biochemistry*  
33 *(Mosc.)* **2011**, *50* (48), 10530–10539.
- 34 (48) Koppisetty, C. A. K.; Frank, M.; Lyubartsev, A. P.; Nyholm, P.-G. Binding Energy  
35 Calculations for Hevein-Carbohydrate Interactions Using Expanded Ensemble  
36 Molecular Dynamics Simulations. *J. Comput. Aided Mol. Des.* **2015**, *29* (1), 13–21.
- 37 (49) Kostlánová, N.; Mitchell, E. P.; Lortat-Jacob, H.; Oscarson, S.; Lahmann, M.; Gilboa-  
38 Garber, N.; Chambat, G.; Wimmerova, M.; Imberty, A. The Fucose-Binding Lectin  
39 from *Ralstonia Solanacearum*. A New Type of Beta-Propeller Architecture Formed by  
40 Oligomerization and Interacting with Fucoside, Fucosyllactose, and Plant Xyloglucan.  
41 *J. Biol. Chem.* **2005**, *280* (30), 27839–27849.
- 42 (50) Schell, M. A. Control of Virulence and Pathogenecity Genes of *Ralstonia*  
43 *Solanacearum* by an Elaborate Sensory Network. *Annu. Rev. Phytopathol.* **2000**, *38*,  
44 263–292.
- 45 (51) Woods, R. J.; and coworkers. GLYCAM Web <http://glycam.org/> (accessed Aug 1,  
46 2014).
- 47 (52) Autieri, E.; Segal, M.; Pederiva, F.; Guella, G. Puckering Free Energy of Pyranoses: A  
48 NMR and Metadynamics-Umbrella Sampling Investigation. *J. Chem. Phys.* **2010**, *133*  
49 (9), 095104.
- 50 (53) Ardèvol, A.; Biarnés, X.; Planas, A.; Rovira, C. The Conformational Free-Energy  
51 Landscape of B-D-Mannopyranose: Evidence for a 1S5 → B2,5 → OS2 Catalytic  
52 Itinerary in B-Mannosidases. *J. Am. Chem. Soc.* **2010**, *132* (45), 16058–16065.
- 53  
54  
55  
56  
57  
58  
59  
60

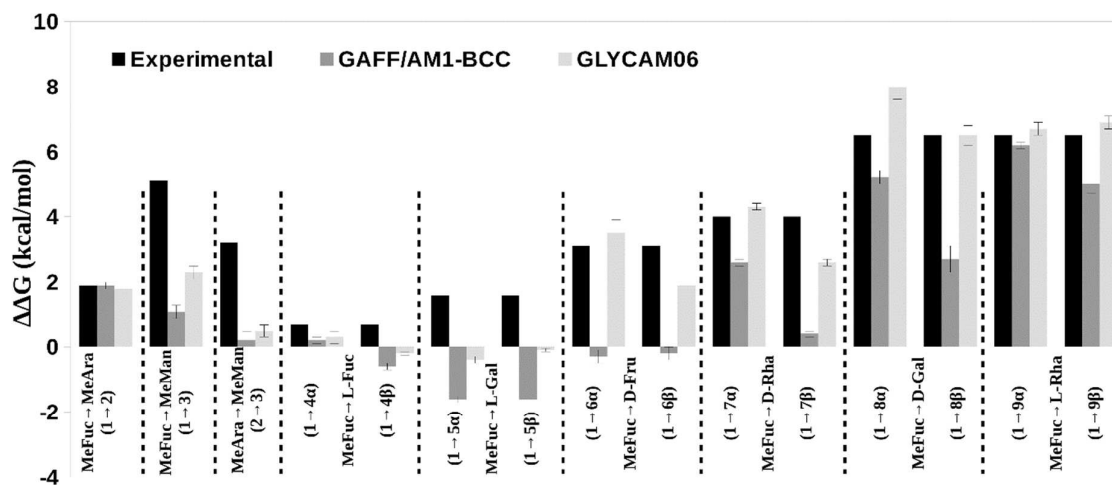
- 1  
2  
3 (54) Loeffler, H. H.; Woods, C. J.; Michel, J. FESetup 1.0.  
4 <http://ccpforge.cse.rl.ac.uk/gf/project/ccpbiosim/>. (accessed on 02/15/2015)
- 5 (55) Case, D.A.; Babin, V.; Berryman, J.T.; Betz, R.M.; Cai, Q.; Cerutti, D.S.; Cheatham,  
6 III, T.E.; Darden, T.A.; Duke, R.E.; Gohlke, H.; Goetz, A.W.; Gusarov, S.; Homeyer,  
7 N.; Janowski, P.; Kaus, J.; Kolossváry, Kovalenko, I. A.; Lee, T.S.; LeGrand, S.;  
8 Luchko, T.; Luo, R.; Madej, B.; Merz, K.M.; Paesani, F.; Roe, D.R.; Roitberg, A.;  
9 Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C.L.; Smith, W.; Swails, J.;  
10 Walker R.C; Wang, J.; Wolf, R.M.; Wu, X. and Kollman P. A. AMBER 14, University  
11 of California, San Francisco, 2014.
- 12  
13  
14 (56) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C.  
15 Comparison of Multiple Amber Force Fields and Development of Improved Protein  
16 Backbone Parameters. *Proteins* **2006**, *65* (3), 712–725.
- 17 (57) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-  
18 Thermodynamic Basis for Computation of Binding Affinities: A Critical Review.  
19 *Biophys. J.* **1997**, *72* (3), 1047–1069.
- 20 (58) Mezei, M. The Finite Difference Thermodynamic Integration, Tested on Calculating the  
21 Hydration Free Energy Difference between Acetone and Dimethylamine in Water. *J.*  
22 *Chem. Phys.* **1987**, *86* (12), 7084–7088.
- 23 (59) Shyu, C.; Ytreberg, F. M. Reducing the Bias and Uncertainty of Free Energy Estimates  
24 by Using Regression to Fit Thermodynamic Integration Data. *J. Comput. Chem.* **2009**,  
25 *30* (14), 2297–2304.
- 26 (60) Wang, J.; Dixon, R.; Kollman, P. A. Ranking Ligand Binding Affinities with Avidin: A  
27 Molecular Dynamics-Based Interaction Energy Study. *Proteins* **1999**, *34* (1), 69–81.
- 28 (61) Michel, J.; Verdonk, M. L.; Essex, J. W. Protein–Ligand Complexes: Computation of  
29 the Relative Free Energy of Different Scaffolds and Binding Modes. *J. Chem. Theory*  
30 *Comput.* **2007**, *3* (5), 1645–1655.
- 31 (62) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure And/or  
32 Temperature. *J. Chem. Phys.* **1980**, *72* (4), 2384.
- 33 (63) Tironi, I. G.; Sperb, R.; Smith, P. E.; Gunsteren, W. F. van. A Generalized Reaction  
34 Field Method for Molecular Dynamics Simulations. *J. Chem. Phys.* **1995**, *102* (13),  
35 5451–5459.
- 36 (64) Fennell, C. J.; Gezelter, J. D. Is the Ewald Summation Still Necessary? Pairwise  
37 Alternatives to the Accepted Standard for Long-Range Electrostatics. *J. Chem. Phys.*  
38 **2006**, *124* (23), 234104.
- 39 (65) Woods, C. J.; Calabro, G.; Michel, J. Sire Molecular Simulation Framework, Revision  
40 2702, 2014, <http://siremol.org/Sire/Home.html>; (accessed on 02/15/2015)
- 41 (66) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.;  
42 Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich,  
43 T.; Klein, C.; Shirts, M. R.; Pande, V. S. OpenMM 4: A Reusable, Extensible, Hardware  
44 Independent Library for High Performance Molecular Simulation. *J. Chem. Theory*  
45 *Comput.* **2012**, *9* (1), 461–469.
- 46 (67) Brown, S. P.; Muchmore, S. W.; Hajduk, P. J. Healthy Skepticism: Assessing Realistic  
47 Model Performance. *Drug Discovery. Today* **2009**, *14* (7–8), 420–427.
- 48 (68) Mikulskis, P.; Genheden, S.; Ryde, U. A Large-Scale Test of Free-Energy Simulation  
49 Estimates of Protein–Ligand Binding Affinities. *J. Chem. Inf. Model.* **2014**, *54* (10),  
50 2794–2806.
- 51 (69) Pearlman, D. A.; Charifson, P. S. Are Free Energy Calculations Useful in Practice? A  
52 Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System. *J.*  
53 *Med. Chem.* **2001**, *44* (21), 3417–3423.
- 54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 (70) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Effects of Water Placement  
4 on Predictions of Binding Affinities for p38 $\alpha$  MAP Kinase Inhibitors. *J. Chem. Theory*  
5 *Comput.* **2010**, *6* (12), 3850–3856.  
6  
7 (71) Morozov, A. V.; Misura, K. M. S.; Tsemekhman, K.; Baker, D. Comparison of  
8 Quantum Mechanics and Molecular Mechanics Dimerization Energy Landscapes for  
9 Pairs of Ring-Containing Amino Acids in Proteins. *J. Phys. Chem. B* **2004**, *108* (24),  
10 8489–8496.  
11 (72) Juaristi, E. *Conformational Behavior of Six-Membered Rings: Analysis, Dynamics and*  
12 *Stereoelectronic Effects*; Wiley-VCH Verlag GmbH: New York, 1995.  
13 (73) Homola, J.; Yee, S. S.; Gauglitz, G. Surface Plasmon Resonance Sensors: Review. *Sens.*  
14 *Actuators B Chem.* **1999**, *54* (1–2), 3–15.  
15 (74) Kozmon, S.; Matuška, R.; Spiwok, V.; Koča, J. Three-Dimensional Potential Energy  
16 Surface of Selected Carbohydrates' CH/ $\pi$  Dispersion Interactions Calculated by High-  
17 Level Quantum Mechanical Methods. *Chem. – Eur. J.* **2011**, *17* (20), 5680–5690.  
18 (75) Wimmerova, M.; Kozmon, S.; Nečasová, I.; Mishra, S. K.; Komárek, J.; Koca, J.  
19 Stacking Interactions between Carbohydrate and Protein Quantified by Combination of  
20 Theoretical and Experimental Methods. *PLoS ONE* **2012**, *7* (10), e46032.  
21 (76) Fennell, C. J.; Wymer, K. L.; Mobley, D. L. A Fixed-Charge Model for Alcohol  
22 Polarization in the Condensed Phase, and Its Role in Small Molecule Hydration. *J.*  
23 *Phys. Chem. B* **2014**, *118* (24), 6438–6446.  
24 (77) Pathiaseril, A.; Woods, R. J. Relative Energies of Binding for Antibody-Carbohydrate-  
25 Antigen Complexes Computed from Free-Energy Simulations. *J. Am. Chem. Soc.* **2000**,  
26 *122* (2), 331–338.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

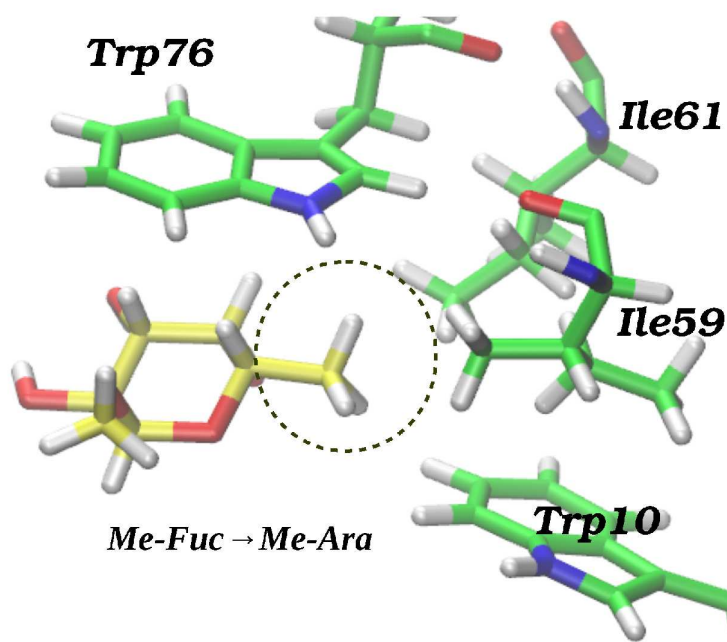


**Figure 1.** Graphical representations of perturbations of Me- $\alpha$ -L-Fuc (1) into different labeled monosaccharides. Red atoms that differ from topologically equivalent atoms in Me- $\alpha$ -L-Fuc. Experimental dissociation constants ( $K_d$ ) of each monosaccharide for RSL lectin are in micromole/L. R1 and R2 for  $\alpha$ - &  $\beta$ - anomers of L-Fuc (4), L-Gal (5), D-Rha (7), D-Gal (8) and L-Rha (9) are specified in top legend and for D-Fru (6) in the lower legend.

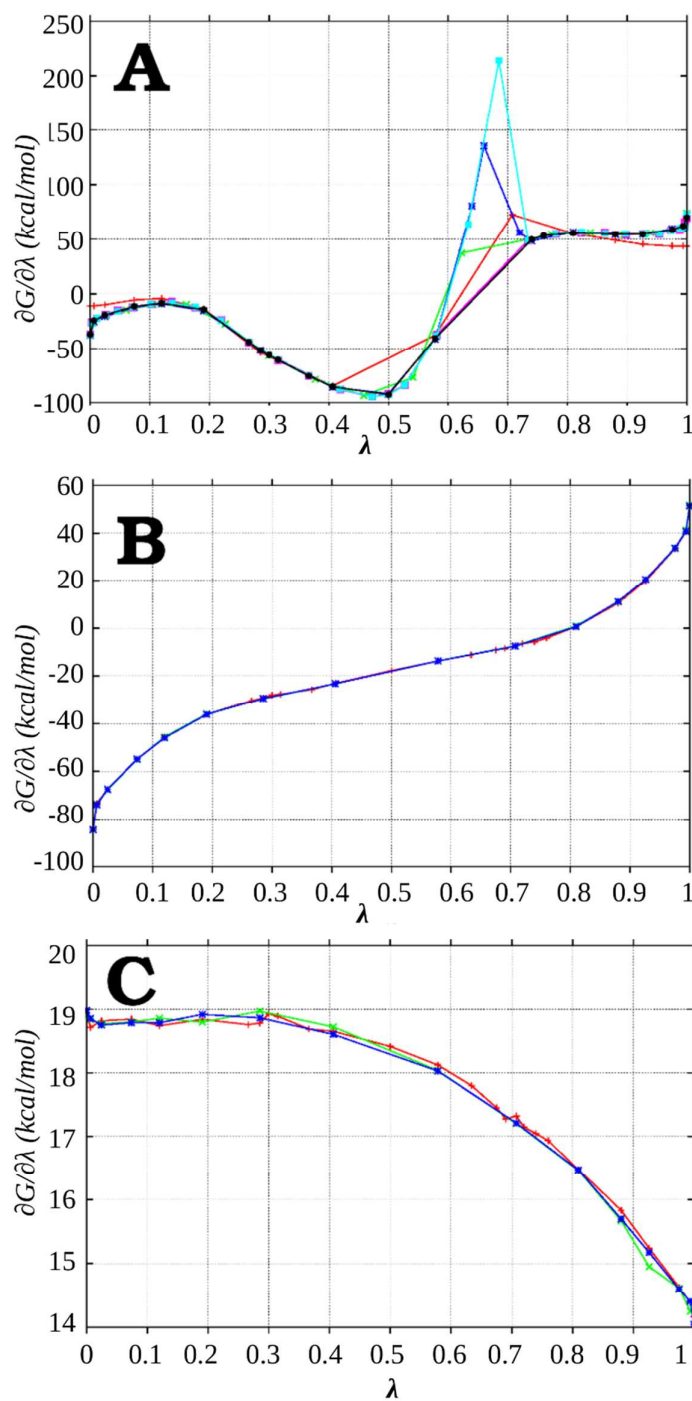




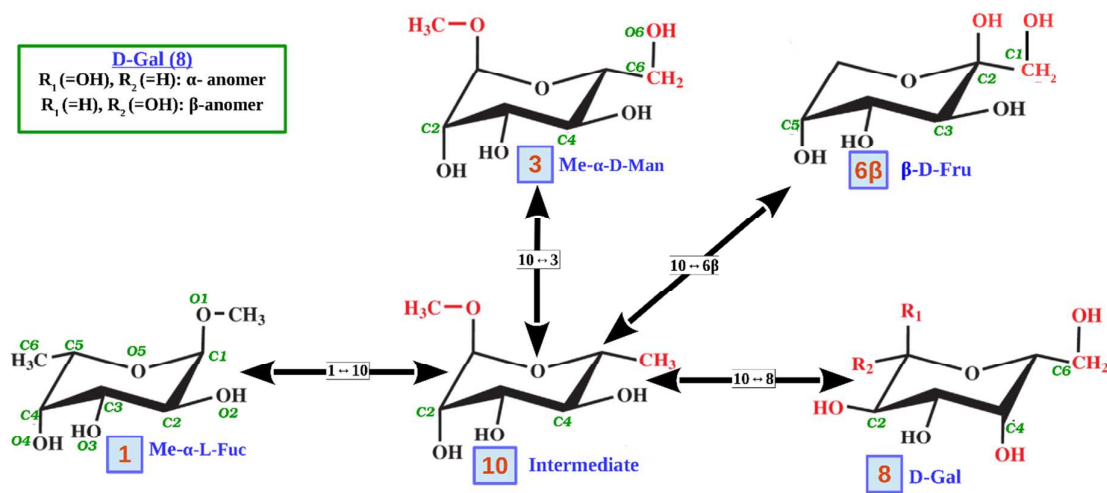
**Figure 2.** Experimental ( $\Delta\Delta G_{b,exp}(X \rightarrow Y)$ ) and calculated ( $\Delta\Delta G_{b,calc}(X \rightarrow Y)$ ) relative binding free-energies using the GAFF1.7/AM1-BCC (GAFF) and GLYCAM06j force fields. Estimated uncertainties ( $\text{err}(\Delta\Delta G_{b,calc}(X \rightarrow Y))$ ) are shown as error bars. The  $\text{err}(\Delta\Delta G_{b,exp}(X \rightarrow Y))$  values are assumed to be  $0.4 \text{ kcal.mol}^{-1}$ .<sup>67,68</sup>



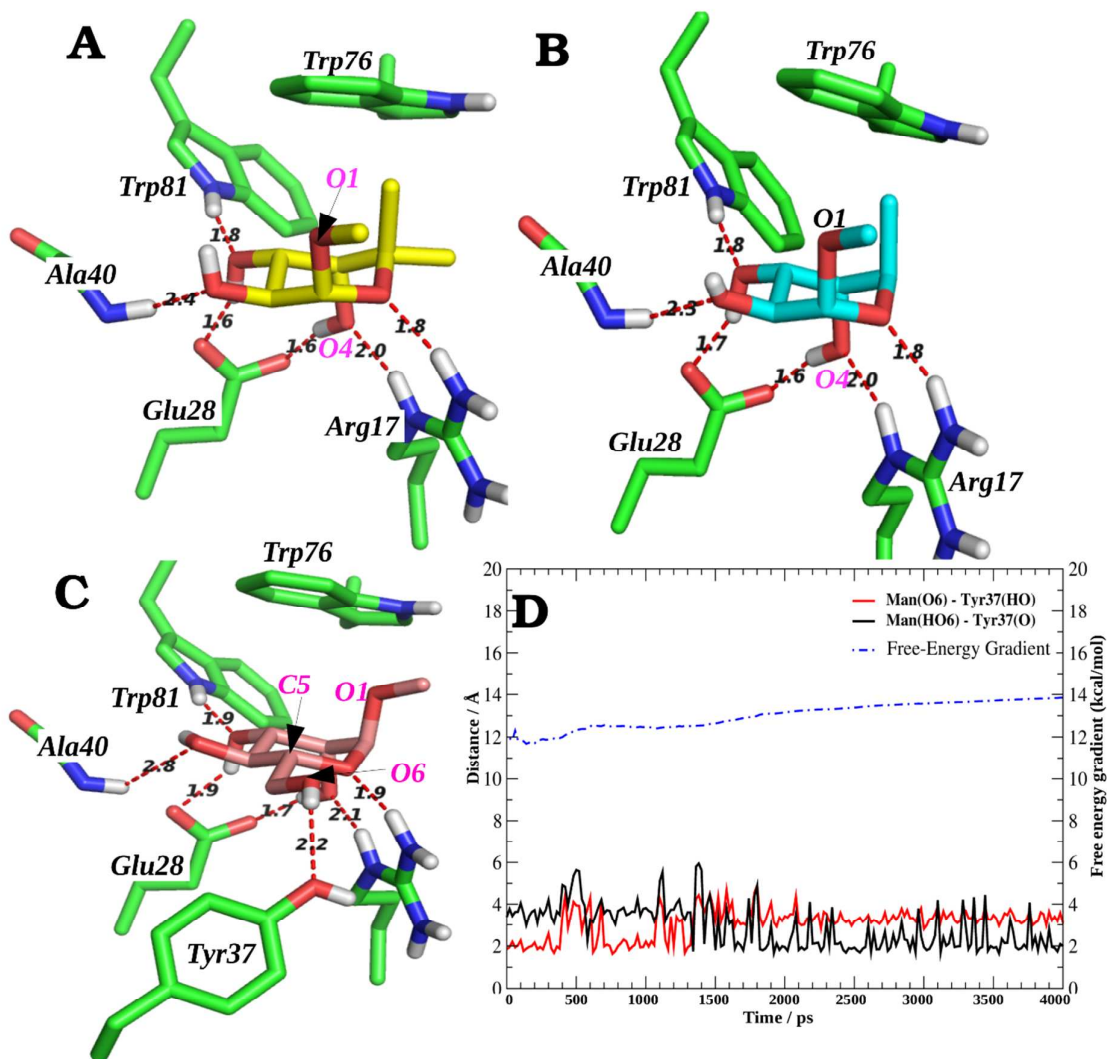
**Figure 3.** The Me- $\alpha$ -L-Fuc to Me- $\beta$ -D-Ara perturbation (1 $\rightarrow$ 2) in the complex with RSL. The methyl group (circle) is replaced by a hydrogen atom within the hydrophobic patch of the protein.



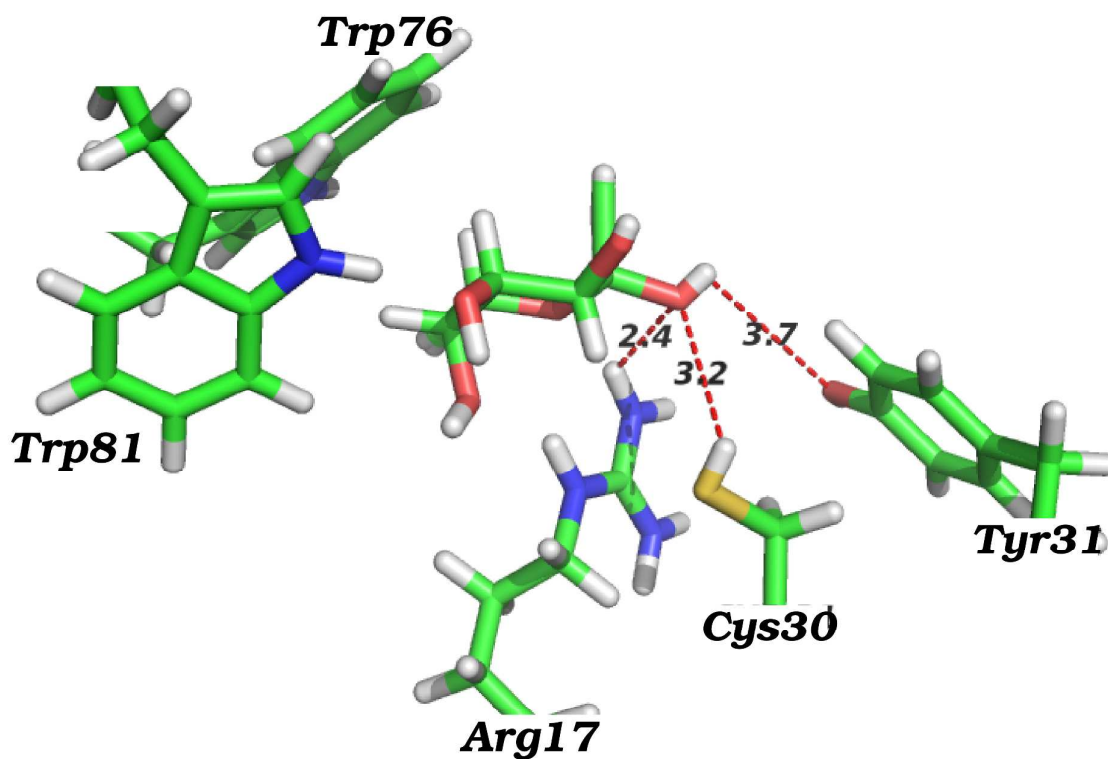
**Figure 4.** The free-energy gradients from Me- $\alpha$ -L-Fuc $\rightarrow$ Me- $\alpha$ -D-Man (**1** $\rightarrow$ **3**) perturbation using GLYCAM. (A) Multiple setups of **1** $\rightarrow$ **3** perturbation in single step. (B&C) Replicates of **1** $\rightarrow$ **10** and **10** $\rightarrow$ **3** perturbation, respectively. Multiple replicates are shown in different color.



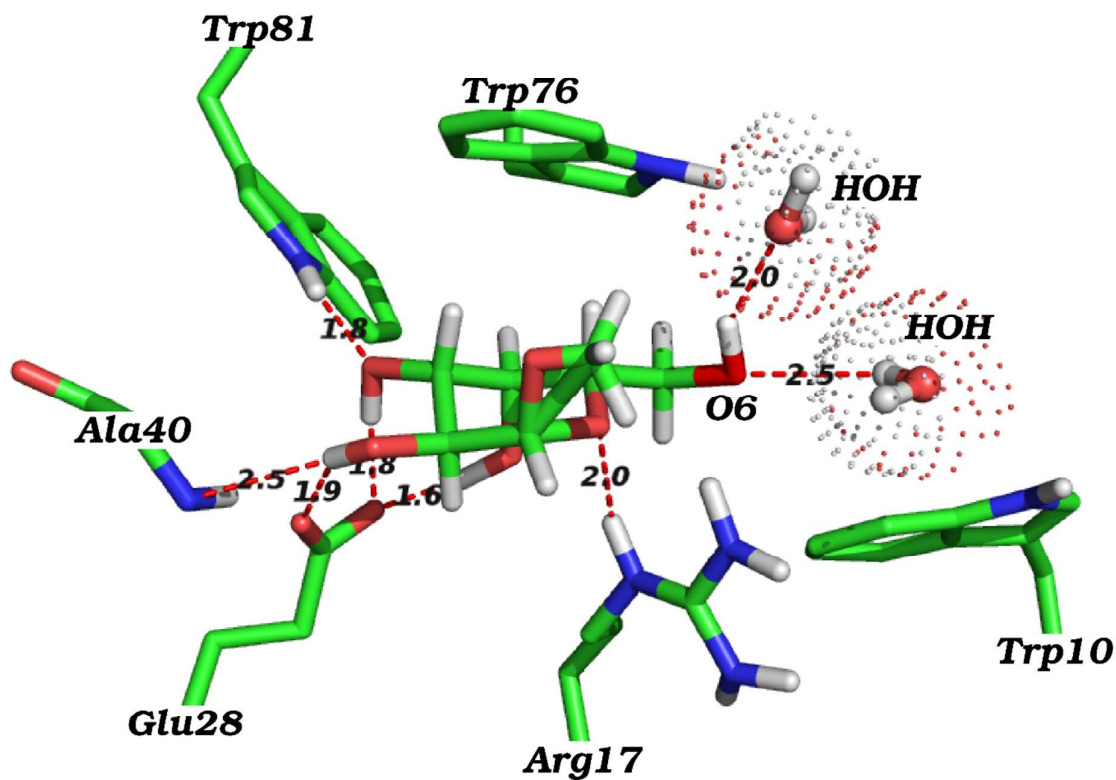
**Figure 5.** Graphical representations of perturbing Me- $\alpha$ -L-Fuc (**1**) into labeled monosaccharides via an intermediate structure (**10**). Red atoms/groups are the ones that differ from **1**.  $R_1$  and  $R_2$  for  $\alpha$ - &  $\beta$ - anomers of D-Gal (**8**) are specified in the legend.



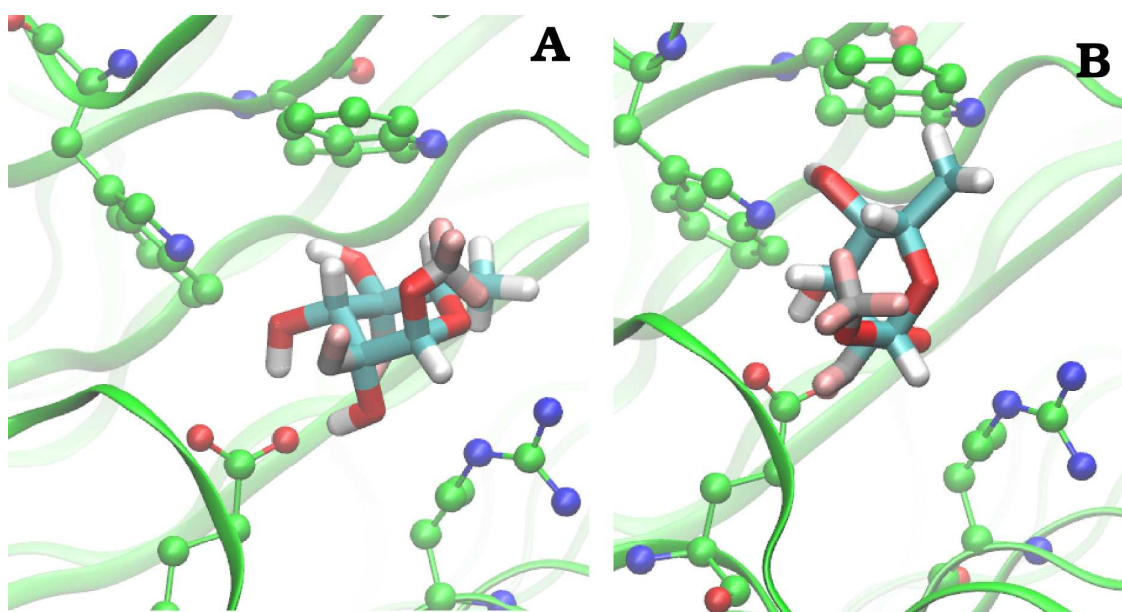
**Figure 6.** Average structures from MD simulation of saccharides in the protein bound state. (A) Me- $\alpha$ -L-Fucoside (1), (B) Me- $\beta$ -D-Arabinoside (2), and (C) Me- $\alpha$ -D-Mannoside (3). All the distances shown in red dotted lines are in angstrom. (D) Time series of the hydrogen bond distance between hydrogen of O6 hydroxyl in 3 & oxygen of hydroxyl in Tyr37 (in black) and oxygen of O6 hydroxyl in 3 & hydrogen of hydroxyl in Tyr37 (in red) with corresponding free energy gradient.



**Figure 7.** Average structure from Me- $\alpha$ -L-Fucose to  $\beta$ -L-Fucose perturbation ( $1 \rightarrow 4\beta$ ) in the complex with RSL. All the distances from O1 hydroxyl shown in red dotted lines are in Angstrom.

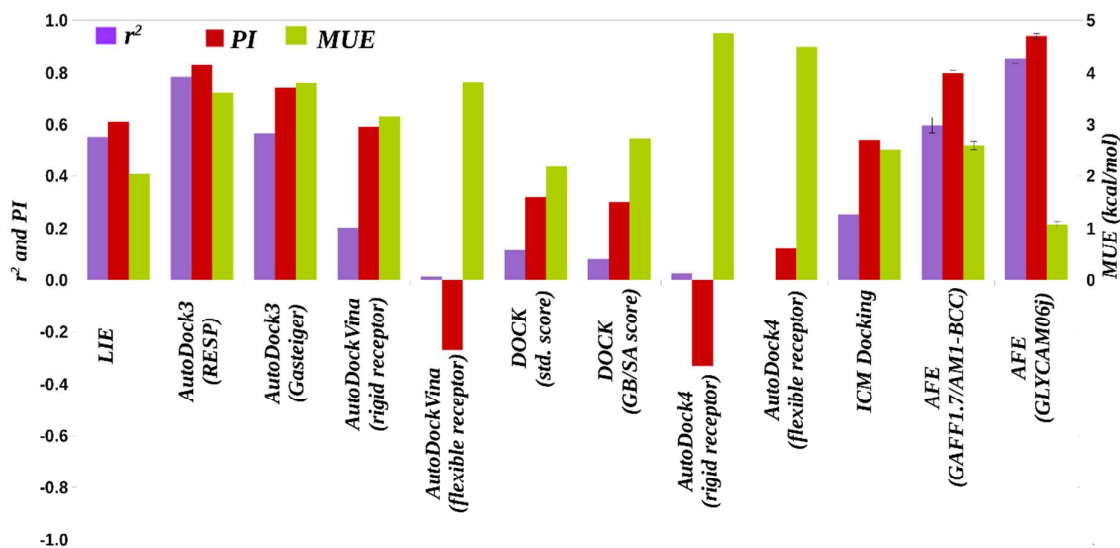


**Figure 8.** Average structure from Me- $\alpha$ -L-Fuc to  $\beta$ -L-Gal perturbation ( $1 \rightarrow 5\beta$ ) in the complex with RSL. All the distances from O6 hydroxyl shown in red dotted lines are in Angstrom. There are two water molecules interacting closely with O6 hydroxyl in  $5\beta$  are shown with dotted van der Waals surface.



**Figure 9.** Two different binding conformations of L-Rha (**9**) during the simulations (**9**→**1**). (A) Initial and well defined binding mode of  $\alpha$ -L-Rha (**9a**) and (B) poorly binding mode from simulation (**9b**).





**Figure 10.** Evaluation of twelve different molecular modelling protocols to predict binding affinities of the dataset of RSL ligands. For each protocol the first and second histograms indicate the determination coefficient  $r^2$  and predictive index PI, whereas the third histogram indicates the mean unsigned error MUE in kcal.mol<sup>-1</sup>. For the last metric, the null hypothesis would give a MUE of 3.55 kcal.mol<sup>-1</sup> on this dataset. Where available, error bars are indicated. Details on the docking and LIE calculations are provided in the SI.

## Graphical TOC

