Analysis of protein-glycan recognition using absolute binding free energies

Sondos Musleh,[a,b] Irfan Alibay,[c,d] Philip C. Biggin,[d] and Richard A. Bryce[a,*]

[a] *Division of Pharmacy and Optometry, The University of Manchester, Manchester, M13 9PT, UK.*

[b] *Department of Medicinal Chemistry and Pharmacognosy, Faculty of Pharmacy, Jordan University of Science and Technology, P.O. Box 3030, Irbid, 22110, Jordan*

[c] *Open Free Energy, Open Molecular Software Foundation, Davis, California 95616, United States*

[d] *Structural Bioinformatics and Computational Biochemistry, Department of Biochemistry, The University of Oxford, South Parks Road, Oxford, OX1 3QU, UK.*

Corresponding Author

*Richard Bryce, Division of Pharmacy and Optometry, School of Health Sciences, University of Manchester, Manchester, M13 9PT, U.K. Email: R.A.Bryce@manchester.ac.uk, Tel: (0)161-275-8345, Fax: (0)161-275-2481; ORCID 0000-0002-8145-2345

Abstract

Carbohydrates are key biological mediators of molecular recognition and signalling processes. In this work, we explore the ability of absolute binding free energy (ABFE) calculations to predict the affinities of a set of five related carbohydrate ligands for the lectin protein, concanavalin A, ranging from 27-atom monosaccharides to a 120-atom complex-type N-linked glycan core pentasaccharide. ABFE calculations quantitatively rank and estimate the affinity of the ligands in relation to microcalorimetry, with a mean signed error in binding free energy of -0.63 ± 0.04 kcal/mol. Consequently, the diminished binding efficiencies of the larger carbohydrate ligands are closely reproduced: the ligand efficiency values from isothermal titration calorimetry for the glycan core pentasaccharide and its constituent trisaccharide and monosaccharide compounds are respectively -0.14 ± 0.00, -0.22 ± 0.00 and -0.41 ± 0.00 kcal/mol per heavy atom. ABFE calculations predict these ligand efficiencies to be -0.14 ± 0.02, -0.24 ± 0.03 and -0.46 ± 0.06 kcal/mol per heavy atom respectively. Consequently, the ABFE method correctly identifies the high affinity of the key anchoring mannose residue and the negligible contribution to binding of both β-GlcNAc arms of the pentasaccharide. While challenges remain in sampling the conformation and interactions of these polar, flexible and weakly bound ligands, we nevertheless here find that the ABFE method performs well. The method shows excellent promise as a quantitative tool for predicting and deconvoluting carbohydrate-protein interactions, with potential application to design of therapeutics, vaccines and diagnostics.

keywords: concanavalin A, carbohydrate-protein interactions, N-linked glycans, alchemical binding free energy calculations, absolute binding free energy calculations, absolute solvation free energy calculations.

1. Introduction

Carbohydrates serve a number of important biological functions, as energy stores, structural elements, and ligands in a range of recognition processes, including cell-cell and cell-pathogen interactions.[1] Targeting of carbohydrate-mediated cell-pathogen interactions is a route to the development of small molecule therapeutics, for example the anti-flu neuraminidase inhibitor, zanamivir,[2] and the anti-diabetic glucosidase inhibitor, miglitol;[3] and vaccines, such as those based on the bacterial polysaccharides of *Streptococcus pneumoniae*[4] and *H. influenza*.[5] Interestingly, glycosylation of proteins can assist pathogens in evading the host immune response but also play a role in stabilising functional states of the protein, as in the case of the spike protein of SARS-Cov-2, where a N-glycan chain at Asn343 on the spike protein was found to facilitate opening of its receptor binding domain.[6]

To guide the design of glycan-related therapeutics, diagnostics and vaccines, the ability to decipher the structure-activity relationship of a carbohydrate for its receptor protein is key. Computational tools are well placed to analyze carbohydrate-protein interactions in atomistic detail, furnishing energetic components and residue contributions to binding not readily accessible to experiment.[7] Methods to compute binding free energies from end-point simulations[8-11] or alchemically[12, 13] have achieved some success in accurate prediction of carbohydrate-protein affinities. For example, a relative binding free energies (RBFE) approach was applied to *R. Solanacearum* lectin, ranking ten of its monosaccharide ligands with a mean absolute error (MAE) of 1.1 ± 0.1 kcal/mol, including correct prediction of the anomeric preference of D-glucose (Glc) and D-mannose (Man).[13] The RBFE method involves alchemically transforming one ligand into another when protein-bound and unbound; the approach is most suited to studying differences in binding of closely related ligand structures,[14, 15] such as comparing monosaccharide anomers or other epimers, due to the need to keep a

conserved common core between end states. Despite recent advances in RBFE methodology,[15, 16] it is still not entirely straightforward to capture the free energy consequences of very large differences in structure, for example when comparing carbohydrate ligands that differ in the number of saccharide residues.

However, recent advances in computing absolute binding free energies (ABFEs), by transforming the ligand into a non-interacting species when bound and when in solution, has enabled some success in reliably estimating ΔG_{bind} for a range of disparate druglike ligand structures.[17-19] A recent meta-analysis studied 853 cases of ABFE calculations of protein-ligand affinities, and found a mean unsigned error in free energy below 3 kcal/mol was achieved in 87% of cases, with a MUE of 1.58 kcal/mol.[20] Carbohydrate ligands are not particularly druglike, though, being considerably higher in complexity than small organic molecules: typically, they are larger and more polar, especially in their oligomeric linear or branched forms, with numerous stereogenic centres and rotatable bonds.[21] Nevertheless, application of ABFEs to computing the protein binding free energies of monosaccharides,[22, 23] disaccharides[22] and, in one case, a trisaccharide[24] has proved encouraging, yielding deviations of 1 – 3 kcal/mol from experiment for these systems.

In this study, we evaluate the ability of ABFE calculations to predict the binding affinities of five carbohydrate substrates to the protein Concanavalin A (Con A). Con A is a glucose/mannose-binding lectin derived from the jack bean (*Canavalia ensiformis*); its carbohydrate complexes have been well-characterized by calorimetry and crystallography and comprise a useful test set for assessing methods for computation of binding affinities.[8-10, 12] Ligands **1** – **5** (Figure 1a) are of increasing complexity, ranging from 27-atom monosaccharides to a 120-atom pentasaccharide.

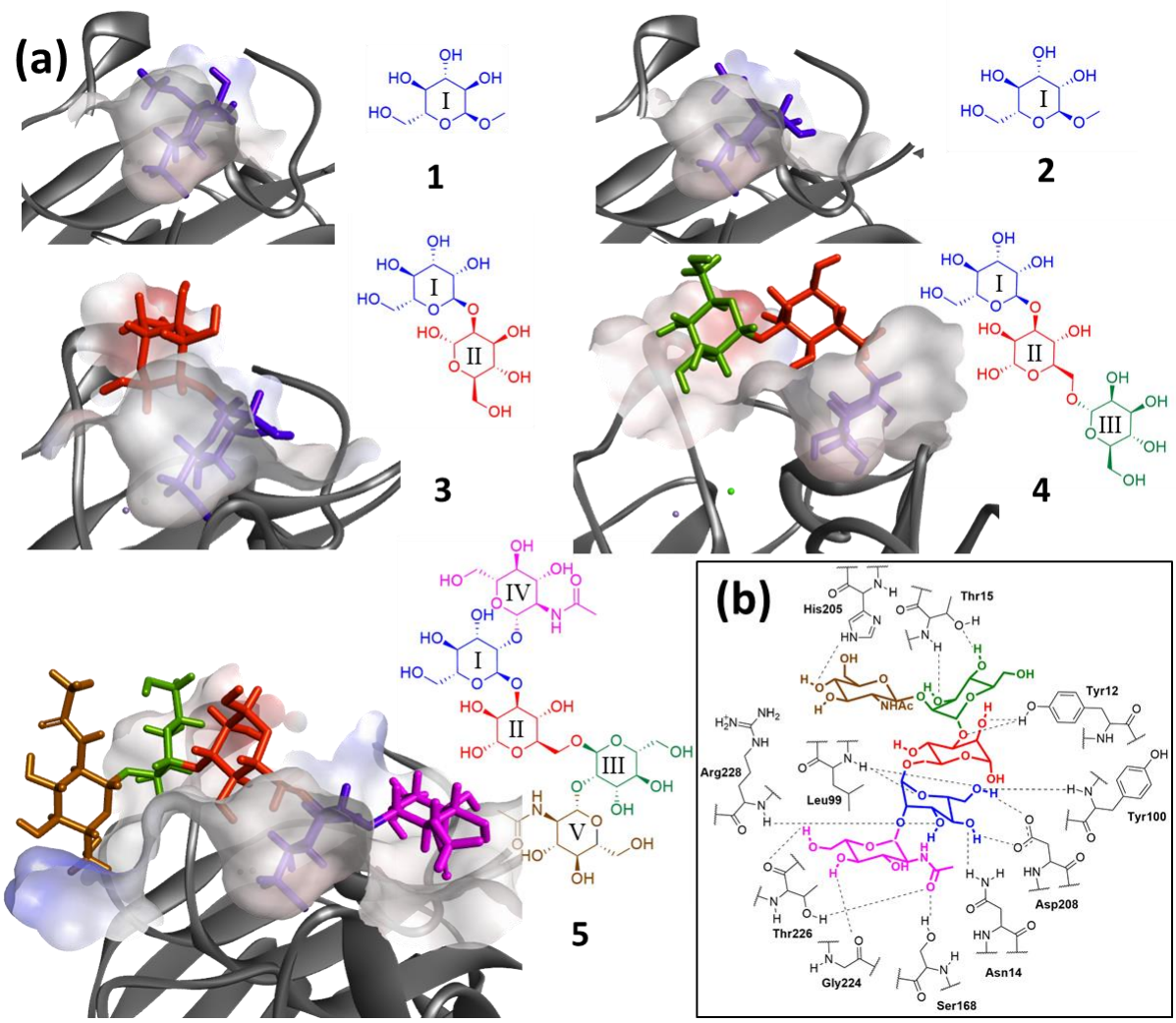

Figure 1 (a) Ligands **1** – **5** and their X-ray bound pose to Con A; these ligands are **(1)** *α*-MeOGlc, **(2)** *α*-MeOMan, **(3)** Man-*α*-(1→2)-Man-*α*-OMe, **(4)** Man-*α*-(1→6)-[Man-*α*-(1→3)]-mannose, and **(5)** *β*-GlcNAc-(1→2)-*α*-Man-(1→3)-[*β*-GlcNAc-(1→2)-*α*-Man-(1→6)]-Man. (b) Polar interactions formed by ligand **5** with Con A in X-ray structure. The high affinity monosaccharide binding site is where the blue ring binds. Color coding used here is not related to glycan symbol nomenclature color coding.[25, 26]

Pentasaccharide **5** has the sequence *β*-GlcNAc-(1→2)-*α*-Man-(1→3)-[*β*-GlcNAc-(1→2)-*α*-Man-(1→6)]-Man and is a common motif of complex-type N-linked glycans. Given its molecular weight of 910 Da, twice that of a large druglike molecule, and its rather weak binding affinity to Con A of -8.38 kcal/mol, ligand **5** is a particularly challenging case for ABFE calculations. Here, we assess the performance of ABFE calculations in recovering the

structure-activity relationship of pentasaccharide **5** to its constituent trisaccharide, Man-α-(1→6)-[Man-α-(1→3)]-mannose **4**; to a disaccharide, Man-*α*-(1→2)-Man-*α*-OMe **3**; and to monosaccharide ligand *α*-MeOMan **2** and its epimer *α*-MeOGlc **1** (Figure 1a).[27] In their crystal structures with Con A, the residues of ligands **1** – **5** occupy the shallow lectin binding groove to varying degrees (Figure 1a). However, for all five ligands, the high affinity mannose binding site within the groove is occupied, lined by the amino acid residues Asn14, Leu99, Tyr100, Asp208 and Arg228 (blue ring I for ligands **1** - **5**, Figure 1b).

2. Materials and Methods

2.1 System preparation and simulation details

Initial models for Con A in complex with ligands **1** - **5** were constructed based on the available X-ray structures, with respective PDB entry codes and resolutions of 1GIC (2.00 Å), 5CNA (2.00 Å), 1I3H (1.20 Å), 1CVN (2.30 Å) and 1TEI (2.70 Å).[28-32] A single subunit of Con A, which can exist as a dimer or tetramer, was retained for simulations. Protonation and tautomeric states were assigned using MOE 2020.09 consistent with physiological pH (pH = 7).[33] All crystal waters for this monomer were kept, including the conserved bound water molecule of Con A that is important for the protein's interaction with ligands **4** and **5**.

Parameters for Con A and its carbohydrate ligands were assigned using the CHARMM36-feb2021 force field via the CHARMM-GUI tool.[34-38] Parameters for the Mn^{2+} ion of Con A were modelled based on CHARMM calcium ion parameters, as adopted elsewhere,[39] given the same charge, coordination pattern, and similar size. All molecular dynamics simulations used the GROMACS 2021.5 software package.[40] The systems were neutralized with sodium ions

and solvated 15 Å beyond the complex using a truncated octahedron with TIP3P water.[41] The resulting systems contained ~19,000 – 22,000 water molecules.

2.2 Absolute free energy calculations

The thermodynamic pathway used to calculate the ABFEs follows the protocol of Aldeghi *et al.*[19, 42, 43] Namely, following equilibration, a partial decoupling scheme is employed to follow the alchemical path from a fully interacting protein-carbohydrate complex to a carbohydrate ligand in solution (Figure S1). This partial decoupling scheme involves annihilating ligand partial charges through 11 windows spaced at λ intervals of 0.1 from each other. This is then followed by 21 van der Waals decoupling windows spaced at 0.05 λ intervals. A soft-core potential for decoupled Van der Waals interactions was used.[44, 45] Additionally, to restrict ligand motion in the complex, an orientational restraint, as defined by Boresch *et al.*[46] was employed, and derived using the MDRestraintsGenerator code.[43, 47] This restraint was applied over 12 windows in the complex decoupling phase. In the solvent phase, the influence of this restraint was accounted for analytically;[46] 31 windows were applied to decouple the ligand from solvent. Therefore, in total, each ABFE calculation corresponded to 75 window simulations.

Simulations were performed using a stochastic leapfrog integrator[48] and a 2 fs time step. The temperature was controlled by Langevin dynamics,[49] with friction constant of 1.00 ps^{-1}. LINCS was applied to constrain bonds involving hydrogen, while water molecules were constrained with the SETTLE algorithm.[50] Periodic boundary conditions were used, with long-range electrostatic and van der Waals interactions treated via the particle mesh Ewald and twin range cut-off schemes respectively,[51, 52] using a short-range cutoff value of 12 Å and switching

distance of 10.0 Å. Coordinates were stored every 2 ps while the free energies were calculated every 200 fs.

At each λ window, the systems were energy minimized and then heated sequentially from 0 to 298 K over 700 ps under NVT conditions, with a restraint of 1000 kJ/(mol nm^2) on all systems' atoms. The systems were then equilibrated under NPT conditions of 1 atm and 298 K in four stages, applying restraints of 1000, 500, and 100 kJ/(mol nm^2) in the first three stages respectively, for 200 ps 200 ps and 300 ps and without restraints in the final stage of 600 ps. The Berendsen barostat[53] was applied in the first three stages and the Parrinello-Rahman barostat [54, 55] in the final stage. Following this, a 20 ns NPT production MD simulation was performed at the given λ.

Using the above simulation protocol, five replica ABFE calculations were obtained for each protein-ligand complex, using independently equilibrated bound poses. For each replica, the initial structure is a frame taken from the preliminary 10 ns MD simulation and represents the structure that is closest to the mean bond, angle and dihedral values of the restraint used in the ABFE calculation. For the solvent leg, the ligand was extracted from the frame then solvated and the solvation free energy calculation was carried out. Estimates of the binding free energies were calculated using the multistate Bennett acceptance ratio (MBAR)[56] via *alchemical-analysis.py*,[57] where the first 2 ns from each production window were excluded from the analysis as extra equilibration time. The protein-ligand binding free energy is reported as the average ABFE over these replicas, with the associated statistical uncertainty taken as the standard deviation.

In addition to computing absolute free energies of binding and solvation, unbiased MD simulations of each complex were run for 500 ns under NPT conditions for ligands **1** - **5**, following the same set up and equilibration protocol over 2 ns discussed earlier. For hydrogen bond analysis, *gmx hbond* routine was employed, where a heavy atom-heavy atom distance and angle cut-off of 3.5 Å and 30° was used respectively. Throughout the manuscript, figures were generated using Discovery Studio 2015 (BIOVIA Software Inc.)[58] and ChemDraw© Ultra 12.0.2.

3. Results and Discussion

3.1 Estimates of absolute and relative binding free energies from ABFE

The absolute binding free energies of carbohydrate ligands **1** - **5** to the lectin protein, Con A, were computed using thermodynamic integration with electrostatic decoupling, based on the available crystal structures of the five complexes. From isothermal titration calorimetry (ITC), the measured binding affinities range from -4.49 kcal/mol for **1**, to -8.38 kcal/mol for **5** (Table 1).[27] The predicted ΔG_{bind} values from ABFE calculations spanned a very similar range to experiment, from -5.15 for **1** to -8.97 kcal/mol for **5** (Table 1). There was a modest systematic overestimation of affinity, with a mean signed error over the five ligands of -0.63 kcal/mol (Table 1). Regarding variation across replicas for a given ligand, the highest sampling error was for the largest ligand, **5**, with a standard deviation of 1.13 kcal/mol (Tables 1 and S1). In accord with the quantitative agreement in binding free energies for **1** - **5**, an accurate estimate of ligand efficiency (LE) is also obtained, with a mean signed error between calculation and experiment of -0.03 kcal/mol per heavy atom (Table 1).

Table 1 Standard binding free energies, ΔG_{bind} (in kcal/mol) for carbohydrate ligands **1** - **5** to Con A from ABFE calculations (calc) and isothermal titration microcalorimetry (expt). Calculated values represent mean of five replicate ABFE calculations, each using 75 λ windows of 20 ns width. Ligand efficiency for computed and experimental binding affinities, LE_{calc} and LE_{expt}, respectively, reported as kcal/mol per heavy atom. Errors stated in parentheses, and mean signed error (MSE) given for binding free energies across ligands, and for $LE_{calc-expt}$.

mol	**ΔG_{bind}**		**$\Delta G^{bind}_{calc-expt}$**	**LE_{calc}**	**LE_{expt}**
	calc	**expt**[27]			
1	-5.15 (0.87)[a]	-4.49 (0.03)[b]	-0.66	-0.40 (0.07)	-0.34 (0.00)
2	-6.03 (0.84)	-5.33 (0.03)	-0.70	-0.46 (0.06)	-0.41 (0.00)
3	-7.02 (0.74)	-6.30 (0.02)	-0.72	-0.30 (0.03)	-0.27 (0.00)
4	-8.00 (1.02)	-7.54 (0.05)	-0.46	-0.24 (0.03)	-0.22 (0.00)
5	-8.97 (1.13)	-8.38 (0.08)	-0.59	-0.14 (0.02)	-0.14 (0.00)
MSE			***-0.63***	***-0.03***	

a Calculated errors are standard deviations for five replicate ABFE calculations

b Experimental errors are those reported in reference 27, namely the standard deviation of fit between binding curve from isothermal titration calorimetry and calculated curve obtained with the fitted thermodynamic parameters.

Given the close agreement in computed and observed absolute values of binding free energy, a strong correlation between calculated and experimental binding affinities of the complexes was also found (Figure 2). The associated Pearson coefficient *r* is 1.00 ± 0.00 and Kendall tau coefficient is 1.00, albeit for a data set of only five ligands. The relationship between binding free energy and saccharide structure appears to be reproduced well: for the stepwise progression from ligands **1** to **5**, corresponding in three of the steps to significant differences in ligand structure and size, the MAE in computed $\Delta\Delta G_{bind}$ is 0.11 kcal/mol with respect to experiment (Figure 3). The internal variation in $\Delta\Delta G_{bind}$ estimates (i.e. combined errors for

replica-based estimates reported in Figure 3) is larger in magnitude than this MAE, ranging from 1.12 kcal/mol for **2→3** to 1.52 kcal for **4→5**, with corresponding standard errors of the mean of 0.50 and 0.69 kcal/mol. The predicted $\Delta\Delta G_{bind}$ values from ABFEs considerably improve upon estimates of $\Delta\Delta G_{bind}$ furnished by a recent MM/PBSA-based study of Con A complexes,[10] which obtained a MAE in $\Delta\Delta G_{bind}$ for the same ligand comparisons of 7.28 kcal/mol. Although here a smaller set of ligands are considered, this error in $\Delta\Delta G$ compares well with a MAE of ~1 kcal/mol from lectin RBFE estimates of ten monosaccharides.[13] The associated sampling error in the latter study however is smaller, given the RBFE protocol and the more modest changes in ligand structure studied, with a value of 0.06 kcal/mol.

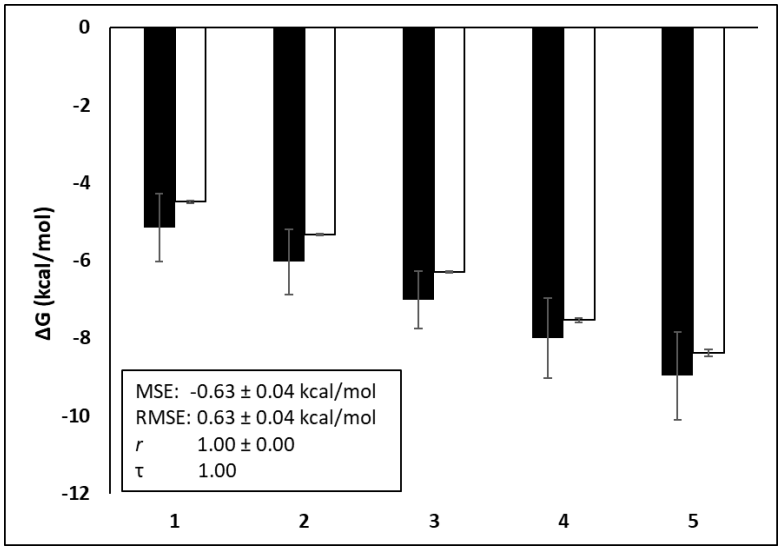

Figure 2 Absolute binding free energies for carbohydrate ligands **1 - 5** to Con A from ABFE calculations (black) and experiment (white). Energies in kcal/mol. Correlation coefficient, MSE, and RMSE calculated from mean estimate values, with error bars obtained as the standard deviation of the means generated through bootstrap resampling (1000 iterations).

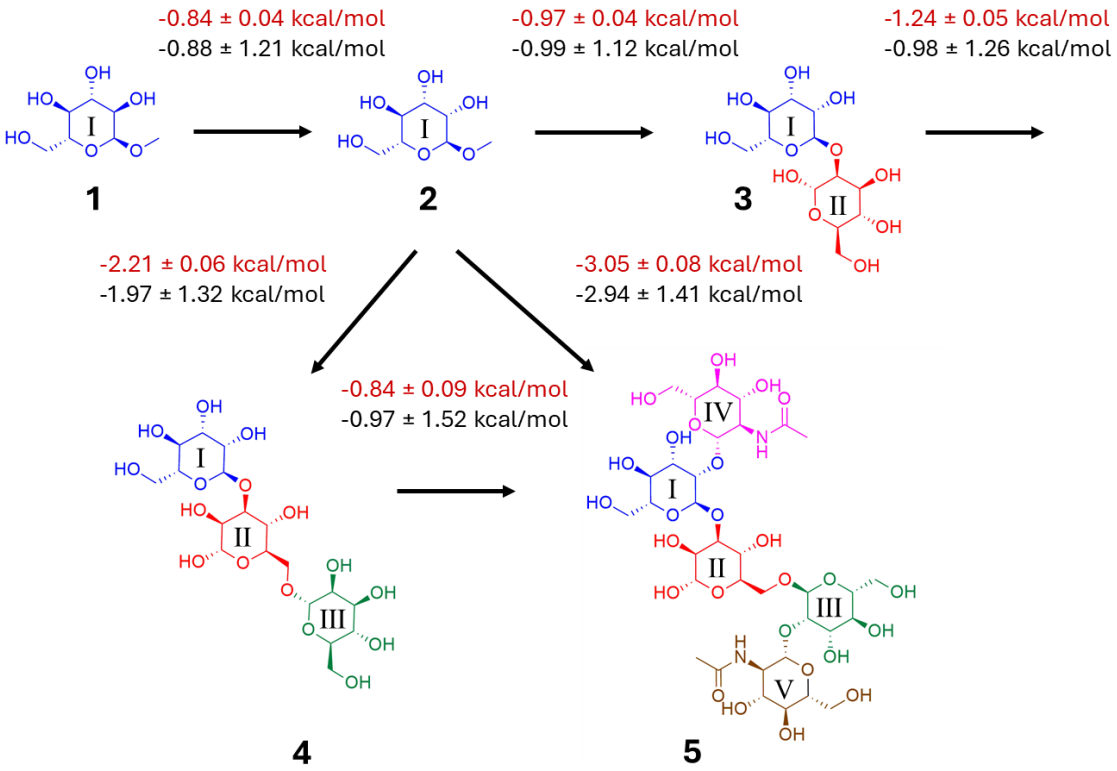

Figure 3 Selected comparative binding free energies of Con A with saccharide ligands **1** – **5**, and corresponding standard deviations across replicates, from experiment (red)[27] and estimated from ABFE calculations (black).

3.2 Structure-activity relationships

We now consider in more detail the ability of ABFE calculations to capture substrate differences in carbohydrate binding to Con A. First, we compare the closely related monosaccharides, α-MeOGlc **1** and α-MeOMan **2** (Figure 3), which differ only by a change in epimeric configuration at the C2 position. Both ligands bind in the same pose to the high affinity monosaccharide binding site of Con A, formed by residues Asn14, Leu99, Tyr100, Asp208 and Arg228 (Figures 1 and S2). The key difference in binding mode is that, for **1**, the equatorial 2-OH projects out into solution, whereas in **2**, the axial 2-OH interacts with protein. Correspondingly, **2** is favoured by 0.84 ± 0.04 kcal/mol over **1** experimentally (Table 1). ABFE calculations provide a $\Delta\Delta G_{bind}$ estimate of 0.88 ± 1.12 kcal/mol (Figure 3); this mean value

indicates the correct preference although with a significant standard deviation. For further insight, we also performed a 500 ns unbiased MD simulation of the complexes of Con A with **1** or **2** bound (Supporting Information, Figures S3 – S4). As expected, these simulations indicate greater overall hydrogen bonding to Con A of **2** over **1** (Figure 4); this mainly arises from 0.46 more hydrogen bonds on average made by the axial 2-OH group of **2** (Table S2).

With a calorimetric ΔG_{bind} of -5.33 kcal/mol, monosaccharide α-MeOMan **2** is the key anchoring residue within larger oligosaccharide forms of complex-type N-linked glycans such as ligand **5**. Consequently, the addition of a second mannose residue at the reducing position of **2**, to give α-(1→2)-linked dimannoside **3**, results in only a modest benefit in ΔG_{bind} by 0.97 ± 0.04 kcal/mol experimentally (Figure 3). Again, ABFE calculations correctly indicate only a small improvement in binding due to this change, with an estimate of 0.99 ± 1.12 kcal/mol. Unbiased MD simulation indicates hydrogen bonding made by the additional ring of **3** (Figures 4 and S5).

The core trimannose ligand, Man-α-(1→6)-[Man-α-(1→3)]-mannose **4**, is experimentally observed to bind with 1.24 ± 0.05 kcal/mol higher affinity than disaccharide **3**. ABFE calculations predict this change to be 0.98 ± 1.26 kcal/mol. Ligand **4** makes favourable hydrogen bonds with Pro13 and Thr15 of Con A (Figure S6); the improved interactions appear facilitated by the change in glycosidic linkage from α-(1→2) for **3** to α-(1→6) for **4** (Figures 1 and S2).

Addition of terminal β-(1→2)-GlcNAc arms to core mannoside **4** yields pentasaccharide **5**. The unusually small increase in observed ΔG_{bind} of 0.84 ± 0.09 kcal/mol accompanying this modification is captured well by the ABFE method, yielding a computed value of 0.97 ± 1.52

kcal/mol (Figure 3). The corresponding reduction in ligand efficiency on proceeding from **4** to **5** is predicted as 0.10 ± 0.04 kcal/mol per heavy atom, compared with an experimental value of 0.08 ± 0.00 (Table 1). This drop in LE is observed despite the additional protein contacts formed by one of the *β*-GlcNAc residues of **5** (ring IV in Figure 3), with amino acid residues Ser168, His205, Gly224, Thr226 and Arg228 (Figure 1b). The second *β*-GlcNAc unit of **5** (ring V, Figure S7) projects out into solution and contributes negligibly to protein-ligand hydrogen bonding (Figure 4). It has been pointed out, however, that the polar interactions of this ring with Con A are not optimal.[8] ABFE calculations appear able to correctly capture the minimal contribution of ring IV to binding affinity of pentasaccharide **5**. Weak hydrogen bonding to Con A by ring IV of **5** is also evidenced by the low population and frequent transitions of hydrogen bond interactions involving this ring over the 500 ns unbiased MD simulation of the **5**/Con A complex (Figure S7).

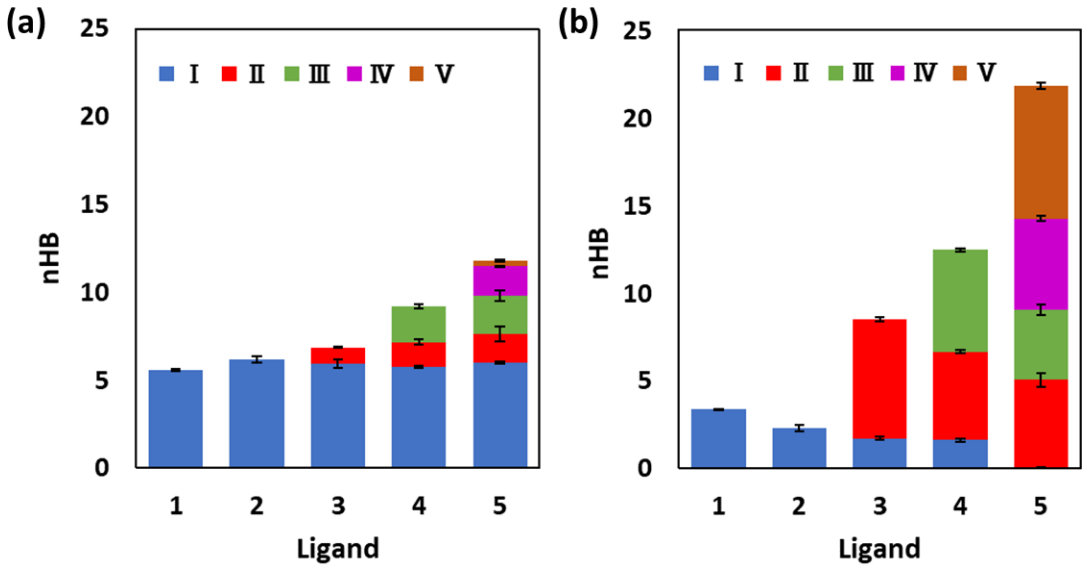

Figure 4 Number of **(a)** protein-ligand and **(b)** solvent-ligand hydrogen bonds, n_{HB}, for rings I - V of ligands **1 - 5**, averaged over 500 ns molecular dynamics simulation. Error bars derived from block averaging.

While the progression from **1** to **5** described above involves only modest changes in experimental binding free energy, larger changes in ΔG_{bind} are also captured well by the ABFE

calculations. For example, the addition of two or four residues to monosaccharide **2**, i.e. **2** → **4** and **2** → **5** result in a reduction in ΔG_{bind} measured by ITC, of 2.21 ± 0.06 and 3.05 ± 0.08 kcal/mol respectively (Figure 3, Table 1). The according ABFE values of 1.97 ± 1.32 and 2.94 ± 1.41 kcal/mol reproduce well the direction and magnitude of these free energy changes. These computed changes for **2** → **4** and **2** → **5** yield LE reductions of 0.16 ± 0.07 and 0.32 ± 0.07 kcal/mol per heavy atom, matching well the corresponding experimental values of 0.14 ± 0.00 and 0.28 ± 0.00 kcal/mol per heavy atom. This level of fidelity in relative free energies and LE from ABFE calculations is comparable with a MAE of ~1 kcal/mol in $\Delta\Delta G_{bind}$ from RBFE estimates of monosaccharide-lectin affinities.[13] In a study of Con A lectin specifically,[10] MM/PBSA-based affinities for ligands including **1** – **5** did not obtain quantitative agreement with experiment: particularly notable is the prediction of **4** → **5** as -12.10 kcal/mol (with a standard error of 4.95 kcal/mol),[10] as opposed to only -0.84 ± 0.09 kcal/mol and -0.97 ± 1.52 kcal/mol from ITC and ABFEs respectively (Figure 3).

3.3 Analysis of errors in ABFE calculations

As discussed above, it appears that ABFE estimates can discriminate the key anchoring residue of pentasaccharide **5** (ring I) from residues which contribute modestly to binding (rings II and III) and those that contribute negligibly to affinity (rings IV and V). While this structure-activity relationship is encouraging, we turn now to comment on potential sources of errors associated with these ABFE estimates, given ligands **1** - **5** are large, flexible, weakly binding ligands. As noted above, the sampling error, estimated through calculating the standard deviation over the five replica ABFE calculations for ligands **1** - **5**, range from 0.74 – 1.13 kcal/mol, with an average of 0.92 kcal/mol (Table 1). These values are considerably higher than the standard deviations associated with the free energies from ITC measurements, which range from 0.02 to 0.08 kcal/mol; these values are very low indeed, when considered in the

context of an experimental reproducibility survey which indicated an experimental root-mean-square error in free energies on average of ~1 kcal/mol.[59] Perhaps unsurprisingly, the highest uncertainty in prediction is for the largest ligand, pentasaccharide **5**, with a standard deviation and range across replicas of 1.13 and 2.76 kcal/mol respectively (Tables 1 and S1). Although initiated from the same crystallographic structure, the five replicas of Con A/**5** equilibrate over the 10 ns preliminary MD simulation to slightly different conformers for each replica (Figure 5a).

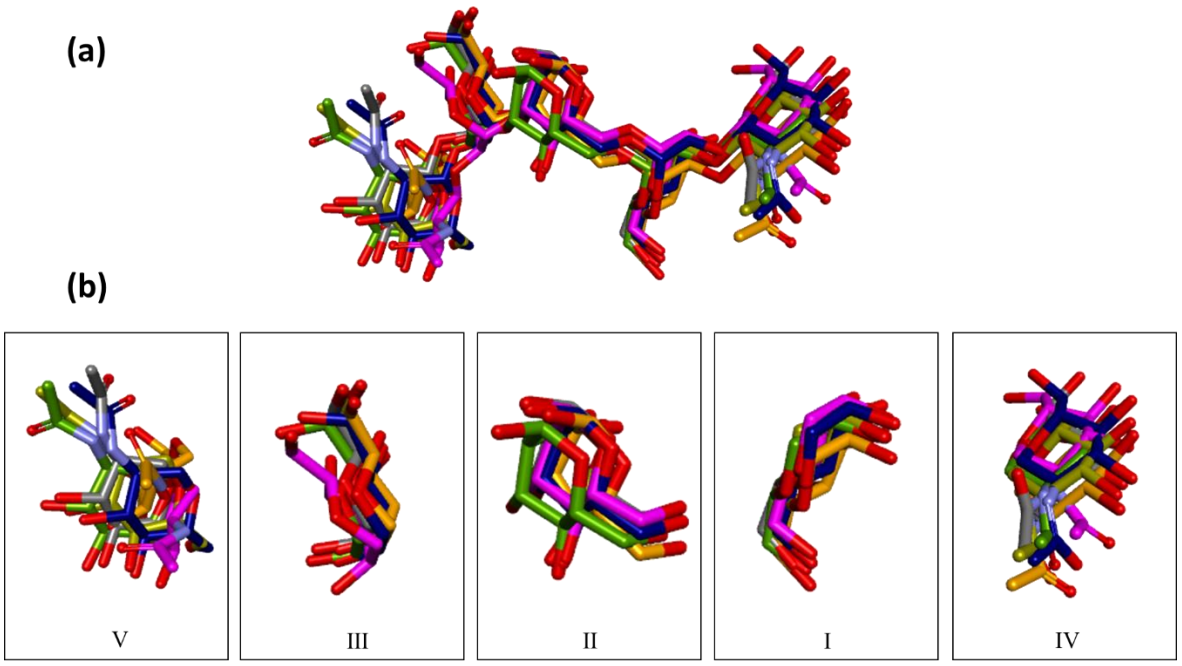

Figure 5 (a) Starting frames of five replicas of ligand **5**, superposed onto the crystal pose based on protein all atoms RMSD. (b) Detail of rings I - V of ligand **5** from this superposition of replicas. Carbon atoms of ligand **5** in replicas *1* - *5* are colored magenta, orange, gold, navy blue, and grey respectively, while those of the crystal structure are colored green. Hydrogen atoms removed for clarity.

To some degree, as might be expected both computationally and experimentally, the lower the LE of the saccharide residue, the larger the structural variation observed in its bound pose: the closest similarity in conformation is found for the anchoring mannose residue, ring I, in the monosaccharide binding site (Figure 5b). Furthermore, the unbiased 500 ns simulation of the

Con A/**5** complex indicated periodic significant changes in ligand pose, as indicated by distance $d_{I,V}$ between the ring O5 atoms of the terminal GlcNAc residues (Figure 6a); and by snapshots taken at 149 and 490 ns superposed onto the crystal structure (Figures 6b and 6c). Interestingly, the key anchoring mannose (ring I) remained firmly attached to Con A throughout the trajectory, as anticipated from its network of favorable interactions with the protein.

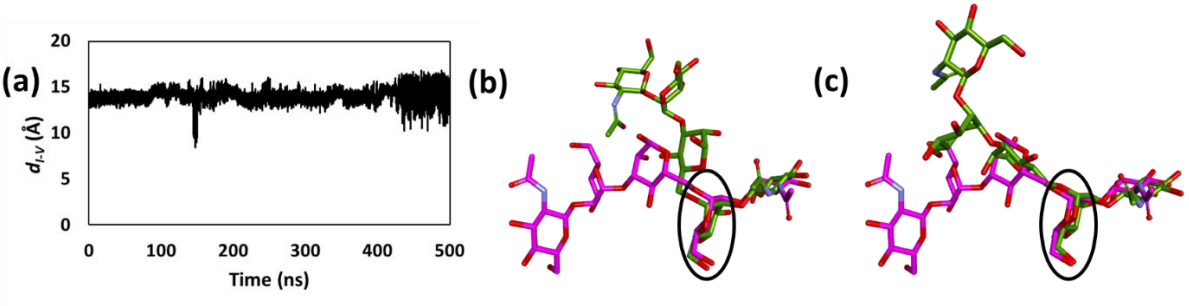

Figure 6 (a) Time series of distance $d_{I\text{-}V}$ between O5 atoms of terminal GlcNAc residues of ligand **5**. Snapshots (in green) of compound **5** MD simulation at (b) 149 ns, and (c) 490 ns, superposed with crystallographic pose of ligand **5** (in magenta). The well-preserved key anchoring mannose residues are labeled by black circles.

While the range in ΔG_{bind} across replicas is highest for pentasaccharide **5**, with a value of 2.76 kcal/mol, a somewhat lower but non-negligible range of 1.51 to 2.44 kcal/mol is found for ligands **1** - **4** (Figure 7, Table S1). Even for anchoring monosaccharide **2**, with the highest LE of the ligands (Table 1), computed ΔG_{bind} values over the five replicas range from -4.77 for replica *4* to -6.99 kcal/mol for replica *5* (Figure 7; Table S1); this indicates incomplete convergence, which is expected given the flexible nature of these ligands. We observe that the overall bound pose of **2** is well preserved across the starting structures of replicas *1* - *5* (Figure 8); however, some hetereogeneity in Con A amino acid residue orientation and proximity is evident, mainly for Tyr100 and Arg228 (Figure 8). A similar variation in local protein environment is found for the other four ligands in their equilibrated structures (Figure S8).

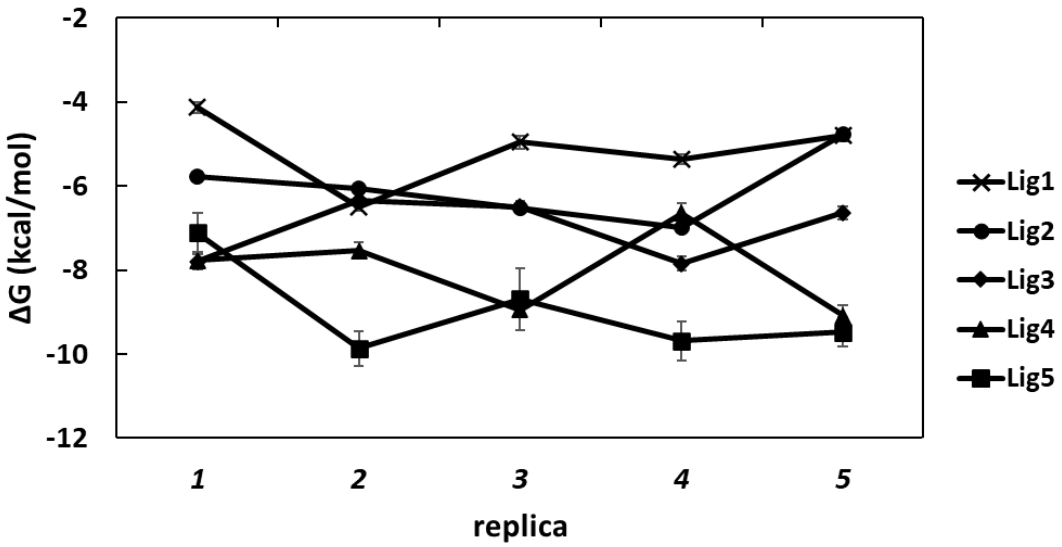

Figure 7 The variation in binding free energies calculated for the five replicas of each of the complexes (in kcal/mol). Error bars from MBAR estimate.

Finally, we also evaluate the impact of the initial crystal structure on the calculated binding free energies. Different initial crystallographic structures of the Con A monomer complex with substrates **2** and **5** were examined. In the preceding calculations, chains D and C were used as the initial structures to compute the ABFE for ligands **2** and **5** respectively, as those chains displayed the fewest geometric outliers in key binding site residues.[29, 32] For comparison, we used chains A and B to calculate the binding free energies of ligands **2** and **5** respectively, employing five replicas per complex (Table S3). This yielded a calculated ABFE of -7.07 ± 0.73 kcal/mol for ligand **2**, with a signed error -1.74 kcal/mol relative to experiment. For ligand **5**, the computed ΔG_{bind} was –10.70 ± 1.37 kcal/mol, with a signed error -2.32 kcal/mol.

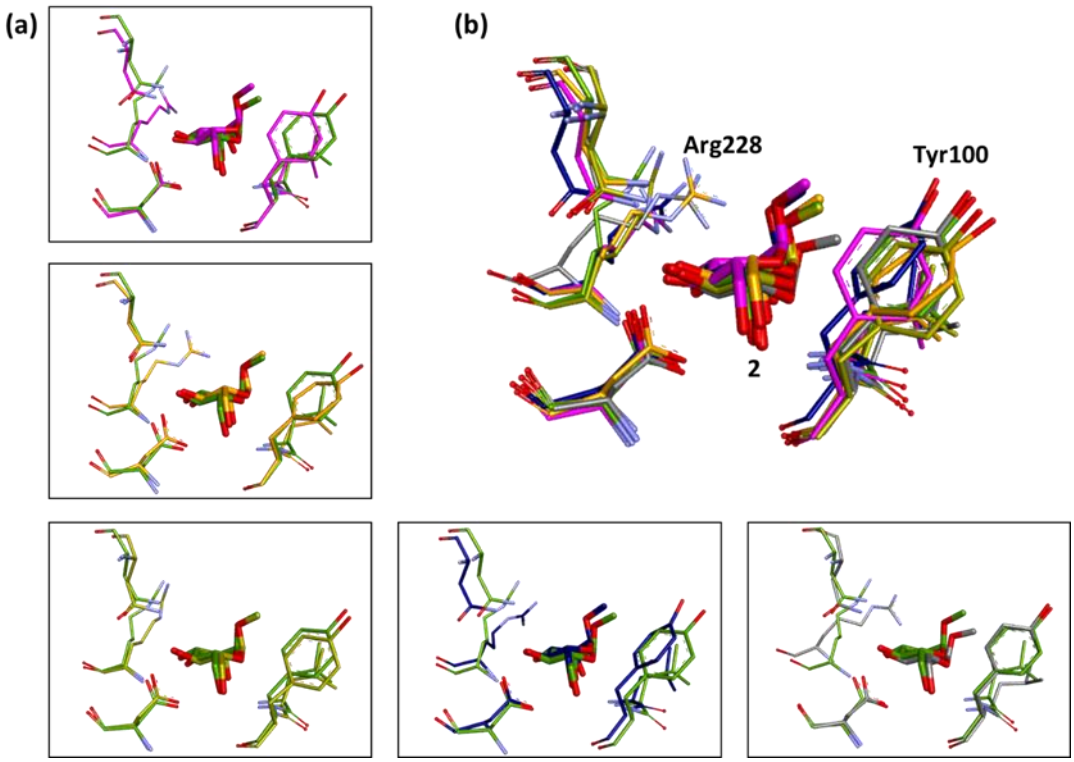

Figure 8 (a) Starting structures of replicas 1 - 5 for ligand **2** each superposed to crystal structure. (b) Starting structures of replicas *1 - 5* superposed to the crystal structure. Carbon atoms of crystal pose colored green, while replicas *1 - 5* are magenta, orange, gold, navy blue, and grey respectively. Hydrogen atoms omitted for clarity.

These deviations are somewhat higher than found for ABFE estimates based on the higher quality chain D and C structures, where the mean signed errors for ligands **2** and **5** respectively were -0.72 and -0.59 kcal/mol (Table 1). Comparing the X-ray structures of the complexes, the poses of the ligands are well preserved in the corresponding pairs of chains, but subtle differences were observed for the crystallographic coordinates of the interacting amino acids, particularly Arg 228 and Tyr100 (Figure 9). Indeed, the Arg228 side-chain did not appear to fit well to the electron density in either chains A and B of ligands **2** and **5** complexes respectively. We note that the uncertainty in the calculated ABFEs, estimated as the standard deviation over the five replica ABFE calculations, gets smaller as the resolution of the

crystallographic structure improves (Table S4); however, the increased complexity of sampling the interactions of larger ligands undoubtedly also plays a role in determining the ABFE error.

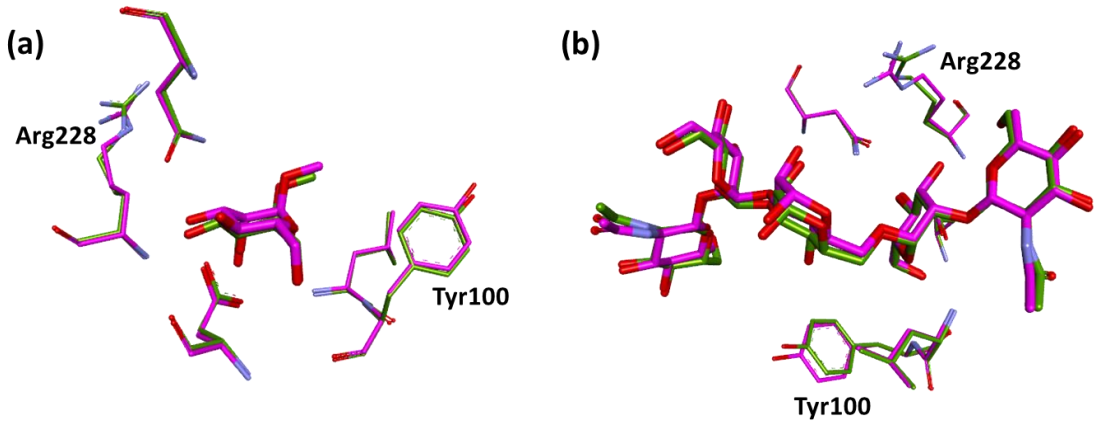

Figure 9 (a) Crystal structures of chain A (green) and D (magenta) of ligand **2**/Con A complex superposed; side-chain torsion angles differ by up to 6° for Arg228 and 20° for Tyr100. (b) Crystal structures of chain B (green) and C (magenta) of ligand **5**/Con A complex superposed; side-chain torsion angles differ by up to 27° for Arg228 and 7° for Tyr100.

4. Conclusions

Here, we have demonstrated the ability of an ABFE approach to quantitatively predict the absolute binding free energies of monosaccharide, disaccharide, trisaccharide and pentasaccharide ligands to the lectin, Con A. The errors in the relative affinity of these ligands are 0.11 kcal/mol from ABFE calculations, which are considerably improved on those obtained from MM/PBSA-based calculations for these complexes (7.28 kcal/mol) and are comparable in accuracy with free energy estimates for much smaller changes in (mono)saccharide structure using a RBFE method.[13] However, for single or hybrid topology RBFE approaches, which typically rely upon close analogy in binding pose and the avoidance of ring breaking or formation,[15] computing a two-residue deletion between ligand **5** and **4** would be challenging to converge.

The error in absolute free energy of 0.63 kcal/mol found here is smaller than that observed in preceding ABFE studies for smaller carbohydrate ligands, where deviations from experiment ranged from 1 – 3 kcal/mol.[13, 23, 24] In this comparison, we note the effect of force field choice on estimated ΔG_{bind}: for a study of ABFEs for up to seven monosaccharide/disaccharide-protein systems,[22] use of the CHARMM36 force field for ligand and protein[60] provided an improved correlation with experimental binding free energies over an AMBER GLYCAM06-j/ff99SB-ILDN potential,[61] although deviations in absolute error in these two calculations were on the order of 3 kcal/mol. In both cases, the TIP3P water model was used, as is the case in the current study. In this work, we use the CHARMM26-feb2021 force field for carbohydrate and protein, obtaining an error of 0.63 kcal/mol, which lies at the lower end of the expected range in error for ligand-protein ABFEs: a recent meta-analysis of 853 protein-ligand absolute binding free energy calculations from 34 different research groups found a MUE of $1.58^{1.83}_{1.34}$ kcal/mol with a variance of $1.71^{2.06}_{1.37}$ kcal^2/mol^2 (here indicating the 95% confidence intervals).[20] These calculations, which include both alchemical and geometrical approaches (e.g. attach-pull-release and confine-and-release[62]), indicate both accuracy and precision in druglike ligand-protein ABFEs.

As large, flexible ligands with low ligand efficiency, adequately sampling the bound poses of carbohydrate ligands is indeed challenging. Ligand flexibility and large binding site conformational space have been noted as challenges for ABFE protocols previously, for example in computing absolute binding affinities for ligands of the MCL-1 receptor.[63, 64] Conformational sampling via several independent replicas, of tens of nanoseconds per window, to obtain more reliable free energy estimates appears to be of particular importance for carbohydrates, which typically bind to their protein receptors with modest affinities. For the

pentasaccharide ligand **5** in particular, we observe long time scale conformational events over a 500 ns MD trajectory; adequately sampling those states within the numerous λ windows of an ABFE calculation is a formidable prospect. We also find that subtle differences in ligand-protein pose appear to contribute to some variation in calculated free energy of binding. This observation was true both for intra-replica variation in ΔG_{bind} estimates for a given chain, and for differences in replica predictions for different choices of chain coordinates.

Nonetheless, the ability of the ABFE method to furnish predictive estimates of $\Delta\Delta G_{bind}$ for carbohydrate ligands such as **5** and its constituent residues, represented by ligands **2** and **4**, is valuable, offering a powerful tool for quantitatively dissecting the residue-wise affinity of oligosaccharide-protein complexes and identification of binding hot spots. These significant changes in LE are closely reproduced: the experimental LEs for monosaccharide **2** (-0.41 ± 0.00 kcal/mol per heavy atom), trisaccharide **4** (-0.22± 0.00) and pentasaccharide **5** (-0.14± 0.00) are predicted as -0.46 ± 0.06, -0.24 ± 0.03 and -0.14 ± 0.02 kcal/mol per heavy atom respectively. Thus, the ABFE approach successfully discerned the anchoring role of the key mannose residue (ring I) from two moderately bound mannose residues (rings II and III); and from a bound *β*-GlcNAc residue (ring IV) that makes crystallographic contacts with Con A but provides almost no benefit in binding free energy. The identification of hot spots can provide valuable insight into mechanism but also inform subsequent ligand design, for the development of potential new therapeutics, diagnostics and vaccines.

Supporting Information. Structural and energetic analyses of carbohydrate-Con A interactions from MD simulations and ABFE calculations.

Acknowledgments

This work made use of the facilities of the N8 Centre of Excellence in Computationally Intensive Research (N8 CIR) provided and funded by the N8 research partnership and EPSRC (Grant No. EP/T022167/1). The Centre is co-ordinated by the Universities of Durham, Manchester, and York. SM thanks Jordan University of Science and Technology, Irbid for their financial sponsorship. Assistance from Research IT and the use of the Computational Shared Facility at the University of Manchester is also acknowledged.

References

1. Dwek, R. A., Glycobiology: Toward Understanding the Function of Sugars. *Chem Rev* **1996**, 96, 683-720.

2. Dreitlein, W. B.; Maratos, J.; Brocavich, J., Zanamivir and oseltamivir: two new options for the treatment and prevention of influenza. *Clin Therap* **2001**, 23, 327-55.

3. Kajimoto, T.; Node, M., Inhibitors Against Glycosidases as Medicines. *Curr Top Med Chem* **2009**, 9, 13-33.

4. Jennings, H., Capsular polysaccharides as vaccine candidates. *Bacterial Capsules* **1990**, 97-127.

5. Jennings, H., Further approaches for optimizing polysaccharide-protein conjugate vaccines for prevention of invasive bacterial disease. *J Infect Dis* **1992**, 165 Suppl 1, S156-9.

6. Sztain, T.; Ahn, S.-H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; Seitz, E.; McCool, R. S.; Kearns, F. L.; Acosta-Reyes, F.; Maji, S.; Mashayekhi, G.; McCammon, J. A.; Ourmazd, A.; Frank, J.; McLellan, J. S.; Chong, L. T.; Amaro, R. E., A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nature Chem* **2021**, 13, 963-968.

7. DeMarco, M. L.; Woods, R. J., Structural glycobiology: A game of snakes and ladders. *Glycobiology* **2008**, 18, 426-440.

8. 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, 1373-1388.

9. Hadden, J. A.; Tessier, M. B.; Fadda, E.; Woods, R. J. Calculating Binding Free Energies for Protein–Carbohydrate Complexes. In *Glycoinformatics*, Lütteke, T.; Frank, M., Eds.; Springer New York: New York, NY, 2015, pp 431-465.

10. Zlotnikov, I. D.; Kudryashova, E. V., Computer simulation of the Receptor–Ligand Interactions of Mannose Receptor CD206 in Comparison with the Lectin Concanavalin A Model. *Biochemistry (Moscow)* **2022**, 87, 54-69.

11. 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, 2340-2350.

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, 16933-16942.

13. Mishra, S. K.; Calabró, G.; Loeffler, H. H.; Michel, J.; Koča, J., Evaluation of Selected Classical Force Fields for Alchemical Binding Free Energy Calculations of Protein-Carbohydrate Complexes. *J Chem Theory Comput* **2015**, 11, 3333-3345.

14. 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**, 137, 2695-2703.

15. Cournia, Z.; Allen, B.; Sherman, W., Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J Chem Inf Model* **2017**, 57, 2911-2937.

16. Song, L. F.; Merz, K. M., Jr., Evolution of Alchemical Free Energy Methods in Drug Discovery. *J Chem Inf Model* **2020**, 60, 5308-5318.

17. Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A., Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. *J Mol Biol* **2007**, 371, 1118-1134.

18. Liang, L.; Liu, H.; Xing, G.; Deng, C.; Hua, Y.; Gu, R.; Lu, T.; Chen, Y.; Zhang, Y., Accurate calculation of absolute free energy of binding for SHP2 allosteric inhibitors using free energy perturbation. *Phys Chem Chem Phys* **2022**, 24, 9904-9920.

19. Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C., Accurate calculation of the absolute free energy of binding for drug molecules. *Chem Sci* **2016**, 7, 207-218.

20. Fu, H.; Zhou, Y.; Jing, X.; Shao, X.; Cai, W., Meta-Analysis Reveals That Absolute Binding Free-Energy Calculations Approach Chemical Accuracy. *J Med Chem* **2022**, 65, 12970-12978.

21. Alibay, I.; Burusco, K. K.; Bruce, N. J.; Bryce, R. A., Identification of Rare Lewis Oligosaccharide Conformers in Aqueous Solution Using Enhanced Sampling Molecular Dynamics. *J Phys Chem B* **2018**, 122, 2462-2474.

22. Plazinska, A.; Plazinski, W., Comparison of Carbohydrate Force Fields in Molecular Dynamics Simulations of Protein–Carbohydrate Complexes. *J Chem Theory Comput* **2021**, 17, 2575-2585.

23. Liu, W.; Jia, X.; Wang, M.; Li, P.; Wang, X.; Hu, W.; Zheng, J.; Mei, Y., Calculations of the absolute binding free energies for Ralstonia solanacearum lectins bound with methyl-α-

l-fucoside at molecular mechanical and quantum mechanical/molecular mechanical levels. *RSC Adv* **2017**, 7, 38570-38580.

24. 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* **2011**, 50, 10530-10539.

25. Varki, A.; Cummings, R. D.; Aebi, M.; Packer, N. H.; Seeberger, P. H.; Esko, J. D.; Stanley, P.; Hart, G.; Darvill, A.; Kinoshita, T.; Prestegard, J. J.; Schnaar, R. L.; Freeze, H. H.; Marth, J. D.; Bertozzi, C. R.; Etzler, M. E.; Frank, M.; Vliegenthart, J. F.; Lütteke, T.; Perez, S.; Bolton, E.; Rudd, P.; Paulson, J.; Kanehisa, M.; Toukach, P.; Aoki-Kinoshita, K. F.; Dell, A.; Narimatsu, H.; York, W.; Taniguchi, N.; Kornfeld, S., Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **2015**, 25, 1323-1324.

26. Neelamegham, S.; Aoki-Kinoshita, K.; Bolton, E.; Frank, M.; Lisacek, F.; Lütteke, T.; O'Boyle, N.; Packer, N. H.; Stanley, P.; Toukach, P.; Varki, A.; Woods, R. J.; Group, T. S. D., Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology* **2019**, 29, 620-624.

27. Mandal, D. K.; Kishore, N.; Brewer, C. F., Thermodynamics of Lectin-Carbohydrate Interactions. Titration Microcalorimetry Measurements of the Binding of N-Linked Carbohydrates and Ovalbumin to Concanavalin A. *Biochemistry* **1994**, 33, 1149-1156.

28. Harrop, S. J.; Helliwell, J. R.; Wan, T. C.; Kalb, A. J.; Tong, L.; Yariv, J., Structure solution of a cubic crystal of concanavalin A complexed with methyl alpha-D-glucopyranoside. *Acta Cryst D* **1996**, 52, 143-55.

29. Naismith, J. H.; Emmerich, C.; Habash, J.; Harrop, S. J.; Helliwell, J. R.; Hunter, W. N.; Raftery, J.; Kalb, A. J.; Yariv, J., Refined structure of concanavalin A complexed with methyl alpha-D-mannopyranoside at 2.0 A resolution and comparison with the saccharide-free structure. *Acta Cryst D* **1994**, 50, 847-58.

30. Sanders, D. A.; Moothoo, D. N.; Raftery, J.; Howard, A. J.; Helliwell, J. R.; Naismith, J. H., The 1.2 A resolution structure of the Con A-dimannose complex. *J Mol Biol* **2001**, 310, 875-84.

31. Naismith, J. H.; Field, R. A., Structural basis of trimannoside recognition by concanavalin A. *J Biol Chem* **1996**, 271, 972-6.

32. Moothoo, D. N.; Naismith, J. H., Concanavalin A distorts the beta-GlcNAc-(1-->2)-Man linkage of beta-GlcNAc-(1-->2)-alpha-Man-(1-->3)-[beta-GlcNAc-(1-->2)-alpha-Man-(1-->6)]-Man upon binding. *Glycobiology* **1998**, 8, 173-81.

33. Chemical Computing Group ULC *Molecular Operating Environment (MOE)*, 2020.09; Montreal, QC, Canada, H3A 2R7, 2020.

34. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem* **2008**, 29, 1859-1865.

35. Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W., CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* **2016**, 12, 405-413.

36. Park, S.-J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W., CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **2019**, 29, 320-331.

37. Jo, S.; Song, K. C.; Desaire, H.; MacKerell Jr., A. D.; Im, W., Glycan reader: Automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *J Comput Chem* **2011**, 32, 3135-3141.

38. Park, S.-J.; Lee, J.; Patel, D. S.; Ma, H.; Lee, H. S.; Jo, S.; Im, W., Glycan Reader is improved to recognize most sugar types and chemical modifications in the Protein Data Bank. *Bioinf* **2017**, 33, 3051-3057.

39. Fadda, E.; Woods, R. J., On the Role of Water Models in Quantifying the Binding Free Energy of Highly Conserved Water Molecules in Proteins: The Case of Concanavalin A. *J Chem Theory Comput* **2011**, 7, 3391-3398.

40. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, 1-2, 19-25.

41. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **1983**, 79, 926-935.

42. Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C., Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations. *J Am Chem Soc* **2017**, 139, 946-957.

43. Alibay, I.; Magarkar, A.; Seeliger, D.; Biggin, P. C., Evaluating the use of absolute binding free energy in the fragment optimisation process. *Commun Chem* **2022**, 5, 105.

44. Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F., Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem Phys Lett* **1994**, 222, 529-539.

45. Pham, T. T.; Shirts, M. R., Identifying low variance pathways for free energy calculations of molecular transformations in solution phase. *J Chem Phys* **2011**, 135, 034114.

46. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M., Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J Phys Chem B* **2003**, 107, 9535-9551.

47. Alibay, I. *IAlibay/MDRestraintsGenerator: v0.2.1*, v0.2.1; Zenodo: 2022.

48. Van Gunsteren, W. F.; Berendsen, H. J. C., A Leap-frog Algorithm for Stochastic Dynamics. *Mol Sim* **1988**, 1, 173-185.

49. Goga, N.; Rzepiela, A. J.; de Vries, A. H.; Marrink, S. J.; Berendsen, H. J. C., Efficient Algorithms for Langevin and DPD Dynamics. *J Chem Theory Comput* **2012**, 8, 3637-3649.

50. Miyamoto, S.; Kollman, P. A., Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* **1992**, 13, 952-962.

51. Fadrná, E.; Hladecková, K.; Koca, J., Long-range electrostatic interactions in molecular dynamics: an endothelin-1 case study. *J Biol Struct Dyn* **2005**, 23, 151-62.

52. Diem, M.; Oostenbrink, C., The Effect of Using a Twin-Range Cutoff Scheme for Nonbonded Interactions: Implications for Force-Field Parametrization? *J Chem Theory Comput* **2020**, 16, 5985-5990.

53. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J Chem Phys* **1984**, 81, 3684-3690.

54. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* **1981**, 52, 7182-7190.

55. Nosé, S.; Klein, M. L., Constant pressure molecular dynamics for molecular systems. *Mol Phys* **1983**, 50, 1055-1076.

56. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* **2008**, 129, 124105.

57. Klimovich, P. V.; Shirts, M. R.; Mobley, D. L., Guidelines for the analysis of free energy calculations. *J Comput-Aid Mol Des* **2015**, 29, 397-411.

58. BIOVIA, D. S. *Discovery Studio*, San Diego: Dassault Systèmes, 2015.

59. Ross, G. A.; Lu, C.; Scarabelli, G.; Albanese, S. K.; Houang, E.; Abel, R.; Harder, E. D.; Wang, L., The maximal and current accuracy of rigorous protein-ligand binding free energy calculations. *Commun Chem* **2023**, 6, 222.

60. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. *J Chem Theory Comput* **2012**, 8, 3257-3273.

61. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct Funct Bioinf* **2010**, 78, 1950-1958.

62. Mobley, D. L.; Chodera, J. D.; Dill, K. A., Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change. *J Chem Theory Comput* **2007**, 3, 1231-1235.

63. Alibay I, M. A., Seeliger D, Biggin P. , Evaluating the use of absolute binding free energy in the fragment optimization process. *ChemRxiv* **2022**.

64. Steinbrecher, T. B.; Dahlgren, M.; Cappel, D.; Lin, T.; Wang, L.; Krilov, G.; Abel, R.; Friesner, R.; Sherman, W., Accurate Binding Free Energy Predictions in Fragment Optimization. *J Chem Inf Model* **2015**, 55, 2411-2420.