Supporting Information

Defining the specificity of carbohydrate-protein interactions by quantifying functional group contributions

*Amika Sood[1], Oksana O. Gerlits[2], Ye Ji[1], Nicolai V. Bovin[3], Leighton Coates[4], Robert J. Woods[1],**

[1]Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, USA

[2]Biology and Soft Matter Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

[3]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997, Russian Federation

[4]Neutron Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

*Author to whom correspondence may be addressed.

Voice: +1-706-542-4454

Email: rwoods@ccrc.uga.edu

Quantification of molecular contributions to affinity

The strengths of the interactions between ECL and its ligands was quantified by performing MM-PB/GBSA energy analyses of the MD simulations. In addition to contributions from direct interactions (van der Waals and electrostatics), the MM-PB/GBSA energies also include estimates of desolvation free energy. Conformational entropies were estimated using three different methods. Firstly, we examined the quasi-harmonic (QH) approach, which derives entropy differences from a covariance analysis of the changes in atomic fluctuations that occur upon ligand binding [1]. Secondly, an analysis of changes in the vibrational normal mode (NM) was performed, which estimates the entropic contributions for binding resulting from changes in the vibrational frequencies as well as global rotational and translational motions [2]. Lastly, the Karplus–Kushick QH approach [3] was employed to account for entropy differences arising from variations in the populations of conformational states of the glycosidic torsion angles in the bound versus free oligosaccharides (Table S5).

In agreement with the experimental data, and independent of the five GBSA desolvation models evaluated, **1** and **2** were always ranked the weakest binders and displayed essentially equivalent interaction energies, while **6** was correctly ranked as the highest affinity ligand. However, each of the models also ranked **5** amongst the best binders, in disagreement with experiment. In contrast, the PBSA desolvation model, in agreement with the experiment, ranked **5** along with **1** and **2** amongst the weakest binders and correctly ranked **6** as the strongest binder. Overall none of the GBSA models could rank all the ligands accurately; on the other hand, the PBSA model could correctly rank every ligand (Figure S6A and Table S6). As expected [4], incorporation of the entropic penalties always reduced the magnitude of the interaction energies. Surprisingly, the combination GBSA energies with QH entropy values did not lead to an improvement in the ranking of the relative affinities of the ligands (Figure S6B

and Table S7). In contrast, incorporation of NM entropies significantly improved the correlation between experimental and theoretical GBSA binding energies, and correctly ranked **3** and **6** amongst the best binders (Figure S6C and Table S9). Remarkably, the addition of either QH or NM entropies to the PBSA binding energies significantly degraded their correlation with the experimental affinities. While the QH entropies had not converged within 100 ns, it was possible to estimate their values at infinite sampling time by extrapolation [5]. In contrast, the extreme computational expense of NM calculations reduced the number of snapshots that could practically be analyzed to 100 from each simulation. Thus, the NM values are less precise and probably less accurate than the QH values. Neither QH nor NM entropy estimates account for the decrease in the number of conformational states in the ligand that can be significant for oligosaccharides. Ligand conformational entropy penalties were therefore computed using the Karplus-Kushick method based on the differences in glycosidic torsional states observed during MD simulations of the free and bound ligands [6]. The addition of relatively modest (< 1.5 kcal/mol) ligand conformational entropic penalties further improved these correlations but did not improve the poor correlations of the MM-PBSA data (Figure S6D and Tables S8 and S10).

Table S1. IC_{50} values for oligosaccharides.

Ligand	IC_{50}[a]
1	0.66 (0.04)
2	0.44 (0.01)
3	0.17 (0.01)
5	0.49 (0.05)
6	0.07 (0.01)
7	No Binding

[a]In mM.

Table S2. Average dihedral angles for the glycosidic linkages from the PDB structures and MD simulations of the six complexes (**1** to **6**) [a.]

		Galβ1-4Glcβ		Fucα1-2Galβ	
	Ligand	φ	ψ	φ	ψ
MD	**1**	-80.5 (22)	-134.0 (19)		
	2	-78.1 (21)	-133.2 (19)		
	3	-78.4 (15)	-131.1 (12)		
	4	-75.1 (7)	-134.0 (10)		
	5	-75.6 (7)	-135.9 (20)	-62.4 (7)	-86.5 (8)
	6	-74.9 (6)	-133.5 (8)	-63.4 (7)	-87.4 (8)
Exptl.	Average[b]	-73.7 (23)	-113.4 (41)	-78.5 (21)	-104.9 (17)

[a]Data in parentheses is the standard deviation.
[b]Average of dihedral angles for the glycosidic linkages from all the structures in the PDB database containing that linkage obtained using glytorsion.

Table S3. Experimental and theoretical hydrogen bonds and pairwise hydrogen bond interaction energies.

Ligand	Protein residue	Ligand residue	Distance[a]		Occupancy (%)	Interaction energy[b]		
			Exptl.	MD		GB^{HCT} (igb = 1)	GB_1^{OBC} (igb = 2)	GB_2^{OBC} (igb = 5)
1	D89-Oδ2	Gal-O4	2.6	2.6	100	-1.0	-3.7	-3.6
2		Gal-O3	2.6	2.8	98	**0.1**[c]	-3.2	-3.1
3		Gal-O3	2.6	2.6	99	**0.1**	-3.9	-3.0
5		Gal-O4	2.6	2.6	99	**0.2**	-3.5	-3.4
1	D89-Oδ1	Gal-O3	2.7	2.7	98	**0.0**	-3.3	-3.3
2		Gal-O4	2.7	2.6	100	-1.1	-3.8	-3.7
3		Gal-O4	2.6	2.7	100	-1.2	-3.1	-3.8
5		Gal-O3	2.7	2.7	100	-0.8	-2.8	-2.6
1	N133-Nδ2	Gal-O3	2.9	2.9	81	-2.5	-2.6	-3.0
2		Gal-O3	4.0	3.0	90	-2.7	-2.8	-3.2
3		Gal-O3	3.1	3.0	94	-2.6	-2.8	-3.1
5		Gal-O3	2.9	3.0	96	-3.0	-3.4	-3.8
1	A218-N	Gal-O4	3.1	3.0	99	-2.0	-2.0	-2.2
2		Gal-O4	3.2	3.1	99	-1.9	-2.0	-2.2
3		Gal-O4	3.1	3.2 (0.2)	97	-1.8	-1.8	-2.0
5		Gal-O4	3.1	3.0	99	-2.1	-2.1	-2.4
1	G107-N	Gal-O3	3.0	3.0	28	-1.6	-1.5	-1.7
2		Gal-O3	3.0	3.0	29	-1.7	-1.5	-1.8
3		Gal-O3	2.9	3.0	39	-1.7	-1.6	-1.7
5		Gal-O3	3.0	3.0	41	-1.8	-1.7	-1.9
1	D219-Nε2	Glc-O3	3.1	4.0 (1.1)	21	-1.3	-1.3	-1.4
2		ManO3	3.0	3.9 (0.9)	33	-1.5	-1.5	-1.6
3		GlcNAc-O3	2.9	3.4 (0.8)	71	-2.7	-2.7	-3.0
5		Glc-O3	3.1	4.0 (1.0)	34	-1.5	-1.5	-1.6
5	N133-Nδ2	Fuc-O2	2.7	3.1 (0.3)	64	-3.6	-3.9	-4.6
5	Y108-OH	Fuc-O4	3.0	4.4 (0.8)	14	-0.7	-0.8	-0.9

[a]In Å, with standard deviations greater than 0.1 shown in parentheses.
[b]In kcal/mol, standard error of mean is less than 0.01 in all cases.
[c]Numbers in bold represent residues with structurally inconsistent values.

Table S4. Theoretical hydrogen bonds and pairwise hydrogen bond interaction energies in modeled complexes.

Ligand	Protein residue	Ligand residue	Distance[a] MD	Occupancy (%)	Interaction energy[b] GB^{HCT} (igb = 1)	GB_1^{OBC} (igb = 2)	GB_2^{OBC} (igb = 5)
4	D89-Oδ2	GalNAc-O3	2.6	99	**0.0**[c]	-3.0	-2.7
6		Gal-O4	2.6	100	-0.9	-3.5	-3.4
4	D89-Oδ1	GalNAc-O4	2.7	100	-0.8	-3.4	-3.3
6		Gal-O3	2.7	100	**0.2**	-2.7	-2.4
4	N133-Nδ2	GalNAc-O3	3.0	97	-2.7	-2.8	-2.4
6		Gal-O3	3.0	97	-3.0	-3.4	-3.9
4	A218-N	GalNAc-O4	3.0	100	-2.2	-2.2	-1.9
6		Gal-O4	3.1	100	-2.0	-2.0	-2.3
4	G107-N	GalNAc-O3	3.1 (0.2)	13	-1.4	-1.3	-1.2
6		Gal-O3	3.0	43	-1.8	-1.7	-2.0
4	D219-Nε2	GlcNAc-O3	3.9 (1.4)	60	-2.3	-2.3	-2.1
6		GlcNAc-O3	3.2 (0.6)	84	-3.1	-3.0	-3.3
6	N133-Nδ2	Fuc-O2	3.0 (0.2)	77	-4.2	-4.5	-5.3
6	Y108-OH	Fuc-O4	4.0 (0.8)	28	-0.9	-1.2	-1.3

[a]In Å, with standard deviations greater than 0.1 shown in parentheses.
[b]In kcal/mol, standard error of mean is less than 0.01 in all cases.
[c]Numbers in bold represent residues with structurally inconsistent values.

Table S5. Entropy contributions (-TΔS) in kcal/mol at 300 K[a].

Ligand	VRT Quasiharmonic[b] (QH)	VRT Normal Mode[c] (NM)	Ligand Conformational Karplus–Kushick[d]
1	14.38 (0.01)	19.0 (0.87)	0.34
2	14.60 (0.02)	20.5 (1.00)	0.03
3	13.98 (0.02)	19.4 (0.95)	0.89
4	13.88 (0.01)	22.3 (0.98)	0.50
5	16.64 (0.01)	26.0 (1.04)	1.08
6	16.02 (0.01)	26.4 (0.96)	1.32

[a]Data in parentheses is the standard error of mean.
[b]Employing100,000 frames interpolated to infinite sampling.
[c]Employing 100 frames.
[d]Employing 100,000 frames.

Table S6. Experimental (BLI) and theoretical (not including entropy corrections) binding free energies[a].

Ligand	Exptl.	GB^{HCT} (igb = 1)	GB_1^{OBC} (igb = 2)	GB_2^{OBC} (igb = 5)	GB*n1* (igb = 7)	GB*n2* (igb = 8)	PBSA
Dielectric Constant (ϵ) = 1							
1	-4.8	-27.2	-30.2	-33.1	-35.7	-26.4	-13.1
2	-5.1	-28.3	-30.7	-33.6	-36.7	-26.5	-13.6
3	-5.7	-30.7	-32.7	-35.5	-37.3	-29.1	-19.0
4		-31.5	-32.2	-34.5	-36.7	-28.1	-14.5
5	-5.0	-36.5	-37.7	-41.3	-43.2	-33.7	-11.5
6	-6.2	-42.3	-41.6	-45.0	-44.4	-38.4	-21.4
	r[b]	0.74	0.7	0.68	0.57	0.74	0.95
ϵ = 4							
1	-4.8	-25.0	-25.7	-26.4	-27.0	-24.8	-13.4
2	-5.1	-25.9	-26.6	-27.3	-28.1	-25.7	-14.0
3	-5.7	-28.0	-28.5	-29.1	-29.6	-27.6	-19.3
4		-30.7	-30.9	-31.4	-31.9	-29.9	-15.3
5	-5.0	-35.8	-36.1	-36.9	-37.4	-35.1	-12.3
6	-6.2	-38.9	-38.7	-39.5	-39.4	-37.9	-21.9
	r	0.63	0.62	0.61	0.59	0.62	0.96

[a]In kcal/mol. All experimental errors less than 0.1 kcal/mol and all theoretical standard error of mean values less than 0.1 kcal/mol.
[b]Pearson correlation coefficient.

Table S7. Experimental (BLI) and theoretical (employing QH entropies) binding free energies[a]

Ligand	Exptl.	GB^{HCT} (igb = 1)	$GB_1{}^{OBC}$ (igb = 2)	$GB_2{}^{OBC}$ (igb = 5)	GBn_1 (igb = 7)	GBn_2 (igb = 8)	PBSA
1	-4.8	-12.8	-15.8	-18.7	-21.4	-12.0	1.3
2	-5.1	-13.7	-16.1	-19.0	-22.1	-11.9	1.0
3	-5.7	-16.8	-18.7	-21.5	-23.4	-15.1	-5.0
4		-17.6	-18.4	-20.6	-22.8	-14.2	-0.6
5	-5.0	-19.8	-21.1	-24.6	-26.5	-17.0	5.1
6	-6.2	-26.3	-25.5	-28.9	-28.3	-22.4	-5.4
	r[b]	0.82	0.80	0.77	0.69	0.83	0.71

[a]In kcal/mol. All experimental errors less than 0.1 kcal/mol and all theoretical standard error of mean values less than 0.1 kcal/mol.
[b]Pearson correlation coefficient

Table S8. Experimental (BLI) and theoretical (employing QH and conformational entropies) binding free energies[a]

Ligand	Exptl.	GB^{HCT} (igb = 1)	$GB_1{}^{OBC}$ (igb = 2)	$GB_2{}^{OBC}$ (igb = 5)	GBn_1 (igb = 7)	GBn_2 (igb = 8)	PBSA
1	-4.8	-12.5	-15.5	-18.4	-21.0	-11.7	1.6
2	-5.1	-13.6	-16.0	-19.0	-22.0	-11.9	1.0
3	-5.7	-15.9	-17.8	-20.6	-22.5	-14.2	-4.1
4		-17.1	-17.9	-20.1	-22.3	-13.7	-0.1
5	-5.0	-18.7	-20.0	-23.6	-25.4	-16.0	6.2
6	-6.2	-24.9	-24.2	-27.6	-27.0	-21.1	-1.2
	r[b]	0.82	0.81	0.77	0.67	0.84	0.63

[a]In kcal/mol. All experimental errors less than 0.1 kcal/mol and all theoretical standard error of mean values less than 0.1 kcal/mol.
[b]Pearson correlation coefficient

Table S9. Experimental (BLI) and theoretical (employing NM entropies) binding free energies[a]

Ligand	Exptl.	GB^{HCT} (igb = 1)	$GB_1{}^{OBC}$ (igb = 2)	$GB_2{}^{OBC}$ (igb = 5)	GBn_1 (igb = 7)	GBn_2 (igb = 8)	PBSA
1	-4.8	-8.2(0.9)	-11.2(0.9)	-14.1(0.9)	-16.7(0.9)	-7.4(0.9)	5.9(0.9)
2	-5.1	-7.8(1.0)	-10.1(1.0)	-13.1(1.0)	-16.2(1.0)	-6.1(1.0)	6.9(1.0)
3	-5.7	-11.3(1.0)	-13.3(1.0)	-16.1(1.0)	-17.9(1.0)	-9.7(1.0)	0.4(1.0)
4		-11.0(1.0)	-11.8(1.0)	-14.0(1.0)	-16.2(1.0)	-7.6(1.0)	7.8(1.0)
5	-5.0	-10.4(1.0)	-11.7(1.0)	-15.3(1.0)	-17.1(1.0)	-7.7(1.0)	14.5(1.0)
6	-6.2	-15.8(1.2)	-15.1(1.2)	-18.5(1.2)	-17.9(1.2)	-12.0(1.2)	4.9(1.2)
	r[b]	0.93	0.93	0.91	0.83	0.93	0.48

[a]In cal/mol. All experimental errors less than 0.1 kcal/mol. Theoretical standard error of mean shown in parentheses.
[b]Pearson correlation coefficient.

Table S10. Experimental (BLI) and theoretical (employing NM and conformational entropies) binding free energies[a]

Ligand	Exptl.	GB^{HCT} (1gb = 1)	$GB_1{}^{OBC}$ (1gb = 2)	$GB_2{}^{OBC}$ (igb = 5)	GBn_1 (igb = 7)	GBn_2 (igb = 8)	PBSA
1	-4.8	-7.8(0.9)	-10.8(0.9)	-13.7(0.9)	-16.4(0.9)	-7.1(0.9)	6.3(0.9)
2	-5.1	-7.7(1.0)	-10.2(1.0)	-13.1(1.0)	-16.2(1.0)	-6.0(1.0)	6.9(1.0)
3	-5.7	-10.4(1.0)	-12.4(1.0)	-15.2(1.0)	-17.1(1.0)	-8.8(1.0)	1.3(1.0)
4		-10.5(1.0)	-11.3(1.0)	-13.5(1.0)	-15.7(1.0)	-7.1(1.0)	8.3(1.0)
5	-5.0	-9.4(1.0)	-10.6(1.0)	-14.2(1.0)	-16.1(1.0)	-6.6(1.0)	15.6(1.0)
6	-6.2	-14.6(1.2)	-13.9(1.2)	-17.3(1.2)	-16.7(1.2)	-10.8(1.2)	6.3(1.2)
	r[b]	0.95	0.95	0.94	0.68	0.94	0.40

[a]In kcal/mol. All experimental errors less than 0.1 kcal/mol. Theoretical standard error of mean shown in parentheses.
[b]Pearson correlation coefficient.

Table S11. The impact of polar desolvation free energy on per-residue MM-PB/GBSA energies with dielectric constant of 4[a].

	GB^{HCT} (igb = 1)	GB_1 ^{OBC} (igb = 2)	GB_2 ^{OBC} (igb = 5)	GBn_1 (igb = 7)	GBn_2 (igb = 8)	PBSA
Residues forming hydrogen bonds with the ligand						
A218	-2.50	-2.35	-2.30	-2.03	-2.43	-2.43
D89	**0.25**[b]	-0.64	-1.01	-1.24	**1.57**	**0.20**
G107	-0.51	-0.38	-0.37	-0.26	-0.51	-0.51
N133	-0.75	-0.54	-0.53	-0.33	-0.45	-0.70
Q219	-2.18	-2.01	-2.02	-1.85	-2.02	-2.12
Residues involved in other interactions with the ligand						
A88	-0.78	-0.72	-0.70	-0.68	-0.84	-0.79
A222	-0.49	-0.50	-0.50	-0.55	-0.49	-0.49
F131	-2.14	-2.21	-2.25	-2.25	-2.18	-2.14
G217	-0.56	-0.32	-0.28	-0.08	-0.36	-0.53
P134	-0.16	-0.17	-0.18	-0.16	-0.17	-0.16
W135	-0.11	-0.14	-0.16	-0.21	-0.13	-0.12
Y106	-1.97	-1.77	-1.77	-1.79	-1.90	-1.95
Y108	-0.13	-0.15	-0.16	-0.17	-0.16	-0.13

[a]In kcal/mol, standard error of mean is less than 0.01 in all cases.
[b]Numbers in bold represent residues with structurally inconsistent values.

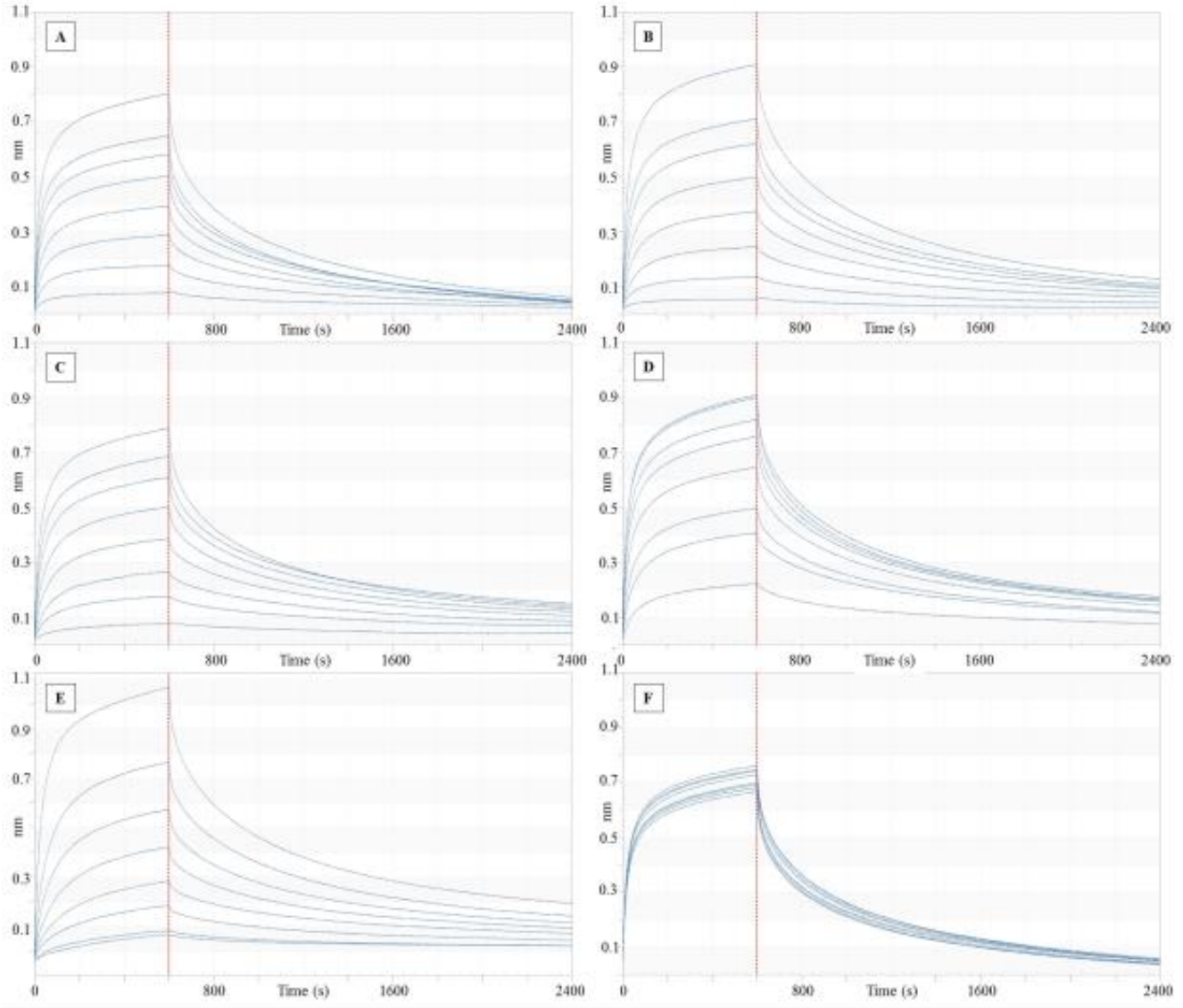

Figure S1. Inhibition of binding of ECL to **3** (immobilized on the BLI biosensor) for six ligands. A. Effect of **1** (concentration 0μM, 62.5μM, 0.125mM, 0.25mM, 0.5mM, 1mM, 2mM, 4mM) on binding. B. Effect of **2** (0μM, 62.5μM, 0.125mM, 0.25mM, 0.5mM, 1mM, 2mM, 4mM) on binding. C. Effect of **3** (0μM, 15.6μM, 31.3μM, 62.5μM, 0.125mM, 0.25mM, 0.5mM, 1mM) on binding. D. Effect of **5** (0μM, 15.6μM, 31.3μM, 62.5μM, 0.125mM, 0.25mM, 0.5mM, 1mM, 2mM) on binding. E. Effect of **6** (0μM, 15.6μM, 31.3μM, 62.5μM, 0.125mM, 0.25mM, 0.5mM, 1mM, 2mM) on binding. F. Effect of **7** (0μM, 0.125mM, 0.25mM, 0.5mM, 1mM, 2mM, 4mM, 8mM) on binding, indicating that **7** did not exhibit any inhibition or dose-dependence.

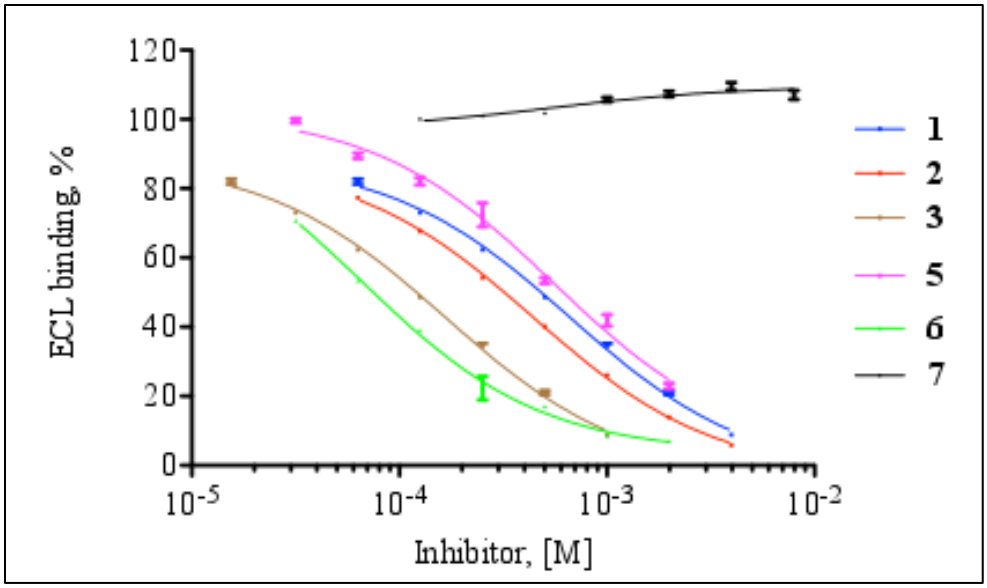

Figure S2. Inhibition curves used to compute IC_{50} values for **1**-**7** inhibiting the binding of ECL to immobilized **3**.

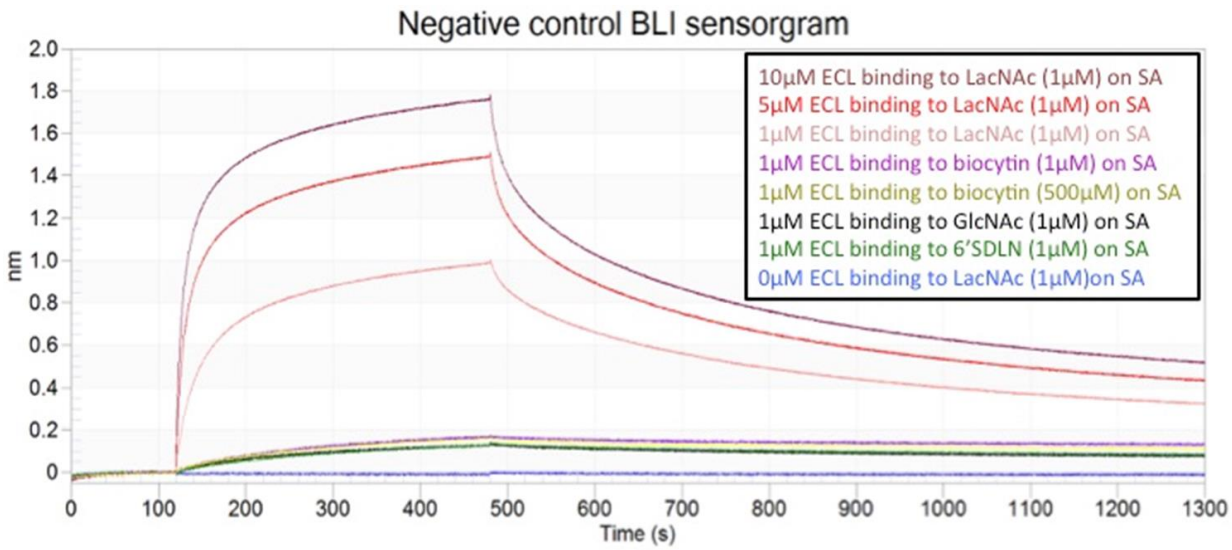

Figure S3. BLI control sensorgrams for ECL binding to positive (LacNAc) and negative controls (GlcNAc, 6'SDLN, and biocytin) surfaces.

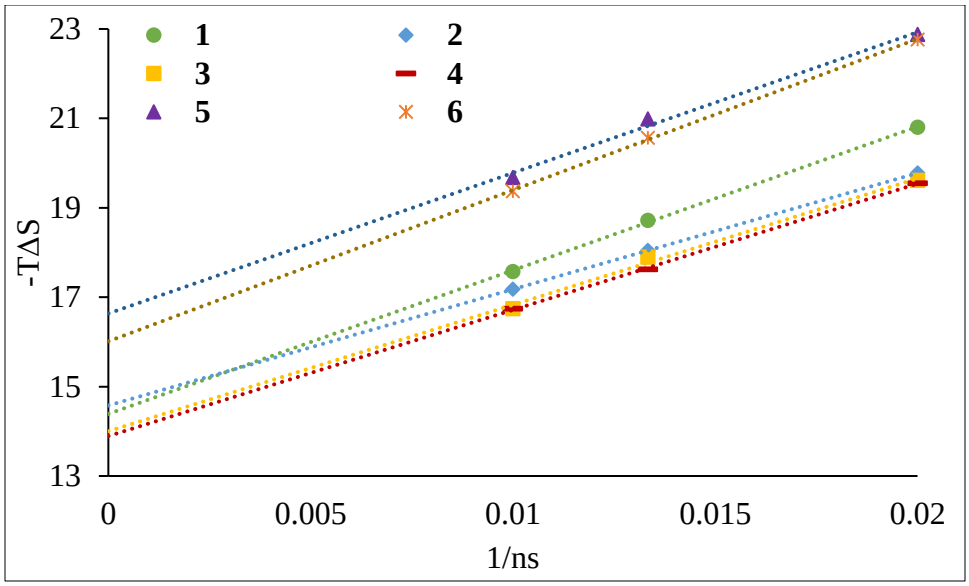

Figure S4. Extrapolation of quasi-harmonic entropy to infinite time for all the ligands. The coefficient of determination (R^2) in all the cases is greater than 0.99.

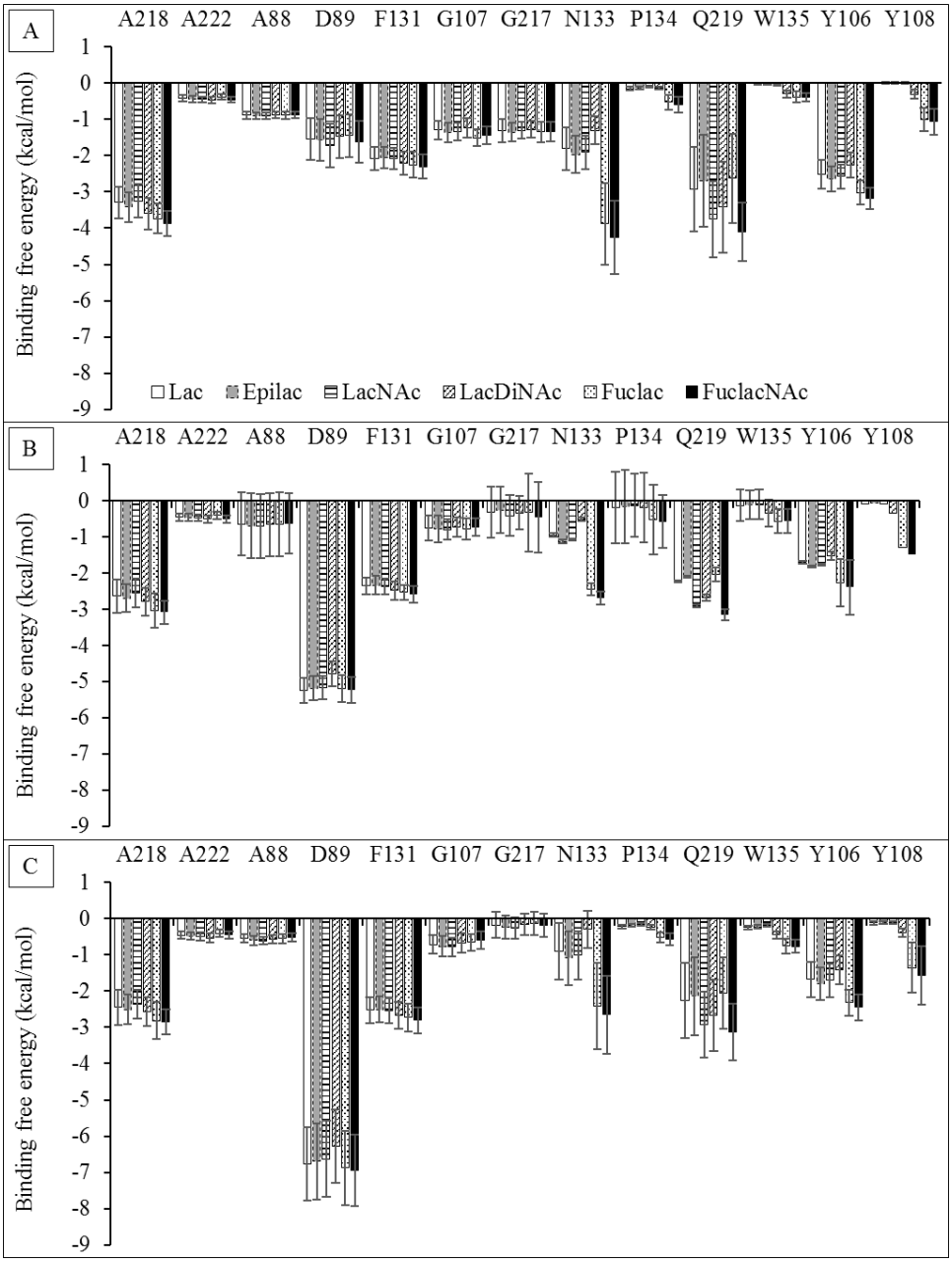

Figure S5. The binding free energy contributions of amino acids making significant interactions with the ligand. The calculations were performed using three different desolvation models. **A**. GB^{HCT}, **B**. $GB_1{}^{OBC}$ and **C**. $GB_2{}^{OBC}$.

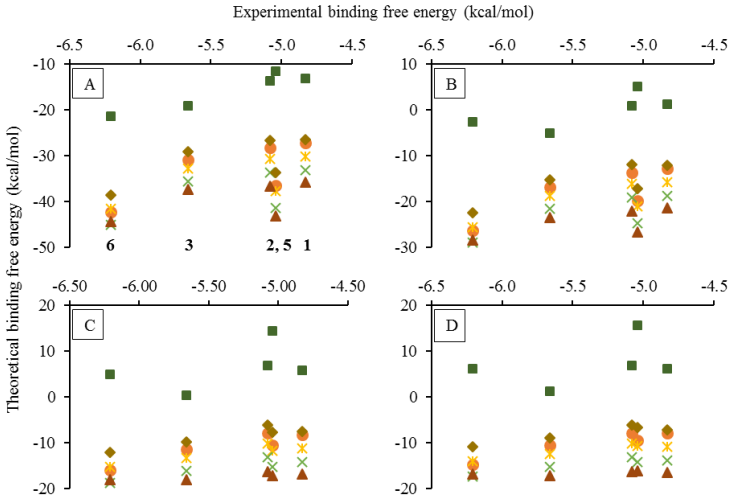

Figure S6. Comparison of theoretical (PBSA (green square), GB^{HCT} (orange circle), GB_1^{OBC} (yellow star), GB_2^{OBC} (green cross), GB*n1* (red triangle) and GB*n2* (brown rhombus)) and experimental binding free energies (Table 1) for five ligands (**1**, **2**, **3**, **5** and **6**). **A**. Binding free energies from MM-GB/PBSA calculation. **B.** Binding free energies from MM-GB/PBSA calculation employing quasi-harmonic entropies. **C.** Binding free energies from MM-GB/PBSA calculation employing normal mode entropies. **D**. Binding free energies from MM-GB/PBSA calculation employing normal mode and conformational entropies.

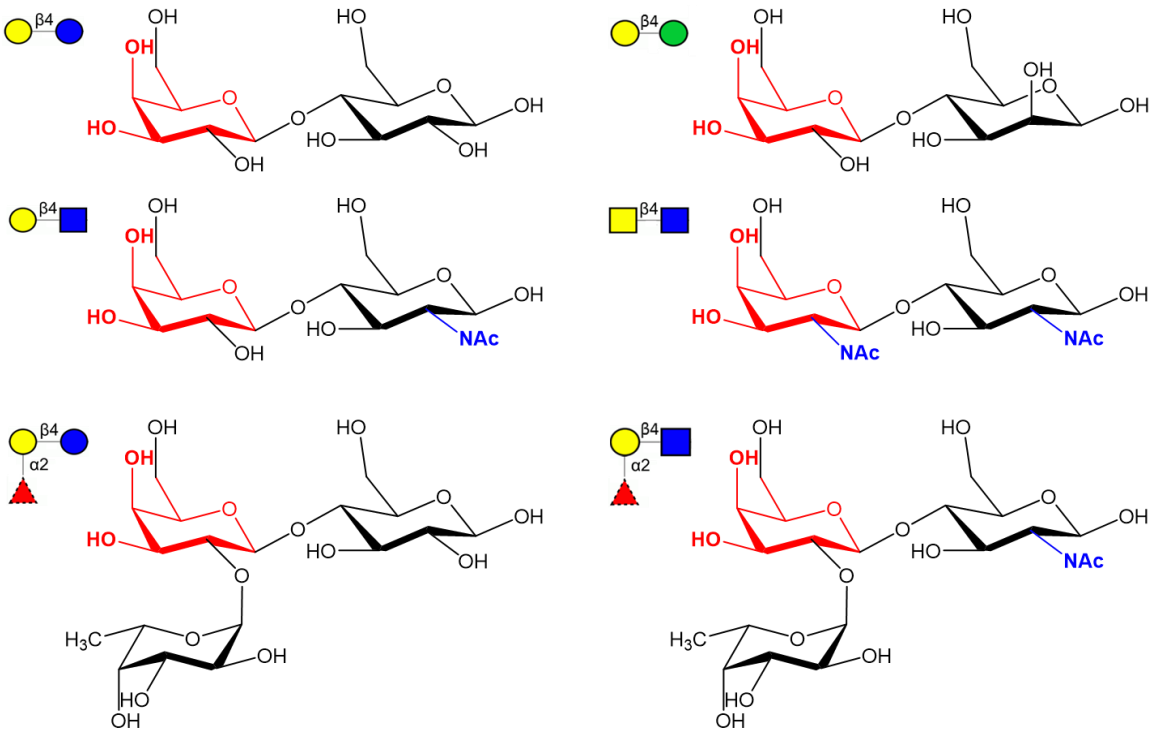

Figure S7. The pharmacophore (red) required for binding to ECL is defined by the spatial orientation of the O3 and O4 hydroxyl groups in the Gal/GalNAc residue along with the atoms forming the ring structure; indicated for six different ligands: Lac (top left), Epilac (top right), LacNAc (middle left), LacDiNAc (middle right), Fuclac (bottom left) and FuclacNAc (bottom right). The NAc groups (blue) enhance the affinity, but are not required for binding. The monosaccharides are represented in SNFG notation [7] as Gal: yellow circle, Glc: blue circle, Man: green circle, GlcNAc: blue square, GalNAc: yellow square and Fuc: red triangle.

References

1. Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D., Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an α-helix. *Macromolecules* **1984**, 17, 1370-1374.
2. Case, D. A., Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* **1994**, 4, 285-290.
3. Karplus, M.; Kushick, J. N., Method for estimating the configurational entropy of macromolecules. *Macromolecules* **1981**, 14, 325-332.
4. Gohlke, H.; Case, D. A., Converging free energy estimates: MM-PB (GB) SA studies on the protein–protein complex Ras–Raf. *J. Comput. Chem.* **2004**, 25, 238-250.
5. Schlitter, J., Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **1993**, 215, 617-621.
6. Bryce, R.; Hillier, I.; Naismith, J., Carbohydrate-protein recognition: molecular dynamics simulations and free energy analysis of oligosaccharide binding to concanavalin A. *Biophys. J.* **2001**, 81, 1373-1388.
7. Varki, A.; Cummings, R. D.; Aebi, M.; Packer, N. H.; Seeberger, P. H.; Esko, J. D.; Stanley, P.; Hart, G.; Darvill, A.; Kinoshita, T., Symbol nomenclature for graphical representations of glycans. *Glycobiology* **2015**, 25, 1323-1324.