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

*Sushil K. Mishra,† Gaetano Calabró,§ Hannes H. Loeffler,‡ Julien Michel,§* Jaroslav Koča †**

† Central European Institute of Technology (CEITEC), and National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice-5, 625 00 Brno, Czech Republic. § EaStCHEM School of Chemistry, Joseph Black Building, King's Buildings, Edinburgh EH9 3JJ, UK.‡ 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 alchemical predictions of binding free-energies with the GAFF1.7/AM1-BCC and GLYCAM06j force fields for modelling protein-carbohydrate complexes. Mean unsigned errors of 1.1±0.06 (GLYCAM06j) and 2.6±0.08 (GAFF1.7/AM1-BCC) kcal•mol$^{-1}$ are achieved for a large dataset of monosaccharide ligands for *Ralstonia solanacearum* lectin (RSL). The level of accuracy provided by GLYCAM06j is sufficient to discriminate potent, moderate and weak binders, a goal that has been difficult to achieve through other scoring approaches. Accordingly the protocols presented here could find useful applications in carbohydrate-based drug and vaccine developments.

## 1. Introduction

The problem of computing the binding free-energy of a ligand for a receptor is a long-standing challenge for computational chemistry.[1–3] Ever since the very first alchemical free-energy (AFE) calculations where reported for ligand binding processes,[4] numerous studies have focused on the binding energetics of organic molecules to proteins.[3,5,6] Despite successes in guiding the design of organic molecules as protein ligands[1,7–9] applications to other classes of biomolecular interactions such as protein-DNA, protein-lipids and protein-carbohydrate complexes have been less explored.[2,10,11] This work is concerned with the validation of parameter sets for accurate modelling of protein-carbohydrate recognition with the aid of alchemical free-energy methods.

Protein-carbohydrate complexes pose specific challenges for molecular modeling due to the large number of hydroxyl groups in the ligands, weak binding affinities, anomerization, ring flexibility, CH...$\pi$ interactions, and frequent role of water[12] and/or ions[13] in receptor binding sites. Progress is necessary owing to the significant role of protein-carbohydrate interactions in biology. A few notable examples includes biological processes like, cell adhesion, differentiation, and metastasis.[14,15] Protein-carbohydrate interactions are also important in medical sciences, e.g. alterations of cell surface glycosylation pattern is linked to the development and progression of specific diseases like cancer.[16] Among protein-carbohydrate complexes, lectin-carbohydrate complexes are of immense interest because lectins have the ability to distinguish between minuscule differences in sugar structures and can be used to detect specific carbohydrate patterns.[17] Thus, understanding the structural and energetic aspects of the lectin-carbohydrate complexes is essential for the elucidation of carbohydrate recognition principles, which should ultimately aid the design of carbohydrate-based pharmaceuticals.[14,18]

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

AFE calculation protocols (e.g. free energy perturbation (FEP), thermodynamic integration (TI)) are attractive alternatives owing to a more rigorous statistical thermodynamics framework.[3,29] However, the reliability of current force fields (FFs) for AFE calculations of protein-carbohydrate complexes is still unclear. The carbohydrate FFs that can be used for simulation in the biomolecular context are mainly CHARMM[30], GROMOS-45A4[31], OPLS-AA-SIE[32] and GLYCAM[33]. Among them GLYCAM is steadily growing and the most cited

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

## 2. Theory and Methods

### 2.1 Preparation of Molecular Models

***Protein setup****:* *Ralstonia solanacearum* lectin (RSL[49]) is a protein isolated from the Gram-negative bacterial pathogen *Ralstonia solanacearum* that causes lethal wilt disease in many agricultural crops all over the world, leading to massive losses in the agricultural industry.[50] RSL is a six-bladed β- propeller trimeric structure, with 90 amino acid residues in each monomeric chain. Each RSL monomer unit contains one fucose binding site located between the two β-sheet blades called the intramonomeric binding site, and the other is formed at the interface between the neighboring monomers called the intermonomeric binding site. Thus, there are a total of six symmetrically arranged binding sites reported in the crystal structure. Isothermal Titration Calorimetry (ITC) has suggested that intermonomeric and intramonomeric binding sites are indistinguishable.[49] The calculated free-energy of Me-α-L-Fucoside (MeFuc) in all six binding sites was statistically equivalent in LIE calculations reported elsewhere.[27] The initial coordinates of RSL bound to MeFuc were obtained from the X-ray crystal structure (PDB ID: 2BT9).[49] A couple of perturbations (**1**↔**2** and **1**↔**4**) were performed in all six RSL binding sites (Table S1). Refer to Fig. 1 for number assigned to each ligand. The other perturbations were performed in the intramonomeric binding site of chain A **(site S1)** only, with the other five binding sites kept empty.

***Protein-Ligand complex setup:*** A full range of monosaccharides spanning from binders, low-affinity binders to non-binders were selected (Fig. 1). The experimental dissociation constants of these monosaccharides have been previously measured using an SPR assay.[23,49] The 3D coordinates of the ligands were modeled using the GLYCAM Carbohydrate Builder webserver.[51] Since there is no evidence that RSL recognizes monosaccharides in furanose form, pyranose form of all the ligands were selected. The starting structures of the RSL-saccharide complexes for the other monosaccharides were prepared manually by superposition with the ring atoms of MeFuc (**1**), keeping the orientation of O2, O3, and O4 hydroxyls unchanged where possible. As the potential energy barrier for ring flips in pyranoses can be ca. 3-5 kcal•mol$^{-1}$ or greater,[52,53] all monosaccharides were kept in their most favorable chair conformation. Input files for alchemical free-energy calculations were prepared with the software FESetup[54] that uses AmberTools14[55] for ligand parameterization. All initial structures were solvated in a rectangular box of TIP3P water molecules extending 12 Å away from the edges of the solute(s) using *tleap*. The total charge of each system was zero and no ions were required to neutralize any of the systems. The protein was described with the Amber ff12SB[56] force field, and saccharides with either the GAFF[34,35] force field (version 1.7) with AM1-BCC[36,37] charges (as computed by antechamber from AmberTools14) or the GLYCAM[33] force field version 06j (GLYCAM06j). The simulation systems were equilibrated by firstly performing 3000 steps of energy minimization to relax unfavorable conformations, followed by a 1 ns 300K NVT simulation with harmonic positional restraints of force constant 5 kcal•mol$^{-1}$•Å$^{-2}$ on all the non-solvent atoms. Finally, an unrestrained 3 ns NPT simulation was performed to equilibrate solvent density. The final snapshot was used as input for subsequent free-energy calculations.

### 2.2 Alchemical free-energy simulations

Neglecting contribution from changes in pressure-volume terms[57], relative binding free-energies (eq 1) were computed as the difference in the free-energy changes for transforming monosaccharide X into Y in the RSL binding site ($\Delta G_P(X \rightarrow Y)$) and in aqueous solution($\Delta G_w(X \rightarrow Y)$):

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

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

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

Where $\lambda$ is a coupling parameter that allows smooth transformation of the potential energy function corresponding to the starting state $X$ ($\lambda$=0) and final state $Y$ ($\lambda$=1). The finite difference thermodynamic integration approach was firstly used to evaluate free-energy gradients at several values of $\lambda$ between 0 and 1.[58] The integral in eq 2 was then numerically approximated by using polynomial regression[59] and setting the polynomial order to seven. Unless stated separately in the text, the free-energy gradients were calculated at 16 non-equidistant $\lambda$ values (0.0, 0.00616, 0.02447, 0.07368, 0.11980, 0.19045, 0.28534, 0.40631, 0.57822, 0.70755, 0.80955, 0.88020, 0.92632, 0.97553, 0.99384, and 1.0).[59] A 4 ns NPT simulation at each $\lambda$ value was performed to collect free-energy gradients. To test for convergence, longer simulations of 10 ns per window were run, or a $\lambda$ schedule with 27 points was applied in selected cases. Additional points were added near noisy parts of the gradient when needed, e.g. **1→3** and **2→3** perturbations. Before collecting free-energy gradients, the systems were further energy minimized (1000 steps) and then equilibrated for 100 ps at the chosen value of $\lambda$. To avoid numerical instabilities, a soft-core[60] potential energy function similar to Michel et al. was used throughout.[61] Free-energy gradients were collected every 200 fs. The first 500 ps of every simulation was discarded to allow for re-equilibration.

A velocity-Verlet integrator was used with a timestep of 2 fs. All simulations were performed in the NPT ensemble. Temperature control was achieved with an Andersen thermostat and a coupling constant of 10 ps$^{-1}$.[62] Pressure control was achieved by attempting isotropic box edge scaling Monte Carlo moves every 25 time-steps. Periodic boundary conditions are used with a 10Å atom-based cutoff distance for the non-bonded interactions. All simulations were performed using an atom-based cutoff of 10 Å with Barker Watts reaction field with dielectric constant set to 78.3.[63] The methodology used here to handle long range electrostatic interactions differs from the parameters used in other studies performed with the Amber ff12SB force field. However, this was deemed acceptable as Fennel and Gezelter have reported that atom-based reaction-field treatments yield energetics and dynamics that reproduce well Particle-Mesh Ewald.[64] Further, in this work the same methodology was used consistently to compare GAFF and GLYCAM results. Production 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[65] and OpenMM-6.0[66]). 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:

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

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:

$$\mathrm{err}\left(\Delta\Delta G_{b,calc}(X \rightarrow Y)\right) = \frac{s}{\sqrt{n}} \qquad (4)$$

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 ($K_d$)[23,49] measured by Surface Plasmon Resonance (SPR) assay at 298.15 K and standard reference concentration ($C^0$) of 1 mol.L$^{-1}$ using eq 5:

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

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$^{-1}$ in $\Delta\Delta G_{b,exp}(X \rightarrow Y)$ was assumed as done by Brown et al. and Mikulskis et al.[67,68]

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.[69] As done elsewhere[70], 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 ±1σ 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-α-L-Fucoside (**1**), Me-β-D-Arabinoside (**2**) and Me-α-D-Mannoside (**3**), and are -8.6, -6.7 and -3.5 kcal•mol$^{-1}$, respectively. Accordingly, a number of relative binding free-energy calculations for MeFuc→MeAra (**1**→**2**), MeFuc→MeMan (**1**→**3**) and MeAra→MeMan (**2**→**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→2**, the $\Delta\Delta G_{b,calc}(\mathbf{1} \rightarrow \mathbf{2})$ values from both GAFF (1.9±0.1 kcal•mol$^{-1}$) and GLYCAM (1.8±0.1 kcal•mol$^{-1}$) are in an excellent agreement with $\Delta\Delta G_{b,exp}(\mathbf{1} \rightarrow \mathbf{2})$ (1.9 kcal•mol$^{-1}$). In **1**→**2**, the equatorial methyl group at position C5 in **1** is replaced by a hydrogen in **2**. This C6-methyl projects into a hydrophobic region lined by the side chains of the residues Ile59, Ile61 and Trp10 of RSL (Fig. 3). The change in the binding free-energy is particularly unfavorable in this case because this scaffold modification results in a loss of hydrophobic interactions with the protein environment.

In the perturbations **1**→**3**, and **2**→**3** larger groups of atoms need to be perturbed. The ring carbon atoms C1, C2, C4 and C5 in **1** and **2** have been mapped to C5, C4, C3 and C2 in **3**, respectively, such that the orientation of the O2, O3 and O4 hydroxyls remain unchanged. However, the axial -$OCH_3$ group (methoxy) and the equatorial hydrogen of C1 in **1** are perturbed into a hydrogen and hydroxymethyl group in **3**, respectively. Additionally, the axial hydrogen of C5 in **1** is perturbed into a methoxy group, and the equatorial methyl at C5 in **1** is also perturbed into a hydrogen in **3** (Fig. 1). For these two calculations, we found serious 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 $\Delta G_P\ (\mathbf{2} \rightarrow \mathbf{3})$. Analysis of the free-energy gradients shows a considerable peak of the free-energy gradients at $\lambda$ ~0.7 for **1**→**3** and **2**→**3**. This is mirrored at $\lambda$ ~0.3 for a perturbation done through the reverse paths (**3**→**1** and **3**→**2**). The free-energy gradients have very large values within these $\lambda$ regions, and the resulting free-energy profile is noisy (Fig. 4A). Increasing the length of the simulation or the number of $\lambda$ points does not improve the precision of the results (Fig. S1 & S2). A careful investigation was undertaken to diagnose the problem. The perturbations were broken down into two sequential calculations involving an intermediate compound **10** so as to minimize the magnitude of the structural changes attempted in one step. Compound **1** and **2** were thus first perturbed into **10** where the equatorial hydrogen of C1 in **1** and **2** is perturbed into a methyl. In the second step, this methyl is then perturbed into the final hydroxymethyl in **3** (Fig. 5). A complication for the GLYCAM force field is that the intermediate structure **10** does not have parameters, and force field parameters were thus manually adapted by analogy from those used to describe **3**.

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

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

Figure 2 shows experimental and calculated change in the free-energy of binding of methylated monosaccharides **3** relative to **1** and **2** calculated by the two-step perturbation protocol using GAFF and GLYCAM force fields. The $\Delta\Delta G_{b,calc}\ (\mathbf{1} \rightarrow \mathbf{3})$ values using GAFF and GLYCAM are 1.1±0.2 and 2.3±0.2 kcal•mol$^{-1}$ respectively, and $\Delta\Delta G_{b,calc}\ (\mathbf{2} \rightarrow \mathbf{3})$ values are 0.2±0.3 and 0.5±0.2 kcal•mol$^{-1}$ respectively. While the two force fields follow a similar trend in values for $\Delta\Delta G_{b,calc}\ (\mathbf{1} \rightarrow \mathbf{3})$ and $\Delta\Delta G_{b,calc}\ (\mathbf{2} \rightarrow \mathbf{3})$, it is clear that the GLYCAM calculations provide better agreement with $\Delta\Delta G_{b,exp}\ (\mathbf{1} \rightarrow \mathbf{3})$ and $\Delta\Delta G_{b,exp}\ (\mathbf{2} \rightarrow \mathbf{3})$ (5.1 and

3.2 kcal•mol$^{-1}$ respectively). Relative binding free-energies for **1→3** and **2→3** are overestimated by ca. 2.5 kcal•mol$^{-1}$. The thermodynamic cycle closure error for **1→10→2→1** and **1→2→10→1** perturbations (Fig. S6) from GAFF is -0.3 kcal•mol$^{-1}$ and 1.4 kcal•mol$^{-1}$, respectively. While thermodynamic cycle closure error from GLYCAM is 0.3 kcal•mol$^{-1}$ in both cases. This indicates that the computed relative binding free-energies with this force field appear well converged and that deviations from experimental data may be attributed to inaccurate in force field parameters for **3** since the relative binding free-energy for **1→2** matches well with experimental data. Indeed, the methoxy group in **3** is placed below Trp76, which is expected to disrupt CH...π stacking that is observed in **1** and **2** (Fig. 6A-6C). Current classical force fields are of limited accuracy for the modeling of CH...π stacking interactions.[71] A close inspection of the GLYCAM simulations also shows that the hydroxymethyl group of C5 in **3** is projected outward from the binding site, and creates a hydrogen bond with the hydroxyl of Tyr37 (Fig. 6D). A similar behavior is seen for perturbation **2→3**.

The resulted cycle closure error for the thermodynamic cycle (shown in Fig. S6) for perturbation **1→2→3→1** is 0 and 1.0 kcal•mol$^{-1}$ for GLYCAM and GAFF simulations, respectively, indicating confidence in the GLYCAM computed free-energies. By contrast, the free-energies computed with the GAFF force field are not consistent. Turning now to the accuracy of the force fields, for carbohydrates, relatively less accurate energies from GAFF/AM1-BCC calculations are expected for of two reasons. Firstly, GAFF is parameterized to cover mostly small organic compounds, and has not been optimized for performance on hexopyranoses or carbohydrate-like structures.[34] Secondly, in asymmetric molecules, such as sugars, the possibility of hydroxyl and hydroxymethyl group rotation leads to ambiguity in selecting a single conformation for charge calculations. Thus, the intramolecular interactions or solution properties are often poorly reproduced, unless conformationally averaged charges are employed.[72]

Interestingly, both GAFF and GLYCAM systematically overestimate the binding affinity of **3** to RSL. Mishra et al. also made a similar observation in a LIE study of this system using the OPLS-AA 2005 force field, where the predicted absolute binding free-energy for **3** was overestimated by approximately -2.0 kcal•mol$^{-1}$.[27] The possibility that experimental artifacts have affected SPR measurements of the weak binding of compound **3** ($K_d$ ~2.5 mM) should not be ruled out.[73]

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

Perturbation of MeFuc into L-Fuc (**1→4**) is computationally demanding because L-Fuc can exist in both anomers in solution and the protein bound state. Thus, we decided to transform MeFuc into both α-L-Fuc (**4α**) and β-L-Fuc (**4β**). To close an additional thermodynamic cycle the **4α→4β** perturbation was also performed. The calculated and experimental changes in the binding free-energy of the **1→4** perturbations are presented in Fig 2.

The $\Delta\Delta G_{b,calc}(\mathbf{1} \rightarrow \mathbf{4\alpha})$ and $\Delta\Delta G_{b,calc}(\mathbf{1} \rightarrow \mathbf{4\beta})$ values using GLYCAM are 0.3±0.3 kcal•mol$^{-1}$ and -0.2±0.1 kcal•mol$^{-1}$, respectively, which is in good agreement with the experiment and well converged as shown by the cycle closure error of the thermodynamic cycle **1→4β→4α→1** that is close to zero. Lower $\Delta\Delta G_{b,calc}\ (\mathbf{1} \rightarrow \mathbf{4\beta})$ values using GAFF (-0.6±0.1 kcal•mol$^{-1}$) and GLYCAM (-0.2±0.1 kcal•mol$^{-1}$) compared to $\Delta\Delta G_{b,calc}\ (\mathbf{1} \rightarrow \mathbf{4\alpha})$ from GAFF (0.2±0.1 kcal•mol$^{-1}$) and GLYCAM (0.3±0.2 kcal•mol$^{-1}$) suggests that RSL will prefer to bind **4β**. Structural details from the computed trajectories show that the higher affinity observed with GAFF may be due to additional electrostatic interactions of the O1 hydroxyl in **4β** with the protein. The O1 hydroxyl in the equatorial position (**4β**) interacts with Arg17, Cys31 and Try37 (Fig. 7), but in the axial position (**4α**) it interacts largely with water molecules. Regardless of whether RSL binds preferably **4α** or **4β** and on the basis of the experimental data, neither anomer can be a better binder than its methylated form **1**. Thus, the GLYCAM calculations are in better agreement with the experiment but the trend for $\Delta\Delta G_{b,calc}$ $(\mathbf{1} \rightarrow \mathbf{4\beta})$ from both GLYCAM and GAFF for **4β** is not correct. However, it must be emphasized that $\Delta\Delta G_{b,exp}$ $(\mathbf{1} \rightarrow \mathbf{4})$ is only 0.7 kcal/mol, thus the magnitude of the deviation from experiment remains within range of the accuracy typically expected from an AFE calculation. It can also be stressed here that the difference between calculated binding free-energies for the **4α**→**4β** perturbation is very small, suggesting that both the anomers may readily bind to RSL. Table S1 shows the $\Delta\Delta G_{b,calc}$ $(\mathbf{1} \rightarrow \mathbf{4})$ on all six binding sites of the RSL. The $\Delta\Delta G_{b,calc}$ for **4α** and **4β** from a single run is slightly different among the binding sites, also indicating an individual preference in each of the six binding sites. However, similar to perturbation **1**→**2**, the mean $\Delta\Delta G_{b,calc}$ for all the six binding sites is comparable to the average from three independent simulations in the binding site S1. Similar free-energy profiles (Fig. S5) in all the six binding sites of RSL suggests that differences in $\Delta\Delta G_{b,calc}$ are merely statistical errors.

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

On the other hand, unlike in the previous perturbations where the performance of GAFF was comparable to GLYCAM, especially in modeling changes in hydrophobic interactions, binding free-energies for **5** from GLYCAM are more accurate (error ca. ~2 kcal•mol$^{-1}$) than those from GAFF (error ca. ~3 kcal•mol$^{-1}$). This is attributed to inacurate energetics for strong CH...$\pi$ stacking interactions between non-polar hydrogens of C6 in **1** and the aromatic amino acid residues of Trp76.[74] It has been shown by Wimmerova et al. that strong stacking interactions between carbohydrate and Trp76 in RSL largely contributes to the binding energy.[75] There are different views on the strength of stacking interactions between amino acids and saccharides, but these interactions can be stronger than a hydrogen bond and measured inetraction energies are ~8.0 kcal•mol$^{-1}$.[74,75] Current force fields incorporate stacking (CH...$\pi$ interactions) partially in the form of van der Waals interaction, but, there is no explicit term in the force fields to model CH...$\pi$ interactions. The CH...$\pi$ interactions are strongest when a CH bond is pointing directly towards the center of the aromatic ring.[55] In **1**, C6 methyl is strongly involved in the CH...$\pi$ interaction with Trp76, but upon perturbation to hydroxymethyl in **5**, this group rearranges to interact with water. The CH groups in hydroxymethyl can still form CH...$\pi$ interaction with the aromatic ring of Trp76, but this is modeled poorly with the present force fields. Hydrogens of C6 in **5** are not perpendicular to the aromatic ring due to dominating O6 hydroxyl-water interactions. This orientation of the O6 hydroxyl group is further stabilized by favorable electrostatic interactions with Arg18. Thus, a plausible explanation for the overestimation of the calculated binding free energies of **5** by ca. 2, and 3 kcal•mol$^{-1}$ from GLYCAM and GAFF respectively, is the inability of these force fields to account for CH…$\pi$ interactions, consequently promoting additional electrostatic interactions of O6 with water and Arg18.

### *3.3 Relative Free-energies of D-Fructose*

The MeFuc to D-Fru perturbation (**1**→**6**) is the largest structural change attempted in this study. D-Fru (**6**) is a weak binder ($\Delta G_{b,exp}$ = -5.5 kcal•mol$^{-1}$). As with the previously discussed non-methylated monosaccharides, **1**→**6** perturbations are performed for both α-D-

Fru (**6α**) and ß-D-Fru (**6ß**) anomers. The **1→6α** and **1→6ß** perturbations do not alter interactions of O2, O3 and O4 hydroxyl of **1**, and the changes are limited to C1 and C5. The axial methoxy and the equatorial hydrogen of C1 in **1** are morphed respectively into hydroxymethyl and hydroxyl groups in **6α** and vice versa in **6ß**. The equatorial methyl of C5 in **1** is perturbed into hydrogen in both **6α** and **6ß** (Fig. 1). The free-energy gradients accumulated for the **1→6α** perturbation are sufficiently converged along the given pathway. However, the computed free-energy gradients for perturbation **1→6ß** show a similar noisy profile as seen for the **1→3** and **2→3** perturbations discussed previously (Fig. S8). Performing the perturbation from **1** to **6ß** in two steps via the intermediate **10** resolved the issue (Fig. S9).

The $\Delta\Delta G_{b,calc}$ values obtained with GLYCAM for both, **1→6α** (3.5±0.4 kcal•mol$^{-1}$) and **1→6ß** (1.9±0.2 kcal•mol$^{-1}$) perturbation are in closer agreement with $\Delta\Delta G_{b,exp}(\mathbf{1} \rightarrow \mathbf{6})$ (3.1 kcal•mol$^{-1}$) than the values obtained with GAFF (-0.3±0.2 kcal•mol$^{-1}$ and -0.2±0.2 kcal•mol$^{-1}$, respectively). Anomer **6ß** is predicted to bind more favorably than **6α** with GLYCAM. In simulations performed with both GAFF and GLYCAM force fields, the equatorial hydroxymethyl of C2 in **6ß** forms a strong hydrogen bond with the Tyr37, which could not be formed with anomer **6α.** This anomer has a hydroxyl group in equatorial position of C2, which is too far to establish direct interaction with Tyr37. This explains why **6ß** binds stronger than **6α**. Further analysis of the computed trajectories is useful to establish why the binding energetics differ between the two force fields (Table S2). A close inspection of the free-energy gradients and trajectories of end-states does not reveal any noticeable features in the geometry or structural interactions of the molecules. A plausible explanation for the poor performance of GAFF here is that AM1-BCC may significantly underestimate the solvation free-energy of the carbohydrate ligand. Others have reported a systematic underestimation of hydration free-energies of ca. 1.5 kcal/mol for alcohols.[76] In saccharides, these errors could be larger due to the greater number of hydroxyl groups present in the structures; compound **6** contains five hydroxyls. The axial group $R_1$ is always solvent exposed and the equatorial group $R_2$ interacts with water in **6α** ($R_2$= –OH), but makes a hydrogen bond with Tyr37 in **6ß** ($R_2$= $-CH_2OH$). This suggests that accurate prediction of hydration energetics for hydroxyls is critical to model carbohydrate-protein binding.

### *3.4 Relative Free-energies of D-Rhamnose*

D-Rhamnose (**7**) is a weak binder ($\Delta G_{b,exp}$= -4.6 kcal•mol$^{-1}$) of RSL. As for other non-methylated monosaccharides, the **1**→**7** perturbation was performed separately for both α (**7α**) and ß (**7ß**) anomers. These perturbations do not affect orientations of O2, O3 and O4 hydroxyls in **1**, but changes are made at C1 and C5 in **1** only. The axial methoxy group and the equatorial hydrogen of C1 in **1** are perturbed into hydrogen and methyl groups respectively, whereas the axial hydrogen and the equatorial methyl of C5 in **1** are morphed into hydroxyl and hydrogen for **7α** and vice versa for **7ß**. The free-energy gradients collected for **1**→**7α** and **1**→**7ß** perturbations are smooth and well converged along the pathway. The trends of $\Delta\Delta G_{b,calc}$ values obtained with GAFF and GLYCAM are in agreement with $\Delta\Delta G_{b,exp}$ (Table S2). While both GAFF and GLYCAM give similar trends, it is found that GLYCAM calculations produced better results compared to GAFF, which overestimates energies by ca. 2.0 kcal/mol.

The $\Delta\Delta G_{b,calc}$ values from GLYCAM for **7α** and **7ß** are 4.3±0.1 and 2.6±0.1 kcal•mol$^{-1}$, respectively, whereas $\Delta\Delta G_{b,exp}(\mathbf{1} \rightarrow \mathbf{7})$ is 4.0 kcal•mol$^{-1}$. Thus, the calculations suggest that the binding of **7ß** is favored over **7α**, and the binding free energy is too favorable by ca. 1.4 kcal•mol$^{-1}$, too. The intermolecular interactions of both anomers during the simulations are quite similar except for electrostatic interactions between the O1 hydroxyl of C1 in **7α** and Trp81. The O1 hydroxyl resides inside the hydrophobic pocket in **7ß**. Interestingly, **7α** also samples a second binding mode during the simulations (Fig. S10). In this alternative binding mode, hydrogen bonds formed by O2, O3 and O4 hydroxyl of **7α** are still maintained, but O1 hydroxyl now interacts with water molecules. Over a 10 ns simulation, **7α** interconverts twice between the primary and secondary binding modes. This suggests that these two binding modes for **7α** are possible, and **7α** may have similar binding energy in both the binding modes as transitions are observed on this short timescale. The overall difference in binding affinities for **7α** and **7ß** relative to **1** is mainly attributed to the loss of hydrophobic interactions mediated by C6 methyl in **1,** and additional contribution from axial hydroxymethyl to hydrogen and equatorial hydrogen to methyl perturbation at C1 in **1**

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

D-Galactose (**8**) and L-Rhamnose (**9**) have not shown any significant binding in SPR experiments. Thus, both **8** and **9** are either non-binders or their binding was so weak that it could not be measured. The lowest $K_d$ measured using SPR measurement is 2.5 mM for **3** (MeMan). As a consequence, it is assumed that **8** and **9** are either non-binders or their $K_d$ is higher than 2.5 mM. Both **8** and **9** are denoted as non-binders in this report. Based on the $K_d$ values for **3,** $\Delta\Delta G_{b,exp}$ is expected to be greater than 6.0 kcal•mol$^{-1}$ for **1→8** and **1→9** perturbations. The **1→8** perturbation involves replacing the methoxy and hydrogen groups of C1 in **1** by hydrogen and hydroxymethyl groups, and methyl and hydrogen of C5 in **1** by hydroxyl and hydrogen groups (Fig. 1). Additionally, both **1→8** and **1→9** perturbations involve hydrogen to hydroxyl and vice-versa transformation at C2 and C4. The free-energy gradients for the **1→8** perturbation were not well converged with the one-step protocol. Transformation of the equatorial H of C1 in **1** to hydroxymethyl displayed again convergence issues. Thus, the same intermediate **10** was used to obtain smooth gradients for **1→8α** and **1→8ß** perturbations. Interestingly **8** and **9** are the only monosaccharides where orientation of O2 and O4 hydroxyl differs as compared to the high affinity binders **1** to **5**. Because of this, these non-binders cannot interact with binding site residues Trp76, Glu28 or Arg17 as observed for more potent ligands. Consequently **8** and **9** adopt binding modes that differ from **1** during the simulations.

The $\Delta\Delta G_{b,calc}$ values for **1→8α** and **1→8ß** are 5.2±0.2 and 2.7±0.4 kcal•mol$^{-1}$ with GAFF and 8.0±0.4 and 6.5±0.3 kcal•mol$^{-1}$ with GLYCAM, respectively. While both GAFF and GLYCAM give results in qualitative agreement with experiment, once again GLYCAM predictions are quantitatively much closer to the expected change of >6.5 kcal•mol$^{-1}$ in binding free energies (Table 1). For the **1→8** perturbation, the GAFF/AM1-BCC energies are quite different from the GLYCAM energies, showing a similar trend of errors as seen for **1→6ß**, which might also be ascribed to systematic errors in hydration free-energies of alcohols, though this cannot be quantified exactly as no experimental hydration free-energies are available in the literature for these monosaccharides.

The MeFuc→L-Rha perturbations (**1→9**) for both **9α** and **9ß** anomers converged well in a single step transformation. The $\Delta\Delta G_{b,calc}$ $(\mathbf{1} \rightarrow \mathbf{9\alpha})$ and $\Delta\Delta G_{b,calc}$ $(\mathbf{1} \rightarrow \mathbf{9\beta})$ values are 6.2±0.1 and 5.0±0.3 kcal•mol$^{-1}$ with GAFF and 6.7±0.2 and 6.9±0.2 kcal•mol$^{-1}$ with GLYCAM respectively. As expected, the calculated change in the free-energy of binding using GLYCAM is greater than 6.0 kcal•mol$^{-1}$ for both anomers. In the simulations, the α/ß L-Rha ligands are quite unstable inside the binding site. Figure 9 shows the starting conformation in green and other conformations sampled at λ=0 during the **9→1** perturbation. During the simulation, both **8** and **9** attains conformations where interaction of the saccharide with protein side-chains differs significantly. Thus, the simulations clearly suggest that both **8** and **9** are non-binders owing to the lack of a defined binding mode.

## 4. Discussion and Conclusion

A procedure that allows an efficient binding free-energy calculation for monosaccharide ligand substituents in protein-carbohydrate complexes using GAFF/AM1-BCC and

GLYCAM force fields is described. The current results show that accurate relative binding free-energies of protein-carbohydrate complexes are possible using AFE calculations.

Moreover, unlike in previous studies performed with alternative modelling methods,[23,27] AFE calculations are here able to reliably discriminate non-binders, weak-binders and potent-binders. On the technical side, setting up AFE calculations for carbohydrates presented a number of difficulties, and in a number of instances, two steps pathways were shown to outperform direct perturbations between pairs of monosaccharides. A two step pathway is recommended when appearing/disappearing hydroxymethyl groups at equatorial position of C1 or C5 carbon of hexopyranose in their $^1C_4$ or $^4C_1$ chair conformation. Provided the perturbation pathways are adequately defined, the AFE calculation setup using 16 non-equally distant λ windows and four ns simulation for each window was sufficient to obtain converged values in most instances.

The main results of this study are a careful assessment of the accuracy of the GAFF/AM1-BCC and GLYCAM force fields in computing free-energies of binding of protein-carbohydrate complexes. To achieve high precision, each relative binding free-energy was estimated from at least 6 independent calculations. In general, GLYCAM outperformed GAFF/AM1-BCC. The results suggest that GAFF/AM1-BCC can be as accurate as GLYCAM when perturbations only affect a change in hydrophobic interaction as for the **1→2** perturbation. GAFF/AM1-BCC proved to be less accurate in modeling the energetics of hydroxyl groups and this is attributed to systematic errors in hydration energetics for this functional group.

By contrast, the GLYCAM force field reproduces experiential changes in the free-energy of binding in most of the cases. Assuming a relative binding free-energy of 6.5 kcal•mol$^{-1}$ for all the non-binders, GLYCAM produces a mean unsigned error MUE = 1.1 ± 0.06 kcal•mol$^{-1}$, a correlation coefficient $R^2 = 0.85 \pm 0.02$ and a predictive index PI = 0.94 ± 0.01. In comparison GAFF/AM1-BCC achieves on this dataset a statistically significant inferior performance, with MUE = 2.60 ± 0.08 kcal•mol$^{-1}$, $R^2$ = 0.59 ± 0.03 and PI = 0.80 ± 0.01. Thus, while GAFF shows reasonable ranking abilities, the mean errors with GAFF are too large to reliably discriminate potent binders, weak binders and non-binders which is important for hit-to-lead and lead optimization purposes. Thus the versatility of the current version of GAFF and ease of parameterization of carbohydrates derived ligands is more than offset by significant performance degradation and it is preferable to focus efforts on extending GLYCAM parameter sets for detailed modelling studies of carbohydrate-based ligands. The present results can be compared with other free-energy studies of carbohydrates reported in the literature. Kadirvelraj et al achieved an accuracy of 0.5 kcal•mol$^{-1}$ using GLYCAM for hydroxyl to methoxy and hydroxyl to hydroxyethyl perturbation[12], and ~0.5 kcal•mol$^{-1}$ for 6 mutations in antibody-carbohydrate-antigen complexes[77] using GLYCAM_98R. Although a similar level of accuracy is obtained using LIE models parameterized on protein-carbohydrate complexes, transferability of parameters to other systems remains a challenge.[27] Errors are higher when using MM-PB/GBSA on protein-carbohydrate complexes.[24,28]

As the same dataset has been studied previously by docking and LIE methodologies, a systematic comparison of the predictive power of a broad range of methodologies is presented in Figure 10. AFE calculations using GLYCAM provides the highest correlation ($R^2$=0.85±0.02), predictive index (PI=0.94±0.01) and lowest MUE (1.1±0.06 kcal.mol$^{-1}$) with experiment data (see also Table S7 for details). Among the several docking programs evaluated only AutoDock3 with RESP did well in terms of correlation ($R^2$=0.78) and predictive index (PI=0.83), but not MUE (3.6 kcal.mol$^{-1}$). Thus, AutoDock3 outperforms GAFF on this dataset. The performance of the tool is sensitive to the charge model used, and results obtained with Gasteiger charges are inferior. All other docking programs are much inferior to GLYCAM and GAFF. The LIE approach (performed with an OPLS-AA 2005 force field) is broadly speaking comparable in accuracy with the present AFE results obtained with GAFF/AM1-BCC ($R^2$=0.60, PI=0.80, MUE 2.6 kcal.mol$^{-1}$). Overall, AFE calculations with GLYCAM outperform all other methods, but the accuracy of GAFF is comparable to LIE or the best performing docking protocols.

The accuracy achieved with GLYCAM in the present study is thus very encouraging, and we expect a broad applicability to other protein-carbohydrate systems albeit within the limits of current force fields. The present study is also the first to include non-binders in the dataset, thus demonstrating the capability of TI to predict reliably a wide range of binding energetics. The only noticeable error with GLYCAM is for the relative binding free-energies of MeMan (**3**) and L-Gal (**5**), which is overestimated using both GAFF/AM1-BCC and GLYCAM. Only ligands **3** and **5** have a hydroxymethyl and methoxy group at ring carbon C5 and C1, which is just below Trp76. Thus, both **3** and **5** remains a challenge presumably because of the inability to quantify CH...π interactions between carbohydrates and aromatic amino acid residues accurately in current force fields. Such errors are also possible for **8** but this interaction was not apparent owing to a lack of defined binding mode during the simulations. Nevertheless, predicted energies for L-Gal (**5**) are in much better agreement than those predicted with MM/PBSA where the absolute binding energy of L-Gal complexed with PA-IIL was overestimated by ~ -20 kcal/mol.[24] While hexopyranoses assume preferentially a chair conformation ($^1C_4$ L-Galacto or $^4C_1$ D-Manno), ring conformation sampling of the hexopyranoses has not been addressed in this work.

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

A**bbreviations:**

MeFuc: Me-α-L-Fuc, MeAra: Me-β-D-Ara, MeMan: Me-α-D-Man:, αFuc: α-L-Fuc, βFuc: β-L-Fuc, βFru: β-D-Fru, αFru: α-D-Fru, W: Water, P: Protein, Cal: Calculated, Exp : Experimental, methoxy: $-OCH_3$, Hydroxymethyl: $-CH_2OH$

## ASSOCIATED CONTENT

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

## AUTHOR INFORMATION

### Corresponding Authors

* Phone: +420-549494947 (JK). E-mail: jkoca@ceitec.cz (J.K)

* Phone: +44-131-6504797 (JM). E-mail: mail@julienmichel.net (JM)

### Authors Contribution

SKM, JM and JK conceived and designed the calculations. SKM performed calculations and analyzed the data. GC, HL and JM provided GLYCAM support in FESetup and Sire/OpenMM calculations. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

### Funding Sources

This work was supported by the Grant Agency of the Czech Republic [13-25401S]. This work was further supported by the project "CEITEC–Central European Institute of

Technology” [CZ.1.05/1.1.00/02.0068] from the European Regional Development Fund. Additional support was provided by the project "Employment of Newly Graduated Doctors of Science for Scientific Excellence" [CZ.1.07/2.3.00/30.0009] to SKM, co-financed from the European Social Fund and the state budget of the Czech Republic. JM is supported by a Royal Society University Research Fellowship. HL acknowledges support from CCP-BioSim for development of the FEsetup software. Access to the MetaCentrum computing facilities provided under the programme “Large Infrastructure Projects for Research, Development, and Innovation" [LM2010005]; and CERIT-SC computing facilities provided by the CERIT Scientific Cloud programme center, part of the Operational Program Research and Development for Innovation [reg. no. CZ.1.05/3.2.00/08.0144] are appreciated.

**Notes**

The authors declare no competing financial interests.

**References:**

(1) De Ruiter, A.; Oostenbrink, C. Free Energy Calculations of Protein-Ligand Interactions. *Curr. Opin. Chem. Biol.* **2011**, *15* (4), 547–552.

(2) Steinbrecher, T.; Labahn, A. Towards Accurate Free Energy Calculations in Ligand Protein-Binding Studies. *Curr. Med. Chem.* **2010**, *17* (8), 767–785.

(3) Michel, J.; Foloppe, N.; Essex, J. W. Rigorous Free Energy Calculations in Structure-Based Drug Design. *Mol. Inform.* **2010**, *29* (8-9), 570–578.

(4) Bash, P.; Singh, U.; Brown, F.; Langridge, R.; Kollman, P. Calculation of the Relative Change in Binding Free Energy of a Protein-Inhibitor Complex. *Science* **1987**, *235* (4788), 574–576.

(5) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93* (7), 2395–2417.

(6) Homeyer, N.; Stoll, F.; Hillisch, A.; Gohlke, H. Binding Free Energy Calculations for Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design Context. *J. Chem. Theory Comput.* **2014**, *10* (8), 3331–3344.

(7) Pan, K.; Deem, M. W. Predicting Fixation Tendencies of the H3N2 Influenza Virus by Free Energy Calculation. *J. Chem. Theory Comput.* **2011**, *7* (5), 1259–1272.

(8) Tzoupis, H.; Leonis, G.; Mavromoustakos, T.; Papadopoulos, M. G. A Comparative Molecular Dynamics, MM-PBSA and Thermodynamic Integration Study of Saquinavir Complexes with Wild-Type HIV-1 PR and L10I, G48V, L63P, A71V, G73S, V82A and I84V Single Mutants. *J. Chem. Theory Comput.* **2013**, *9* (3), 1754–1764.

(9) Michel, J. Current and Emerging Opportunities for Molecular Simulations in Structure-Based Drug Design. *Phys. Chem. Chem. Phys.* **2014**, *16* (10), 4465–4477.

(10) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, *10* (7), 2632–2647.

(11) Shirts, M. Best Practices in Free Energy Calculations for Drug Design. *Methods Mol. Biol. Clifton NJ* **2012**, *819*, 425–467.

(12) Kadirvelraj, R.; Foley, B. L.; Dyekjær, J. D.; Woods, R. J. Involvement of Water in Carbohydrate−Protein Binding: Concanavalin A Revisited. *J. Am. Chem. Soc.* **2008**, *130* (50), 16933–16942.

(13) Abhilash, J.; Dileep, K. V.; Palanimuthu, M.; Geethanandan, K.; Sadasivan, C.; Haridas, M. Metal Ions in Sugar Binding, Sugar Specificity and Structural Stability of Spatholobus Parviflorus Seed Lectin. *J. Mol. Model.* **2013**, *19* (8), 3271–3278.

(14) Krivan, H. C.; Plosila, L.; Zhang, L.; Holt, V.; Kyogashima, M. Cell Surface Carbohydrates as Adhesion Receptors for Many Pathogenic and Opportunistic Microorganisms. In *Microbial Adhesion and Invasion*; Hook, M., Switalski, L., Eds.; Springer New York, 1992; pp 1–13.

(15) Jones, C. Vaccines Based on the Cell Surface Carbohydrates of Pathogenic Bacteria. *An. Acad. Bras. Ciênc.* **2005**, *77* (2), 293–324.

(16) Tuccillo, F. M.; de Laurentiis, A.; Palmieri, C.; Fiume, G.; Bonelli, P.; Borrelli, A.; Tassone, P.; Scala, I.; Buonaguro, F. M.; Quinto, I.; Scala, G. Aberrant Glycosylation as Biomarker for Cancer: Focus on CD43. *BioMed Res. Int.* **2014**, *2014*, e742831.

(17) Gabius, H.-J.; Siebert, H.-C.; André, S.; Jiménez-Barbero, J.; Rüdiger, H. Chemical Biology of the Sugar Code. *Chembiochem Eur. J. Chem. Biol.* **2004**, *5* (6), 740–764.

(18) Albersheim, P.; Anderson-Prouty, A. J. Carbohydrates, Proteins, Cell Surfaces, and the Biochemistry of Pathogenesis. *Annu. Rev. Plant Physiol.* **1975**, *26* (1), 31–52.

(19) Neumann, D.; Kohlbacher, O.; Lenhof, H.-P.; Lehr, C.-M. Lectin-Sugar Interaction. Calculated versus Experimental Binding Energies. *Eur. J. Biochem. FEBS* **2002**, *269* (5), 1518–1524.

(20) Kerzmann, A.; Neumann, D.; Kohlbacher, O. SLICK--Scoring and Energy Functions for Protein-Carbohydrate Interactions. *J. Chem. Inf. Model.* **2006**, *46* (4), 1635–1642.

(21) Kerzmann, A.; Fuhrmann, J.; Kohlbacher, O.; Neumann, D. BALLDock/SLICK: A New Method for Protein-Carbohydrate Docking. *J. Chem. Inf. Model.* **2008**, *48* (8), 1616–1625.

(22) Adam, J.; Kříž, Z.; Prokop, M.; Wimmerová, M.; Koča, J. In Silico Mutagenesis and Docking Studies of Pseudomonas Aeruginosa PA-IIL Lectin — Predicting Binding Modes and Energies. *J. Chem. Inf. Model.* **2008**, *48* (11), 2234–2242.

(23) Mishra, S. K.; Adam, J.; Wimmerova, M.; Koca, J. In Silico Mutagenesis and Docking Study of Ralstonia Solanacearum RSL Lectin: Performance of Docking Software To Predict Saccharide Binding. *J. Chem. Inf. Model.* **2012**, *52* (5), 1250–1261.

(24) Mishra, N. K.; Kríz, Z.; Wimmerová, M.; Koca, J. Recognition of Selected Monosaccharides by Pseudomonas Aeruginosa Lectin II Analyzed by Molecular Dynamics and Free Energy Calculations. *Carbohydr. Res.* **2010**, *345* (10), 1432–1441.

(25) Fadda, E.; Woods, R. J. Molecular Simulations of Carbohydrates and Protein-Carbohydrate Interactions: Motivation, Issues and Prospects. *Drug Discovery. Today* **2010**, *15* (15-16), 596–609.

(26) Bryce, R. A.; Hillier, I. H.; Naismith, J. H. Carbohydrate-Protein Recognition: Molecular Dynamics Simulations and Free Energy Analysis of Oligosaccharide Binding to Concanavalin A. *Biophys. J.* **2001**, *81* (3), 1373–1388.

(27) Mishra, S. K.; Sund, J.; Åqvist, J.; Koča, J. Computational Prediction of Monosaccharide Binding Free Energies to Lectins with Linear Interaction Energy Models. *J. Comput. Chem.* **2012**, *33* (29), 2340–2350.

(28) Topin, J.; Arnaud, J.; Sarkar, A.; Audfray, A.; Gillon, E.; Perez, S.; Jamet, H.; Varrot, A.; Imberty, A.; Thomas, A. Deciphering the Glycan Preference of Bacterial Lectins by Glycan Array and Molecular Docking with Validation by Microcalorimetry and Crystallography. *PloS One* **2013**, *8* (8), e71149.

(29) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**.

(30) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. Additive Empirical Force Field for Hexopyranose Monosaccharides. *J. Comput. Chem.* **2008**, *29* (15), 2543–2564.

(31) Lins, R. D.; Hünenberger, P. H. A New GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J. Comput. Chem.* **2005**, *26* (13), 1400–1412.

(32) Kony, D.; Damm, W.; Stoll, S.; Van Gunsteren, W. F. An Improved OPLS-AA Force Field for Carbohydrates. *J. Comput. Chem.* **2002**, *23* (15), 1416–1429.

(33) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29* (4), 622–655.

(34) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.

(35) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Modell.* **2006**, *25* (2), 247–260.

(36) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21* (2), 132–146.

(37) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23* (16), 1623–1641.

(38) Rangarajan, E. S.; Proteau, A.; Cui, Q.; Logan, S. M.; Potetinova, Z.; Whitfield, D.; Purisima, E. O.; Cygler, M.; Matte, A.; Sulea, T.; Schoenhofen, I. C. Structural and Functional Analysis of Campylobacter Jejuni PseG: A Udp-Sugar Hydrolase from the Pseudaminic Acid Biosynthetic Pathway. *J. Biol. Chem.* **2009**, *284* (31), 20989–21000.

(39) Hendrickx, P. M. S.; Corzana, F.; Depraetere, S.; Tourwé, D. A.; Augustyns, K.; Martins, J. C. The Use of Time-Averaged 3JHH Restrained Molecular Dynamics (tar-MD) Simulations for the Conformational Analysis of Five-Membered Ring Systems: Methodology and Applications. *J. Comput. Chem.* **2010**, *31* (3), 561–572.

(40) Cruz, L.; Brás, N. F.; Teixeira, N.; Mateus, N.; Ramos, M. J.; Dangles, O.; De Freitas, V. Vinylcatechin Dimers Are Much Better Copigments for Anthocyanins than Catechin Dimer Procyanidin B3. *J. Agric. Food Chem.* **2010**, *58* (5), 3159–3166.

(41) Sommer, R.; Exner, T. E.; Titz, A. A Biophysical Study with Carbohydrate Derivatives Explains the Molecular Basis of Monosaccharide Selectivity of the Pseudomonas Aeruginosa Lectin LecB. *PloS One* **2014**, *9* (11), e112822.

(42) Ernst, B.; Magnani, J. L. From Carbohydrate Leads to Glycomimetic Drugs. *Nat. Rev. Drug Discov.* **2009**, *8* (8), 661–677.

(43) Campbell, L. K.; Baker, D. E.; Campbell, R. K. Miglitol: Assessment of Its Role in the Treatment of Patients with Diabetes Mellitus. *Ann. Pharmacother.* **2000**, *34* (11), 1291–1301.

(44) Chen, X.; Zheng, Y.; Shen, Y. Voglibose (Basen, AO-128), One of the Most Important Alpha-Glucosidase Inhibitors. *Curr. Med. Chem.* **2006**, *13* (1), 109–116.

(45) Weinreb, N. J.; Barranger, J. A.; Charrow, J.; Grabowski, G. A.; Mankin, H. J.; Mistry, P. Guidance on the Use of Miglustat for Treating Patients with Type 1 Gaucher Disease. *Am. J. Hematol.* **2005**, *80* (3), 223–229.

(46) Zacharias, M.; Straatsma, T. P.; McCammon, J. A.; Quiocho, F. A. Inversion of Receptor Binding Preferences by Mutagenesis: Free Energy Thermodynamic Integration Studies on Sugar Binding to L-Arabinose Binding Proteins. *Biochemistry (Mosc.)* **1993**, *32* (29), 7428–7434.

(47) Bucher, D.; Grant, B. J.; McCammon, J. A. Induced Fit or Conformational Selection? The Role of the Semi-Closed State in the Maltose Binding Protein. *Biochemistry (Mosc.)* **2011**, *50* (48), 10530–10539.

(48) Koppisetty, C. A. K.; Frank, M.; Lyubartsev, A. P.; Nyholm, P.-G. Binding Energy Calculations for Hevein-Carbohydrate Interactions Using Expanded Ensemble Molecular Dynamics Simulations. *J. Comput. Aided Mol. Des.* **2015**, *29* (1), 13–21.

(49) Kostlánová, N.; Mitchell, E. P.; Lortat-Jacob, H.; Oscarson, S.; Lahmann, M.; Gilboa-Garber, N.; Chambat, G.; Wimmerova, M.; Imberty, A. The Fucose-Binding Lectin from Ralstonia Solanacearum. A New Type of Beta-Propeller Architecture Formed by Oligomerization and Interacting with Fucoside, Fucosyllactose, and Plant Xyloglucan. *J. Biol. Chem.* **2005**, *280* (30), 27839–27849.

(50) Schell, M. A. Control of Virulence and Pathogenecty Genes of Ralstonia Solanacearum by an Elaborate Sensory Network. *Annu. Rev. Phytopathol.* **2000**, *38*, 263–292.

(51) Woods, R. J.; and coworkers. GLYCAM Web http://glycam.org/ (accessed Aug 1, 2014).

(52) Autieri, E.; Sega, M.; Pederiva, F.; Guella, G. Puckering Free Energy of Pyranoses: A NMR and Metadynamics-Umbrella Sampling Investigation. *J. Chem. Phys.* **2010**, *133* (9), 095104.

(53) Ardèvol, A.; Biarnés, X.; Planas, A.; Rovira, C. The Conformational Free-Energy Landscape of B-D-Mannopyranose: Evidence for a 1S5 → B2,5 → OS2 Catalytic Itinerary in B-Mannosidases. *J. Am. Chem. Soc.* **2010**, *132* (45), 16058–16065.

(54) Loeffler, H. H.; Woods, C. J.; Michel, J. FESetup 1.0. *http://ccpforge.cse.rl.ac.uk/gf/project/ccpbiosim/.* (accessed on 02/15/2015)

(55) Case, D.A.; Babin, V.; Berryman, J.T.; Betz, R.M.; Cai, Q.; Cerutti, D.S.; Cheatham, III, T.E.; Darden, T.A.; Duke, R.E.; Gohlke, H.; Goetz, A.W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, Kovalenko,I. A.; Lee, T.S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K.M.; Paesani, F.; Roe, D.R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C.L.; Smith, W.; Swails, J.; Walker R.C; Wang, J.; Wolf, R.M.; Wu, X. and Kollman P. A. AMBER 14, University of California, San Francisco, 2014.

(56) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65* (3), 712–725.

(57) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72* (3), 1047–1069.

(58) Mezei, M. The Finite Difference Thermodynamic Integration, Tested on Calculating the Hydration Free Energy Difference between Acetone and Dimethylamine in Water. *J. Chem. Phys.* **1987**, *86* (12), 7084–7088.

(59) Shyu, C.; Ytreberg, F. M. Reducing the Bias and Uncertainty of Free Energy Estimates by Using Regression to Fit Thermodynamic Integration Data. *J. Comput. Chem.* **2009**, *30* (14), 2297–2304.

(60) Wang, J.; Dixon, R.; Kollman, P. A. Ranking Ligand Binding Affinities with Avidin: A Molecular Dynamics-Based Interaction Energy Study. *Proteins* **1999**, *34* (1), 69–81.

(61) Michel, J.; Verdonk, M. L.; Essex, J. W. Protein−Ligand Complexes: Computation of the Relative Free Energy of Different Scaffolds and Binding Modes. *J. Chem. Theory Comput.* **2007**, *3* (5), 1645–1655.

(62) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure And/or Temperature. *J. Chem. Phys.* **1980**, *72* (4), 2384.

(63) Tironi, I. G.; Sperb, R.; Smith, P. E.; Gunsteren, W. F. van. A Generalized Reaction Field Method for Molecular Dynamics Simulations. *J. Chem. Phys.* **1995**, *102* (13), 5451–5459.

(64) Fennell, C. J.; Gezelter, J. D. Is the Ewald Summation Still Necessary? Pairwise Alternatives to the Accepted Standard for Long-Range Electrostatics. *J. Chem. Phys.* **2006**, *124* (23), 234104.

(65) Woods, C. J.; Calabro, G.; Michel, J. Sire Molecular Simulation Framework, Revision 2702, 2014, http://siremol.org/Sire/Home.html; (accessed on 02/15/2015)

(66) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2012**, *9* (1), 461–469.

(67) Brown, S. P.; Muchmore, S. W.; Hajduk, P. J. Healthy Skepticism: Assessing Realistic Model Performance. *Drug Discovery. Today* **2009**, *14* (7–8), 420–427.

(68) Mikulskis, P.; Genheden, S.; Ryde, U. A Large-Scale Test of Free-Energy Simulation Estimates of Protein–Ligand Binding Affinities. *J. Chem. Inf. Model.* **2014**, *54* (10), 2794–2806.

(69) Pearlman, D. A.; Charifson, P. S. Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System. *J. Med. Chem.* **2001**, *44* (21), 3417–3423.

(70) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Effects of Water Placement on Predictions of Binding Affinities for p38α MAP Kinase Inhibitors. *J. Chem. Theory Comput.* **2010**, *6* (12), 3850–3856.

(71) Morozov, A. V.; Misura, K. M. S.; Tsemekhman, K.; Baker, D. Comparison of Quantum Mechanics and Molecular Mechanics Dimerization Energy Landscapes for Pairs of Ring-Containing Amino Acids in Proteins. *J. Phys. Chem. B* **2004**, *108* (24), 8489–8496.

(72) Juaristi, E. *Conformational Behavior of Six-Membered Rings: Analysis, Dynamics and Stereoelectronic Effects*; Wiley-VCH Verlag GmbH: New York, 1995.

(73) Homola, J.; Yee, S. S.; Gauglitz, G. Surface Plasmon Resonance Sensors: Review. *Sens. Actuators B Chem.* **1999**, *54* (1–2), 3–15.

(74) Kozmon, S.; Matuška, R.; Spiwok, V.; Koča, J. Three-Dimensional Potential Energy Surface of Selected Carbohydrates' CH/π Dispersion Interactions Calculated by High-Level Quantum Mechanical Methods. *Chem. – Eur. J.* **2011**, *17* (20), 5680–5690.

(75) Wimmerova, M.; Kozmon, S.; Nečasová, I.; Mishra, S. K.; Komárek, J.; Koca, J. Stacking Interactions between Carbohydrate and Protein Quantified by Combination of Theoretical and Experimental Methods. *PLoS ONE* **2012**, *7* (10), e46032.

(76) Fennell, C. J.; Wymer, K. L.; Mobley, D. L. A Fixed-Charge Model for Alcohol Polarization in the Condensed Phase, and Its Role in Small Molecule Hydration. *J. Phys. Chem. B* **2014**, *118* (24), 6438–6446.

(77) Pathiaseril, A.; Woods, R. J. Relative Energies of Binding for Antibody-Carbohydrate-Antigen Complexes Computed from Free-Energy Simulations. *J. Am. Chem. Soc.* **2000**, *122* (2), 331–338.

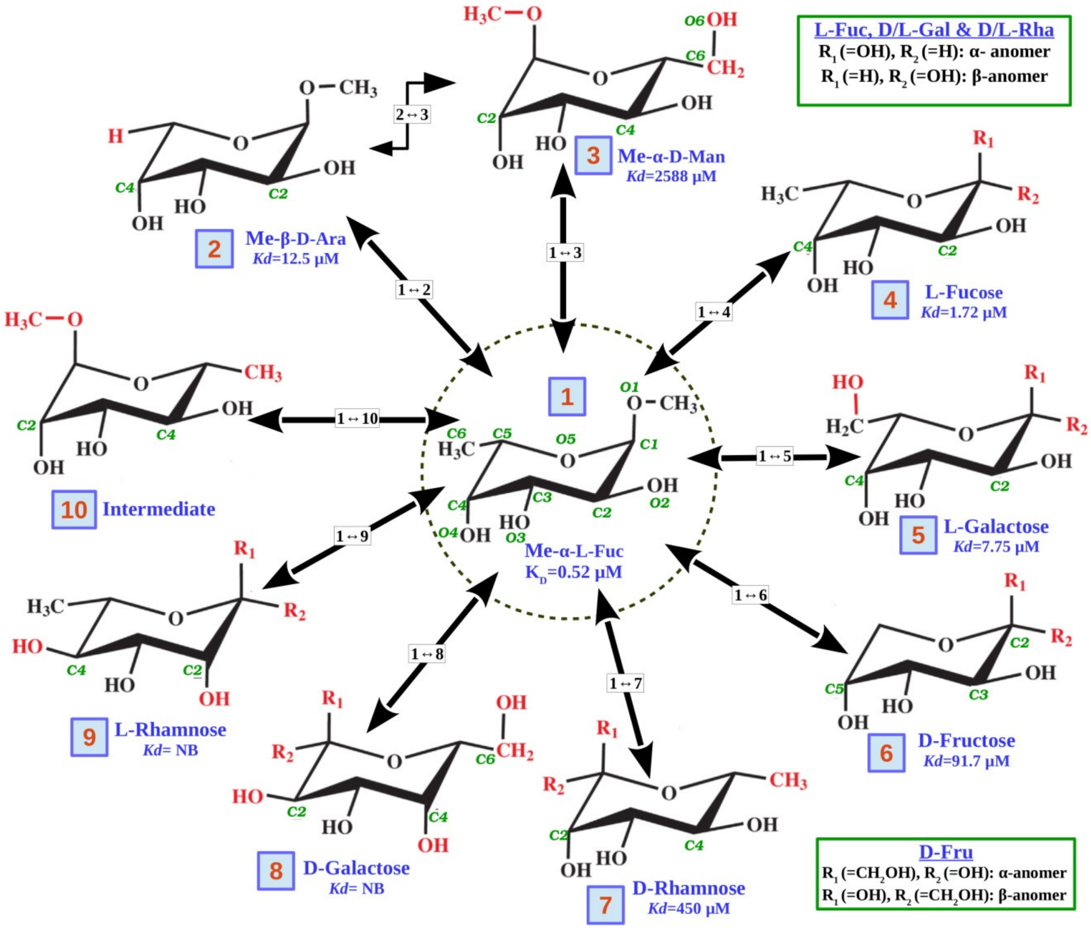


**Figure 1**. Graphical representations of perturbations of Me-α-L-Fuc (**1**) into different labeled monosaccharides. Red atoms that differ from topologically equivalent atoms in Me-α-L-Fuc. Experimental dissociation constants ($K_d$) of each monosaccharide for RSL lectin are in micromole/L. R1 and R2 for α- & β- 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.

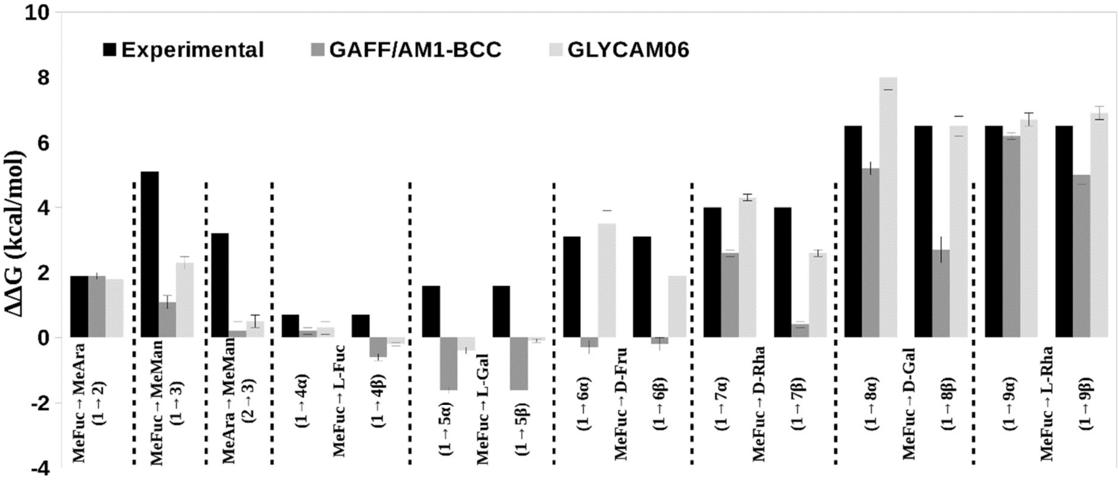


**Figure 2**. Experimental $\left(\Delta\Delta G_{b,exp}(X \rightarrow Y)\right)$ and calculated $\left(\Delta\Delta G_{b,calc}(X \rightarrow Y)\right)$ relative binding free-energies using the GAFF1.7/AM1-BCC (GAFF) and GLYCAM06j force fields. Estimated uncertainties (err $\left(\Delta\Delta G_{b,calc}(X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

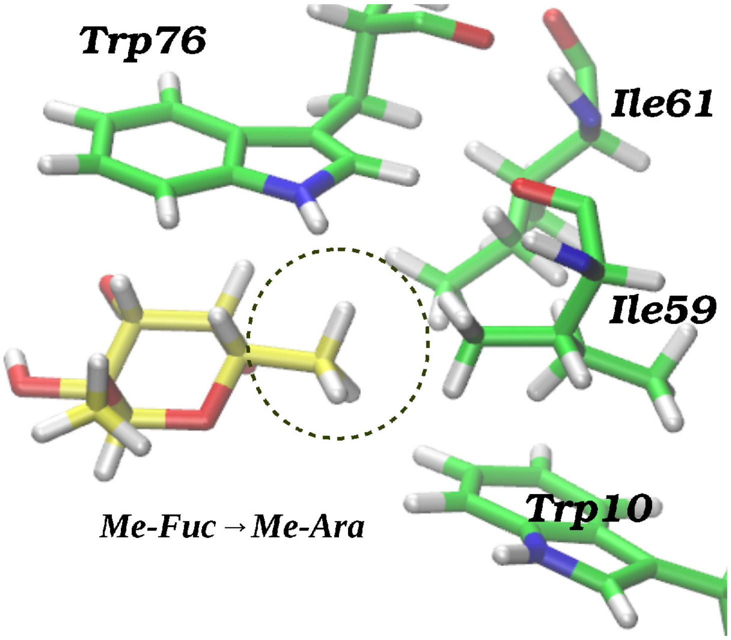


**Figure 3.** The Me-α-L-Fuc to Me-β-D-Ara perturbation (**1**→**2**) in the complex with RSL. The methyl group (circle) is replaced by a hydrogen atom within the hydrophobic patch of the protein.

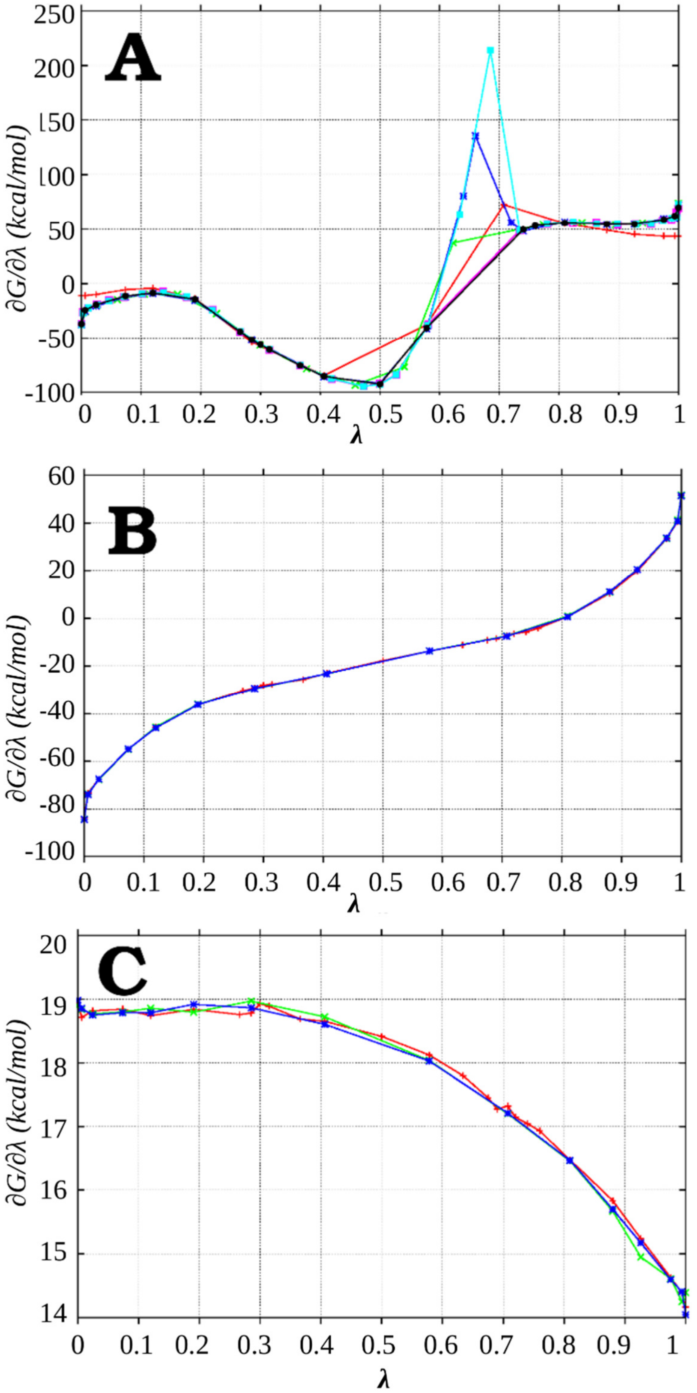


**Figure 4.** The free-energy gradients from Me-α-L-Fuc→Me-α-D-Man (**1→3**) perturbation using GLYCAM. (A) Multiple setups of **1→3** perturbation in single step. (B&C) Replicates of **1→10** and **10→3** perturbation, respectively. Multiple replicates are shown in different color.

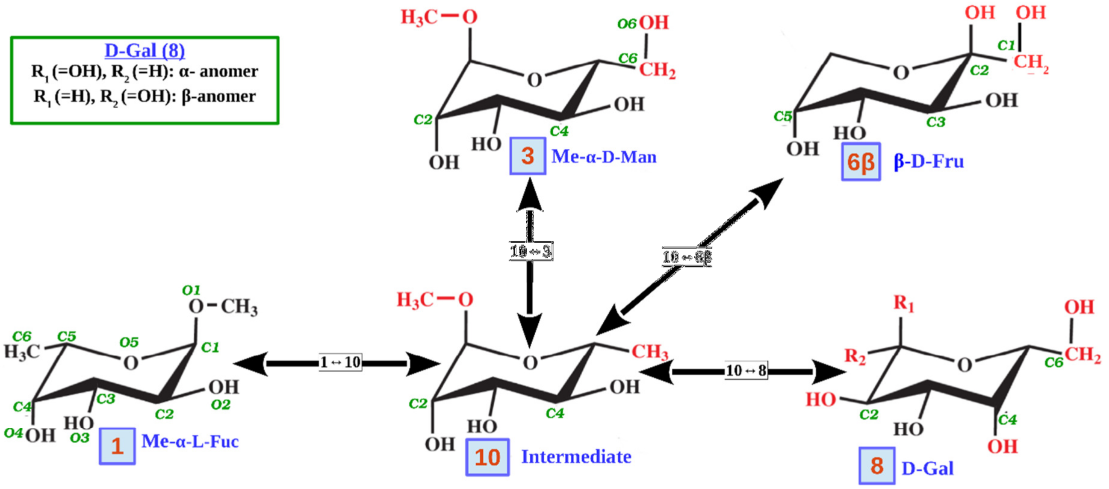


**Figure 5**. Graphical representations of perturbing Me-α-L-Fuc (**1**) into labeled monosaccharides via an intermediate structure (**10**). Red atoms/groups are the ones that differ from **1**. R1 and R2 for α- & β- anomers of D-Gal (**8**) are specified in the legend.

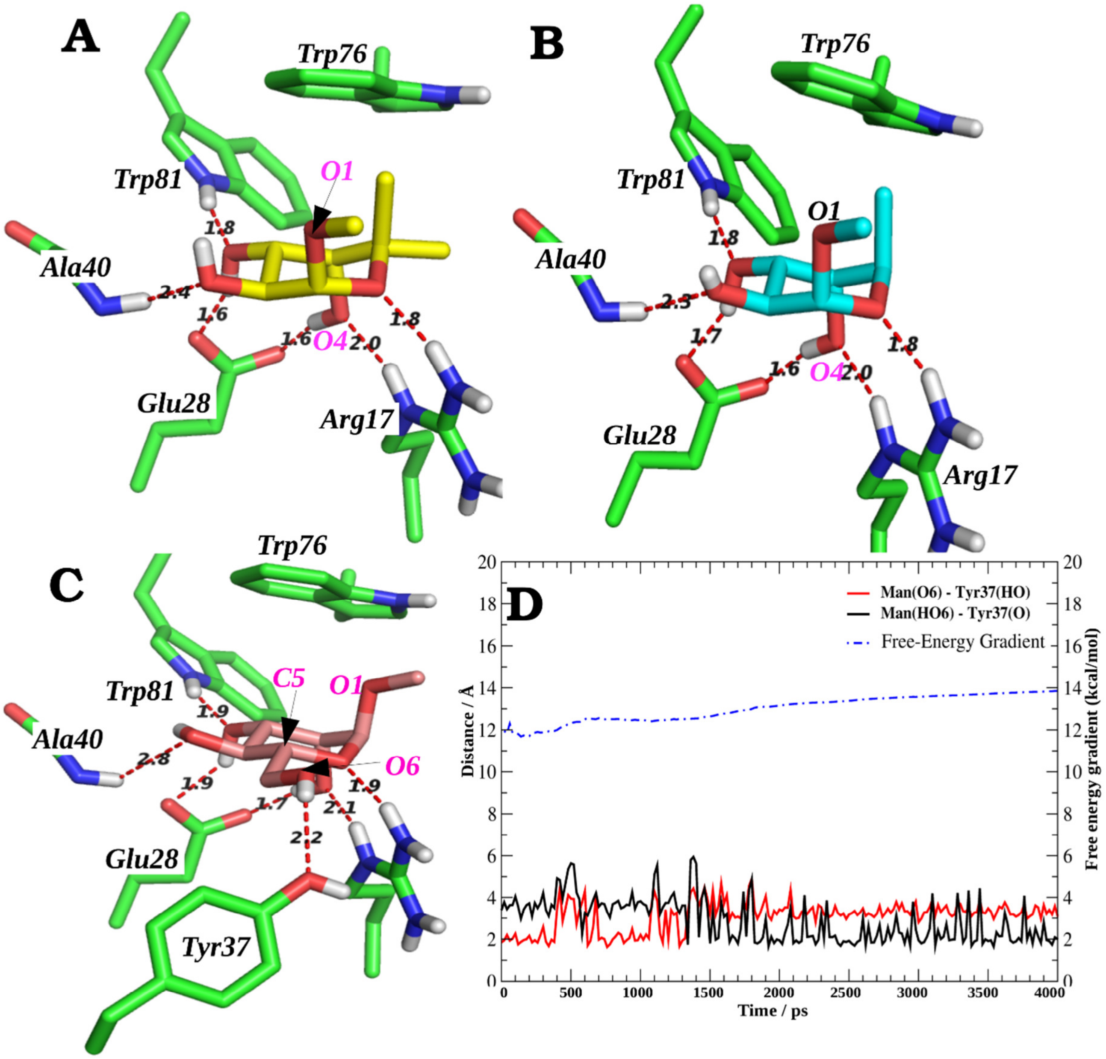


**Figure 6.** Average structures from MD simulation of saccharides in the protein bound state. (A) Me-α-L-Fucoside (**1**), (B) Me-β-D-Arabinoside (**2**), and (C) Me-α-D-Mannoside (**3**). All the distances shown in red dotted lines are in angstrom. (D) Time series of the hydrogen bond distance between 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.

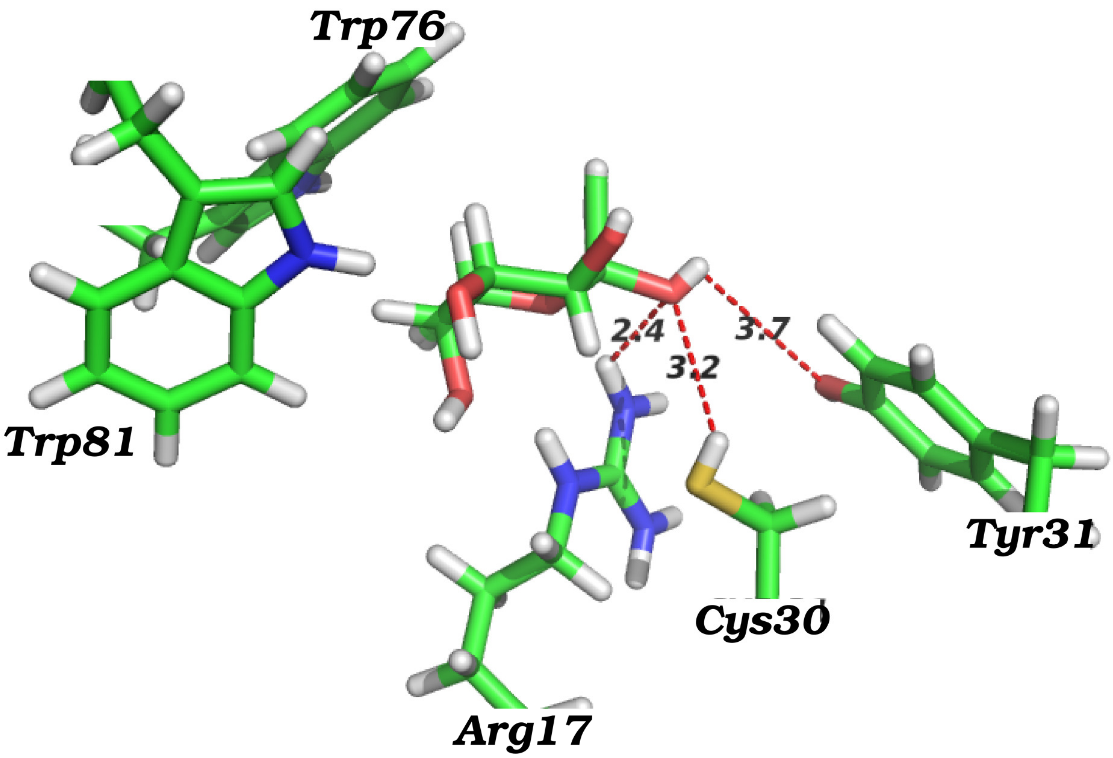


**Figure 7.** Average structure from Me-α-L-Fucose to β-L-Fucose perturbation (**1→4β**) in the complex with RSL. All the distances from O1 hydroxyl shown in red dotted lines are in Angstrom.

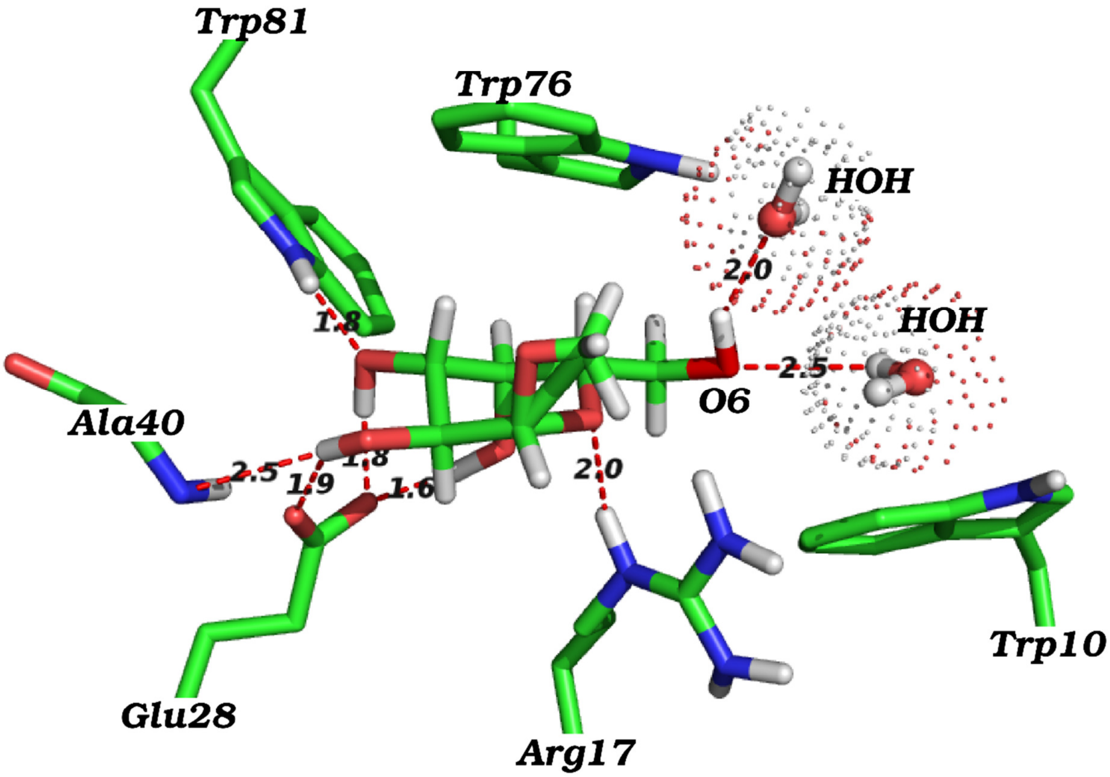


**Figure 8.** Average structure from Me-α-L-Fuc to β-L-Gal perturbation (**1→5β**) in the complex with RSL. All the distances from O6 hydroxyl shown in red dotted lines are in Angstrom. There are two water molecules interacting closely with O6 hydroxyl in **5β** are shown with dotted van der Waals surface.

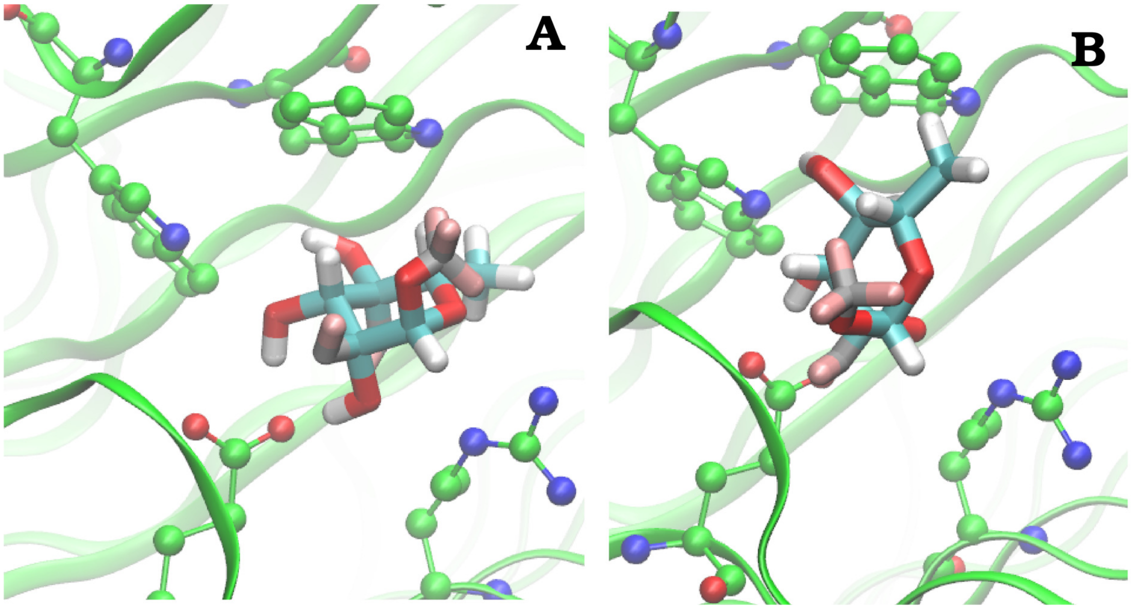


**Figure 9**. Two different binding conformations of L-Rha (**9**) during the simulations (**9→1**). (A) Initial and well defined binding mode of α-L-Rha (**9α**) and (B) poorly binding mode from simulation (**9β**).

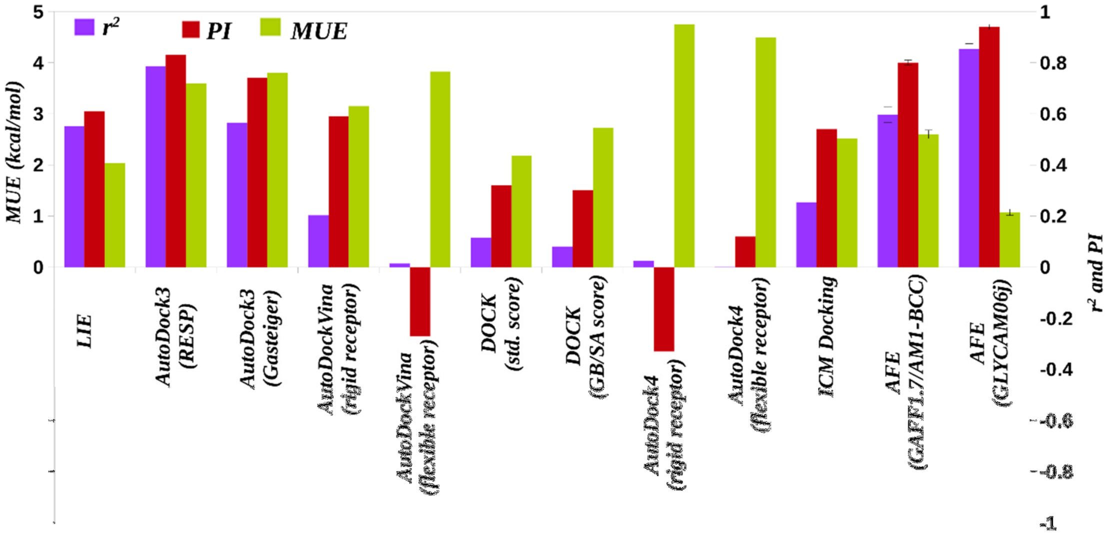
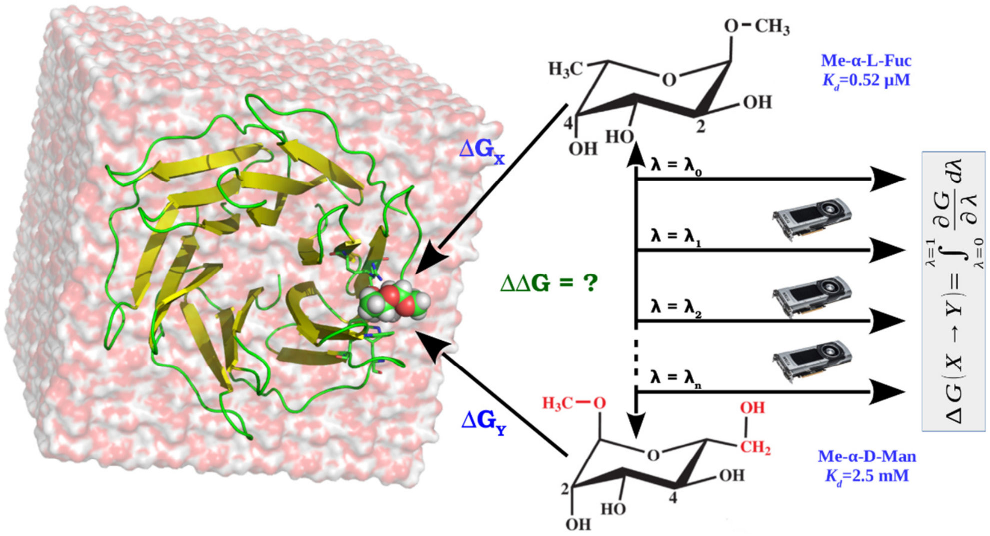


**Figure 10.** Evaluation of twelve different molecular modelling protocols to predict binding affinities of the dataset of RSL ligands. For each protocol the first and second histograms indicate the determination coefficient $r^2$ and predictive index PI, whereas the third histogram indicates the mean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

## Graphical TOC


$O-CH_3$
$H_3C$
O
OH
4
2
OH
HO
Me-α-L-Fuc
$K_d$=0.52 µM
$\Delta G_X$
$\lambda = \lambda_0$
$\lambda = \lambda_1$
$\lambda = \lambda_2$
$\lambda = \lambda_n$
$\Delta\Delta G$ = ?
$\Delta G(X \rightarrow Y) = \int_{\lambda=0}^{\lambda=1} \frac{\partial G}{\partial \lambda} d\lambda$
$\Delta G_Y$
$H_3C-O$
OH
$CH_2$
O
OH
2
HO
4
OH
Me-α-D-Man
$K_d$=2.5 mM