Supporting Information for

Inhibition of the Mammalian Glycoprotein YKL-40: Identification of the Physiological Ligand

Abhishek A. Kognole[1] and Christina M. Payne[1*]

[1] Department of Chemical and Materials Engineering, University of Kentucky, Lexington, KY 40506, USA

*To whom correspondence should be addressed: Christina M. Payne, Department of Chemical and Materials Engineering, University of Kentucky, 177 F. Paul Anderson Tower, Lexington, KY 40506, USA. Tel.: +1-859-257-2902; Fax: +1-859-323-1929; E-mail: christy.payne@uky.edu (CMP)

Detailed Methods

Molecular Dynamics (MD) Simulation

Protonation states of all the titratable residues were determined according to the corresponding pKa values calculated by the H++ web server (1). The protein, structural waters, and ligands were constructed in a vacuum using CHARMM (2). The system was minimized for 1000 steps in vacuum using the Steepest Descent (SD) algorithm followed by another 1000 steps of minimization with the adopted basis Newton-Raphson (ABNR) algorithm. This procedure reduces the number of bad contacts prior to solvation of the solute. The polysaccharide systems were solvated in 100 Å × 100 Å × 100 Å cubic boxes, and the collagen-like peptide systems were solvated in 120 Å × 120 Å × 120 Å cubic boxes. Sodium or chloride ions were added to the solution to ensure overall charge neutrality. For neutral ligands, six chloride ions were required to neutralize the charge of YKL-40 titratable residues. The charged ligands, hyaluronan (-3), heparan sulfate (-12), and chondroitin sulfate (-9), required 3 chloride ions, 6 sodium ions, and 3 sodium ions for charge neutrality, respectively. After solvation, the systems were minimized again in the following sequence: 1000 steps of SD with the protein and ligand restrained, 1000 steps of SD with only the protein restrained, and 2000 steps of SD and 2000 steps of ABNR with no harmonic restraints. Extensive minimization, up to 10000 steps of SD, was carried out for systems bound to highly sulfated polysaccharides and collagen.

The solvated and minimized systems were then equilibrated prior to production MD simulations. The systems were heated from 100 K to 300 K in 50-K increments over 20 ps in the canonical ensemble. The system density was then equilibrated in the *NPT* ensemble at 300 K and 1 atm (101325 Pa) for 100 ps. The Nosé-Hoover thermostat and barostat were used to control temperature and pressure in CHARMM (3,4).

Production MD simulations of 250 ns were performed in the canonical ensemble at 300 K using NAMD (5). Temperature was controlled using Langevin thermostat (6). The SHAKE algorithm was used to fix the bond distances to all hydrogen atoms (7). Non-bonded interactions were truncated with a cutoff distance of 10 Å, a switching distance of 9 Å, and a non-bonded pair list distance of 12 Å. Long range electrostatics were described using the Particle Mesh Ewald method with a 6^{th} order b-spline, a Gaussian distribution width of 0.320 Å, and a 1 Å grid spacing (8). The velocity Verlet multiple time-stepping integration scheme was used to evaluate non-bonded interactions every 1 time step, electrostatics every 3 time steps, and 6 time steps between atom reassignments. All simulations used a 2-fs time step. The CHARMM36 force field with the CMAP correction (2,9,10) was used to describe YKL-40 and the collagen ligands. The parameters for hydroxyproline were determined using ParamChem, which determines force field parameters based on analogy with CHARMM General Force Field (CGenFF)

program version 0.9.7 beta (11). The CMAP corrections for hydroxyproline were adopted simply based on the analogy between proline and hydroxyproline residues. The polysaccharides were described using the CHARMM36 carbohydrate force field (12-14). Water was modeled using the TIP3P force field (15,16). All simulations used explicit solvent.

A complete list of simulations and calculations performed to meet the objectives of this study is given in Table S1. As described in the manuscript, collagen docking calculations indicated two potential binding surfaces; for these cases, the description in Table S1 lists both site and ligand. The length of each MD simulation is also given, as not all simulation lengths were the same; several of the hypothesized ligands dissociated from the binding cleft, and the simulation was halted to conserve computational resources. The free energy calculations performed are also indicated. If a ligand did not remain in the binding cleft throughout the entirety of the MD simulation, a free energy calculation was not performed.

In addition to these protein-ligand or protein-protein complexes, oligo-saccharides and collagen models were solvated in water separately, without YKL-40. These ligand-only simulations were required as input to the free energy calculations. Several additional system configurations beyond those originally proposed were also developed, as described below, in order to study the effect of ligand position on conformational changes and to understand the statistical significance of observed interactions with the putative heparin-binding subsite.

Table S1. Simulations and calculations performed in the investigation of the binding of polysaccharides and collagen ligands to YKL-40.

Case No.	System	MD simulation	Free Energy Calculation
1	Apo YKL-40	250 ns	--
2	YKL-40 + chitohexaose	250 ns	FEP/λ-REMD
3	YKL-40 + cellohexaose	250 ns	FEP/λ-REMD
4	YKL-40 + hyaluronan	250 ns	FEP/λ-REMD
5 [a]	YKL-40 + heparin (fully sulfated)	50 ns	--
6	YKL-40 + heparan sulfate (unsulfated)	50 ns	--
7	YKL-40 + chondroitin sulfate	50 ns	--
8 & 9	YKL-40 + collagen (1CAG) at site A & B	250 ns	--
10 & 11	YKL-40 + collagen (native 1CAG) at site A & B	250 ns	Umbrella Sampling
12 & 13	YKL-40 + collagen (1BKV) at site A & B	250 ns	--
14 & 15	YKL-40 + collagen (1Q7D) at site A & B	250 ns	Umbrella Sampling

[a] Four YKL-40 + heparin systems were constructed: two with heparin initially in the primary polysaccharide binding cleft and two with heparin initially located in bulk solution (Figure 5).

Modeling of heparin in this study required development of new force-field parameters for GlcNAc (Figure S1) where the acetyl group was replaced by SO_3^{-1}. ParamChem was used to obtain an initial set of parameters (11,17). As the sulfamate anions were not explicitly supported, parameters obtained for –NHSO3 group by analogy required optimization. The Force Field Toolkit (ffTK) Plugin Version 1.0 in VMD, developed by Mayne et al. (18), was used to optimize the partial charges, bonds, angles, and dihedrals as described in the reference publication and provided examples. Parameters obtained using this approach are given in Table S2.

Table S2. CHARMM-additive parameters for GlcNS optimized using the ffTK v.1.0 plugin in VMD. The atom labels are as illustrated in Figure S1.

Bonds	**K_b**		**b_0**
C2 – N	271.158		1.464
N – S1	332.175		1.823
N – HN	440.214		1.029
S1 – O2	540.346		1.452
Angles	**K_{theta}**		**$Theta_0$**
C1/C3 – C2 –N	91.721		112.507
N – C2 –H2	114.884		111.824
C2 – N – S1	124.591		117.44
C2 – N – HN	79.624		107.895
S1 – N – HN	74.629		129.979
N – S1 – O2	152.857		109.282
O2 – S1 – O7	103.66		105.957
Dihedrals	**K_{chi}**	**n**	**Delta**
N – C2 – C1 – O5	0.2	3	0
N – C2 – C3 – O3	0.2	3	0
N – C2 – C1 – O1	0.2	3	0
C4 – C3 – C2 – N	0.2	3	0
N – C2 – C3 – H3	0.2	3	0
N – C2 – C1 – H1	0.2	3	0
C1/C3 – C2 – N – S1	1.12	3	180
H2 – C2 – N – HN	0.527	3	180
H2 – C2 – N – S1	2.994	3	0
C2 – N – S1 – O2	1.048	3	180
NH – N – S1 – O2	0.831	3	0
C1/C3 – C2 – N – HN	1.575	1	0
O4* – C1 – C2 – N	0.2	3	0

*this O4 is from the glycosidic linkage this residue will be involved in.

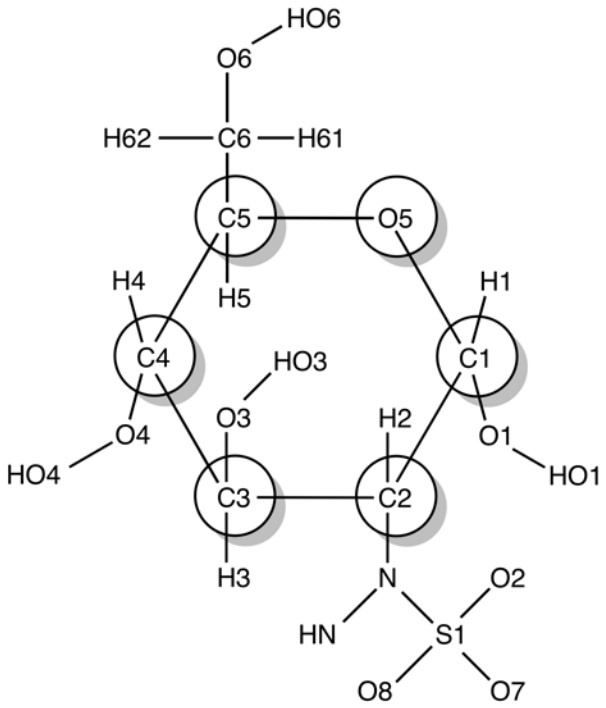

Figure S1. Atom labels of N-sulfo-α-D-glucosamine structure used for optimization of missing force-field parameters. The only missing parameters were the ones around N-S1 bond as documented in Table S2.

Free Energy Calculations

FEP/ λ-REMD. For the free energy calculation, structures of the YKL-40·chitohexaose, YKL-40·cellohexaose, and YKL-40·hyaluronan complexes were obtained from 25 ns snapshots of the MD simulations. A solvated polysaccharide ligand simulation was also generated to determine ΔG_2 (Figure S2). For each of these systems, 20 consecutive 0.1-ns free energy perturbation calculations were performed with NAMD 2.9, using a replica exchange frequency of 1 exchange per 100 steps. The final 10 calculations (1 ns) were averaged to determine the binding free energy. The simulations used a total of 128 free energy perturbation replica windows: 72 dispersive, 24 repulsive, and 32 electrostatic. The oligosaccharide ligands were restrained in the ligand-binding pose using a harmonic restraint on the distance between the center of mass of the protein and the center of mass of the ligand. The harmonic restraint force constant was 41.84 kJ/mol/Å^2. This restraint bias was removed from the free energy calculation according to the approach outlined by Deng and Roux (19). Multistate Bennett Acceptance Ratio (MBAR) was used to determine electrostatic, repulsive, and dispersive contributions to free energy (20). Standard deviation of the final 1 ns free energy values serves as the error estimate. All simulation parameters in the free energy calculations mimic those described in the *MD simulations* section. The progress towards the convergence of free energy calculations for cellohexaose, chitohexaose and hyaluronan systems are shown in Figure S3.

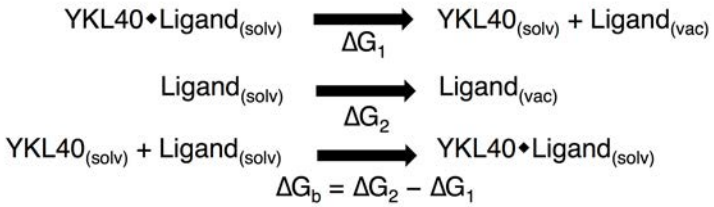

Figure S2. Thermodynamic cycle used to determine ΔG with FEP/λ-REMD method. 'solv' refers to the solvated state and 'vac' refers to the gas-phase state.

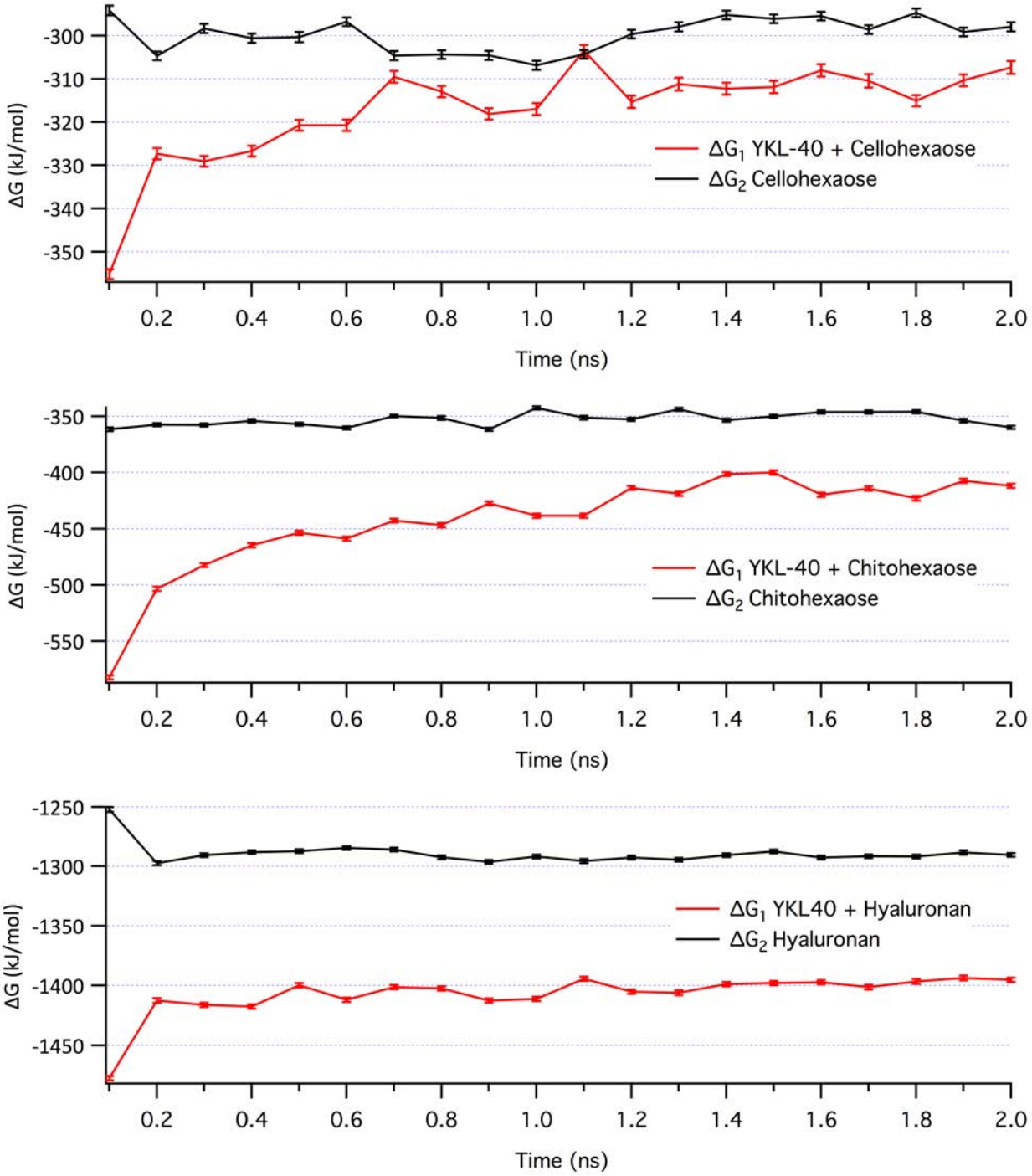

Figure S3. Convergence of ΔG over 20 consecutive 0.1-ns free energy perturbation calculations using the FEP/λ-REMD method.

Additional results and discussion

RMSD and RMSF of protein in polysaccharide binding dynamics

The RMSD of the protein (FigureS4a) is a measure of deviation over the course of the simulation from the initial configuration, which was the first frame of the simulation following *NPT* density equilibration. The relatively consistent RMSD of the protein backbones suggests the simulations reached a local equilibrium. The magnitude of the RMSD change over 250 ns is small given the significant chemical differences in the three ligands examined, which indicates the primary YKL-40 binding site is forgiving of small charged side chains such as the carboxylate of hyaluronan. The RMSF fluctuation of the protein backbone similarly describes fluctuation of a given protein residue from the average position over the course of the entire simulation. As with the RMSD calculation, the RMSF of the protein backbone suggests the binding of chitohexaose and cellohexaose does little to disturb the overall protein conformation (Figure S4b). In the case of hyaluronan binding, we observe increased fluctuation in residues 178-189, 225-235, and 300-325 over that of cellohexaose and chitohexaose bound YKL-40. Both loops 225-235 and 300-325 are located away from the primary carbohydrate-binding site; the increase in flexibility in these loops appears to be related to solvent exposed polar residues sampling bulk solution and is likely unrelated to hyaluronan binding. Segment 178-189, comprising part of a β-sheet and a small α-helix just beneath the +1 and +2 binding sites, becomes increasingly mobile as its interaction with hyaluronan is lost in the formation of the sharp V-shape. Despite localized increases in backbone flexibility, the overall protein structure largely remains in the same initial conformation, as evidenced by the similarity in RMSD (Figure S4a).

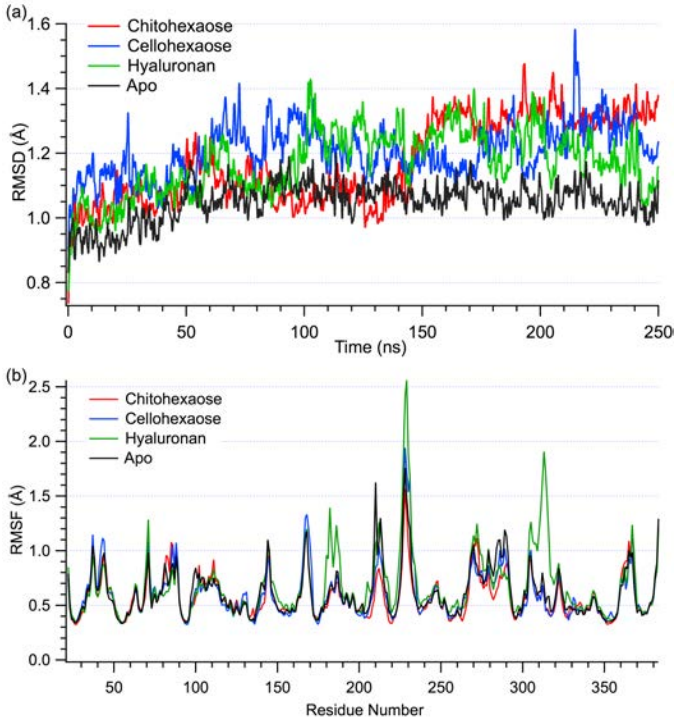

Figure S4. (a) Root-mean-square deviation over 250-ns MD simulations and (b) root-mean-square fluctuation of YKL-40 without a ligand (apo) and bound to chitohexaose, cellohexaose, and hyaluronan. Binding of chitohexaose, cellohexaose, and hyaluronan do not significantly alter the dynamics of YKL-40.

Conformational changes in the YKL-40 binding site

Crystal structures of YKL-40 bound with chito-oligosaccharides suggest that YKL-40 undergoes a conformational change upon chitin ligand binding (21), contrary to suggestions that lectin binding sites, in general, are "pre-formed" to accommodate their natural substrates and undergo little change upon sugar binding (22). Houston et al. reported that the residues forming a loop (residues 209 to 213) near the primary YKL-40 binding cleft occupy an unusual conformation in apo YKL-40 when compared to the ligand bound YKL-40 structure, where Trp 212 lines the +2 and +3 subsites (21). However, a second structural investigation published concurrently did not observe a similar conformation change in either of two crystal structures (1NWR and 1NWS), where no ligand occupied either the +2 or +3 subsites (23). Additionally, the positioning of Trp99 at the +1 site in both apo structures of human YKL-40 (1HJX and 1NWR) and the homologous MGP-40 (1LJY) differs from that of holo-YKL-40 and homologous mammalian lectin Ym1 (1E9L) (21,23-25), with the tryptophan blocking the binding cleft in the apo form. This conformational variation as a function of binding site occupancy has been proposed as a tryptophan-mediated gating mechanism for ligand binding in chitolectins (26).

Based on MD simulations we did not observe data suggesting binding cleft rearrangement is important in polysaccharide binding to YKL-40. To investigate possible loop rearrangement upon ligand unbinding, the apo YKL-40 simulation was prepared by undocking the bound chitin oligomer. One can reasonably expect that over the course of a 250-ns MD simulation, the 5-amino acid residue loop would, at a minimum, sample a variety of conformations indicating flexibility in this region. However, in examining the trajectory of this loop with respect to its initial position, we did not observe the peptide loop returning to the unusual conformation in a single frame (Figure S5). This suggests that the crystallographic apo conformation may have resulted from serendipitous crystal packing interactions and may not represent a typical conformational behavior. Additionally, the phenomenon of tryptophan mediated gating, according to which one would expect the Trp99 to return to the "pinched" conformation of the apo state, was not observed. Though, we note the likelihood of observing that the latter behavior, i.e., returning to a "pinched" conformation, in an unbiased MD simulation is low and may require overcoming an energy barrier through enhanced sampling approaches.

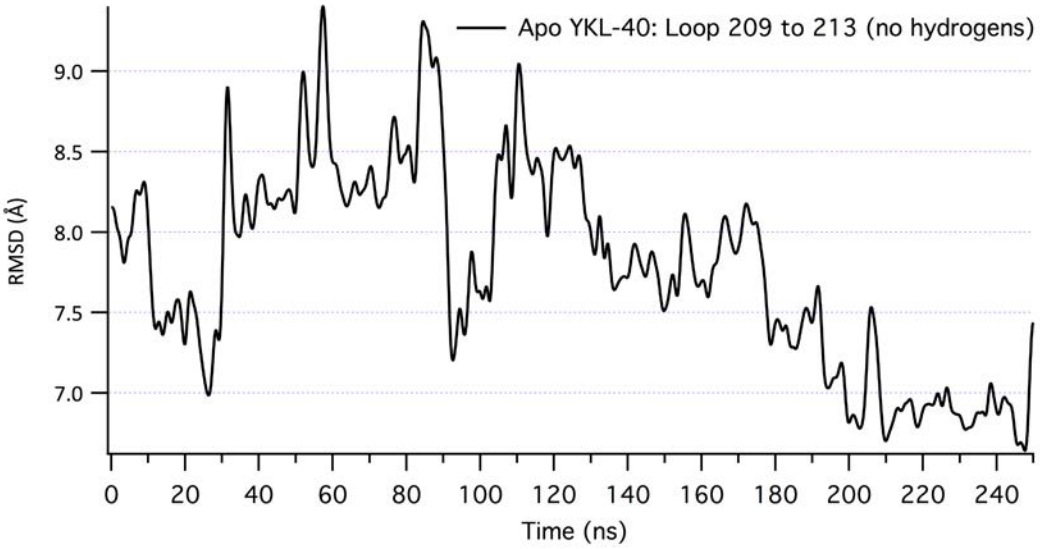

Figure S5. Root mean square deviation of loop of residues 209 to 213 from the unusual configuration in apo YKL-40 crystal structure during 250-ns MD simulation of apo YKL-40 prepared by removing the bound ligand from holo crystal structure.

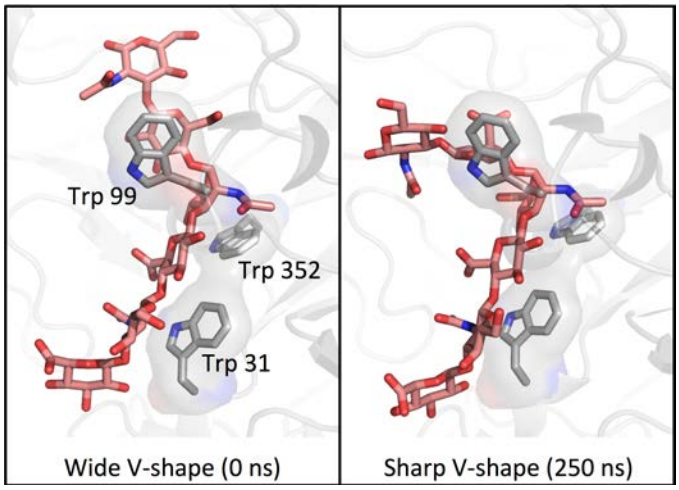

Figure S6. Hyaluronan in YKL-40 binding site at 0 ns (left) and at 250 ns (right) illustrating difference between V-shape conformations of hyaluronan.

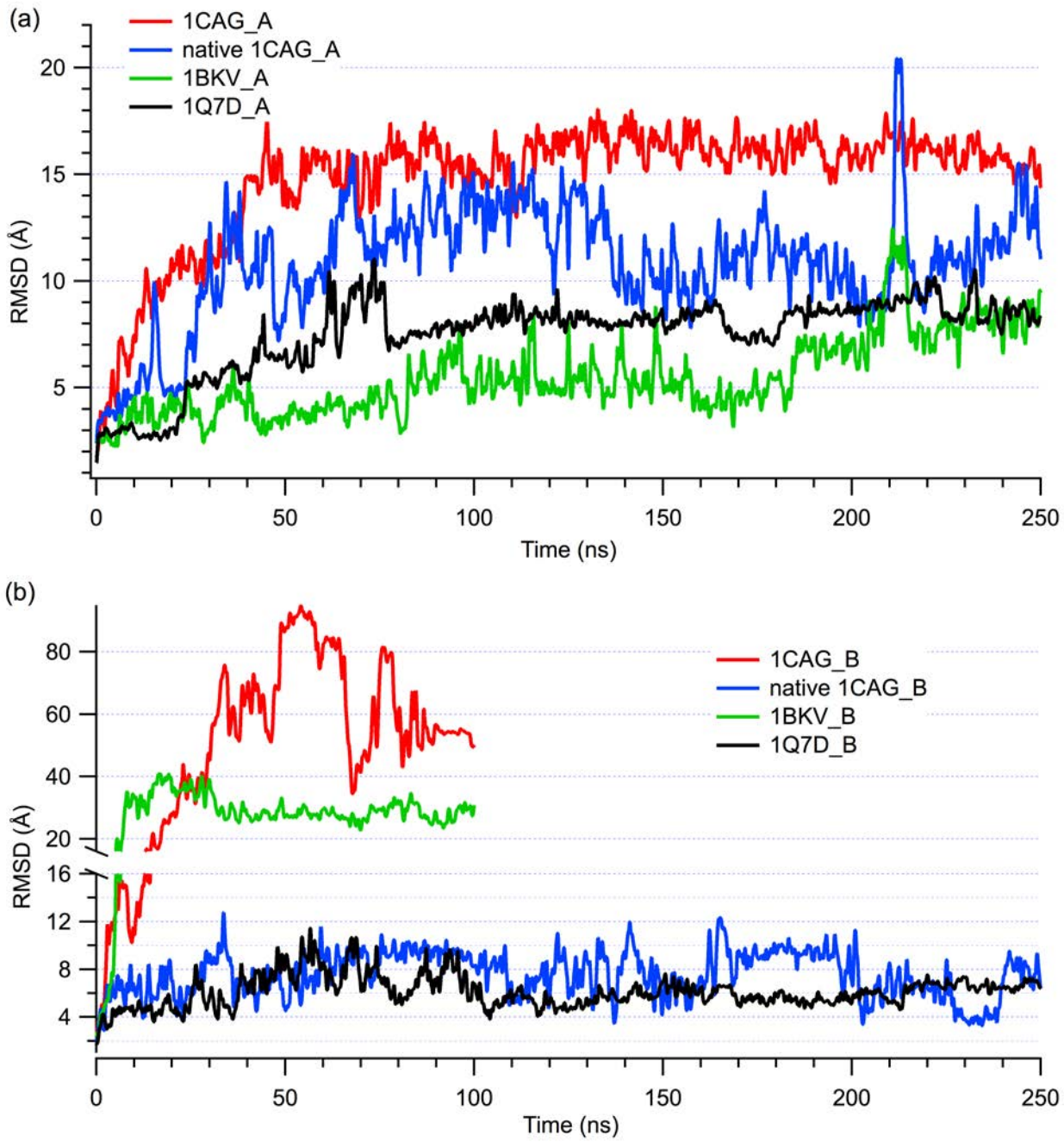

Figure S7. Root-mean-square deviation of collagen-like peptides over the course of 250-ns MD simulations at (a) collagen binding site A and (b) collagen binding site B. Each of the four collagen model peptides are shown.

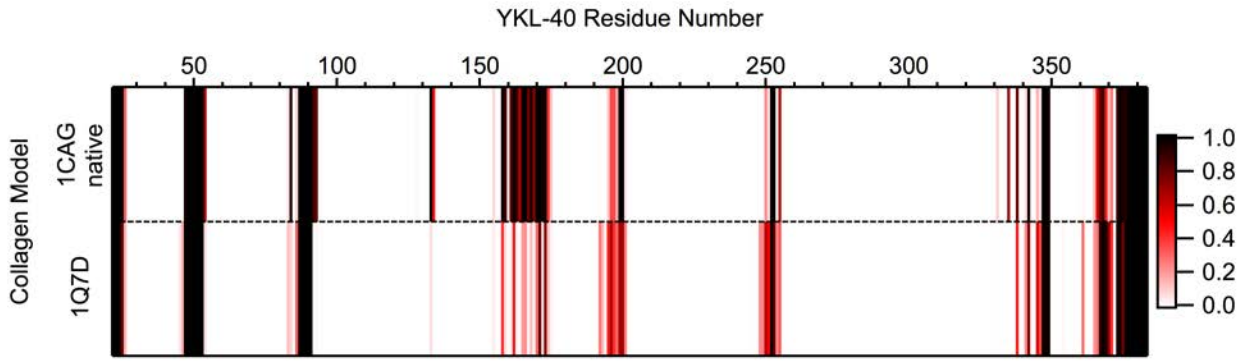

Figure S8. Native contact analysis of each collagen-like peptide model binding to YKL-40 at site B. The color scale represents the normalized frequency (i.e., fractional percentage of frames in which the contact was formed) of the respective YKL-40 residue as a native contact. A native contact was defined as anytime a collagen residue was within 12 Å of a YKL-40 residue, where distance was defined by the center of geometry of a given residue. Only frames from the last 100 ns simulation, following the period of equilibration, were considered in this analysis.

Table S3. Interaction energy between each YKL-40 residue and heparin oligosaccharide averaged over the trajectory after heparin non-specifically interacts with the putative surface-binding site. The table shows the data for residues with the most favorable total average interaction energies. All values are in kJ/mol.

Residue	VdW-Avg	Elec-Avg	Total Avg
ARG145	-9.12	-468.99	-478.11
LYS193	-9.89	-438.47	-448.35
LYS155	-2.18	-330.12	-332.30
ARG144	-8.85	-307.80	-316.65
LYS147	-1.81	-262.19	-263.99
HIS148	-13.41	-50.79	-64.20

Table S4. Hydrogen bonding pairs from polysaccharide-bound molecular dynamics simulations. A hydrogen bond was defined as a polar atom having a donor-acceptor distance of 3.4 Å and a 60° cutoff angle. Occupancy refers to the percent of the simulation during which the hydrogen bond was formed. Occupancies less than 10% have not been reported unless relevant in comparison.

Binding Site	Cellohexaose			Chitohexaose			Hyaluronan		
	Donor	**Acceptor**	**Occupancy**	**Donor**	**Acceptor**	**Occupancy**	**Donor**	**Acceptor**	**Occupancy**
-4	BGLC1-SC	GLU70-SC	56.28%	NAG1-SC	GLU36-SC	9.32%	LYS289-SC	GCU1-SC	13.40%
				LYS289-SC	NAG1-MC	8.48%	GCU1-SC	TRP31-MC	12.55%
-3	TRP69-SC	BGLC2-SC	53.32%	NAG2-SC	GLU290-SC	69.52%	ASN100-SC	NAG1-SC	35.71%
	BGLC2-SC	GLU70-SC	34.76%	ASN100-SC	NAG2-MC	67.68%	TRP69-SC	NAG1-SC	9.76%
	ASN100-SC	BGLC2-SC	21.80%						
-2	ASN100-MC	BGLC3-SC	87.40%	TRP352-SC	NAG3-MC	93.24%	TRP31-SC	GCU2-SC	41.46%
				ASN100-MC	NAG3-SC	66.00%	ASN100-MC	GCU2-SC	22.27%
				NAG3-SC	GLU290-SC	30.32%	TRP99-MC	GCU2-SC	16.03%
				NAG3-SC	ASN100-SC	13.44%	ASN100-SC	GCU2-SC	13.60%
-1	TRP99-MC	BGLC4-SC	76.20%	TYR206-SC	NAG4-MC	75.16%	TRP99-MC	NAG2-MC	86.76%
				TRP99-MC	NAG4-SC	39.56%			
				TYR206-SC	NAG4-SC	16.52%			
				NAG4-SC	ASP207-SC	15.16%			
+1	BGLC5-SC	TYR141-SC	32.32%	NAG5-MC	ASP207-SC	74.08%	GCU3-SC	ASP207-SC	96.19%
	BGLC5-SC	ASP207-SC	18.08%	NAG5-SC	TYR141-SC	17.00%	ARG263-SC	GCU3-SC	76.88%
	TYR141-SC	BGLC5-SC	13.52%	TYR141-SC	NAG5-SC	15.04%	TYR141-SC	GCU3-SC	62.75%
				ARG263-SC	NAG5-MC	14.28%			
+2	TYR141-SC	BGLC6-SC	52.04%	NAG6-MC	TYR141-SC	45.68%	TRP99-SC	NAG3-SC	48.14%
				TYR141-SC	NAG6-SC	18.88%			
				TRP99-SC	NAG6-SC	10.28%			

SC – Side chain; MC – Main chain; BGLC – β-D-glucose; NAG – N-acetyl-α-D-glucosamine; GCU – β-D-glucuronic acid.

Table S5. Interaction energies of YKL-40 residues with collagen peptides. The values are reported in terms of average interaction energy between major YKL-40 residues and collagen as a whole. van der Waals and electrostatic contributions are also provided separately. Residues with total average interaction energy greater than -4.18 kJ/mol have not been reported unless relevant to discussion. Energies are given in kJ/mol.

	Residue #	VdW-Avg	Elec-Avg	Total Avg
1Q7D - binding site A	ARG263	0.330	-23.566	-23.239
	THR184	-1.129	-16.381	-17.514
	LYS182	-0.723	-16.695	-17.418
	TRP212	-13.667	-1.957	-15.624
	ASP207	-0.732	-14.353	-15.084
	TYR141	-4.437	-9.677	-14.118
	GLU70	0.042	-13.370	-13.328
	GLU290	-3.383	-8.109	-11.492
	ARG145	0.138	-10.292	-10.154
	TYR34	-7.783	-0.815	-8.598
	ASN100	-6.821	-1.644	-8.464
	TRP99	-5.708	-2.308	-8.017
	PRO142	-0.381	-5.562	-5.943
	VAL183	-6.273	0.598	-5.675
	GLU36	-1.451	-3.919	-5.366

	Residue #	VdW-Avg	Elec-Avg	Total Avg
Native 1CAG - binding site A	ASP232	0.000	-27.814	-27.814
	TRP99	-13.592	-7.553	-21.144
	TRP212	-15.666	0.351	-15.314
	VAL183	-9.899	-1.618	-11.517
	PHE234	-10.225	-0.084	-10.309
	ASN100	-5.290	-2.534	-7.825
	GLU290	-0.694	-4.943	-5.637
	THR184	-2.835	-2.459	-5.294
	GLN104	-2.196	-2.275	-4.466
	TYR141	-3.801	-0.385	-4.186
	ASP207	-0.661	-3.220	-3.881

	Residue #	VdW-Avg	Elec-Avg	Total Avg
1CAG - binding site A	GLU70	-0.882	-22.512	-23.394
	TRP99	-13.081	-4.040	-17.121
	GLU290	-2.409	-9.656	-12.061
	ASN100	-8.243	-2.530	-10.773
	TRP69	-6.357	-3.191	-9.548
	TRP71	-8.975	-0.443	-9.418
	ALA211	-5.796	-2.802	-8.598
	TRP212	-6.620	-1.255	-7.871
	ASP207	-0.309	-5.591	-5.901
	TYR34	-5.474	0.744	-4.730
	TRP31	-3.358	-0.506	-3.864

	Residue #	VdW-Avg	Elec-Avg	Total Avg
1BKV - binding site A	ASP207	-4.057	-59.811	-63.868
	PHE208	-5.294	-14.210	-19.501
	ALA180	-1.957	-17.447	-19.404
	TYR141	-5.445	-12.680	-18.125
	TRP99	-13.031	-4.563	-17.594
	HIS209	-7.373	-7.419	-14.792
	TRP212	-8.247	-4.015	-12.257
	LYS182	-1.577	-9.890	-11.467
	SER179	-1.752	-9.656	-11.408
	GLU290	-4.985	-5.805	-10.790
	ARG213	-1.811	-8.811	-10.622
	TYR206	-1.033	-8.607	-9.640
	GLY210	-2.727	-6.888	-9.614
	ALA211	-0.815	-7.891	-8.707
	TYR34	-6.942	-1.246	-8.188
	GLU36	-1.853	-4.328	-6.181
	VAL183	-5.943	0.146	-5.796
	ASN100	-3.960	-1.455	-5.420
	TRP31	-4.583	-0.372	-4.951

	Residue #	VdW-Avg	Elec-Avg	Total Avg
1Q7D - binding site B	LYS23	-0.795	-86.229	-87.023
	TYR22	-4.788	-46.654	-51.443
	LYS91	0.046	-38.608	-38.562
	PHE49	-15.130	-1.878	-17.004
	ASP367	-0.259	-16.097	-16.360
	LYS377	-5.361	-9.142	-14.507
	THR52	-3.178	-10.120	-13.295
	ASP47	-0.949	-8.310	-9.263
	LYS253	-1.171	-6.821	-7.992
	ASN89	-6.917	-0.719	-7.632
	ASP378	-0.941	-5.717	-6.658
	ALA381	-5.228	0.322	-4.910

	Residue #	VdW-Avg	Elec-Avg	Total Avg
Native 1CAG - binding site B	ASN89	-11.965	-26.439	-38.399
	LYS377	-7.122	-26.133	-33.255
	ASP378	-5.119	-15.812	-20.927
	ALA381	-9.368	-3.538	-12.906
	GLN166	-6.411	-4.199	-10.606
	THR52	-8.163	-2.308	-10.472
	GLN171	-6.382	-3.015	-9.397
	PHE49	-7.365	-0.912	-8.280
	TYR22	-4.412	-2.639	-7.051
	LYS91	-3.726	-2.881	-6.612
	LEU50	-2.831	-2.865	-5.696
	HIS53	-2.798	-2.262	-5.060
	ASP199	-0.544	-3.680	-4.224

Table S6. Hydrogen bonding pairs between YKL-40 and collagen model peptides at binding site A, including percentage occupancy, over 250-ns MD simulations. A hydrogen bond was considered to be a polar atom having a donor-acceptor distance of 3.4 Å and a 60° cutoff angle. Occupancies above 100% mean that the same pair was involved in more than one type of hydrogen bond.

1Q7D - binding site A			Native 1CAG - binding site A		
Donor	Acceptor	Occupancy	Donor	Acceptor	Occupancy
ARG263-SC	GLU11-SC	164.84%	ARG213-SC	HYP8-SC	79.68%
ARG12-SC	ASP207-SC	126.36%	HYP8-SC	ASP232-SC	76.56%
ARG12-SC	THR184-SC	51.96%	GLN104-MC	HYP14-SC	19.60%
ARG12-SC	ALA291-MC	26.76%	SER103-MC	HYP14-SC	16.24%
HYP9-SC	GLU290-SC	25.56%	ARG233-SC	HYP2-SC	13.44%
HYP6-SC	GLU70-SC	18.80%	HYP17-SC	ASN100-SC	10.32%
TYR141-SC	GLU11-SC	17.28%	other pairs		123.68%
ASN100-SC	HYP9-SC	14.56%			
ARG12-SC	SER179-SC	13.44%			
HYP6-SC	TYR34-MC	13.24%			
ARG12-SC	ASP207-MC	12.04%			
other pairs		100.20%			
Total		585.04%	Total		339.52%

1CAG - binding site A			1BKV - binding site A		
Donor	Acceptor	Occupancy	Donor	Acceptor	Occupancy
HYP20-SC	GLU70-SC	75.72%	TRP99-SC	ALA17-MC	64.92%
ASN100-SC	HYP17-MC	35.84%	ARG14-MC	PHE218-MC	56.88%
GLY214-MC	HYP5-MC	29.24%	ARG11-SC	SER179-SC	54.08%
ARG213-SC	HYP5-SC	26.20%	LYS182-SC	THR8-SC	45.60%
HYP14-SC	ALA291-MC	17.80%	ARG11-SC	TYR141-SC	42.56%
GLY214-MC	GLY6-MC	11.00%	ARG11-SC	ALA180-MC	40.04%
other pairs		112.08%	ARG11-MC	THR184-SC	33.00%
			ARG11-SC	ASP207-MC	28.64%
			ARG11-SC	ASP207-SC	21.56%
			ARG213-SC	THR11-MC	20.24%
			ARG14-SC	GLY210-MC	15.92%
			ARG11-SC	TYR206-MC	13.72%
			TYR141-SC	GLY12-MC	13.04%
			TRP212-SC	GLY12-MC	12.76%
			ARG14-SC	ALA211-MC	11.04%
			other pairs		80.28%
Total		307.88%	Total		554.28%

1Q7D - binding site B			Native 1CAG - binding site B		
Donor	Acceptor	Occupancy	Donor	Acceptor	Occupancy
LYS23-SC	GLU11-SC	120.84%	ASN89-SC	GLY18-MC	94.68%
ASN87-SC	HYP6-SC	61.20%	HYP17-SC	ASN89-MC	84.72%
ASN89-SC	HYP6-MC	55.32%	LYS377-SC	HYP20-MC	59.92%
LYS91-SC	HYP9-SC	23.80%	ASN89-SC	HYP17-MC	47.84%
LYS377-SC	HYP3-SC	21.12%	HYP23-SC	ASP378-SC	32.80%
LYS91-SC	GLU11-SC	18.68%	LYS377-SC	GLY21-MC	30.20%
LYS377-SC	GLY1-MC	16.52%	GLN166-SC	HYP8-MC	18.40%
TYR22-MC	GLU11-SC	15.48%	HYP20-SC	ALA381-MC	16.92%
GLN171-SC	HYP15-SC	11.76%	ASN87-SC	HYP20-SC	15.68%
GLN171-SC	HYP18-SC	11.16%	HYP11-SC	LYS169-MC	15.28%
THR52-SC	GLU11-SC	10.00%	GLN171-SC	HYP11-MC	15.04%
other pairs		129.24%	LYS91-SC	HYP17-SC	12.80%
			other pairs		98.36%
Total		495.12%	Total		542.64%

References

1. Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005) H++: a server for estimating pK(a)s and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33**, W368-W371
2. Brooks, B. R., Brooks, C. L., MacKerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: The biomolecular simulation program. *J. Comp. Chem.* **30**, 1545-1614
3. Nose, S., and Klein, M. L. (1983) Constant pressure molecular-dynamics for molecular-systems. *Mol. Phys.* **50**, 1055-1076
4. Hoover, W. G. (1985) Canonical dynamics - equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695-1697
5. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comp. Chem.* **26**, 1781-1802
6. Schneider, T., and Stoll, E. (1978) Molecular-dynamics study of a 3-dimensional one-component model for distortive phase-transitions. *Phys. Rev. B* **17**, 1302-1322
7. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of N-alkanes. *J. Comput. Phys.* **23**, 327-341
8. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577-8593
9. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586-3616
10. MacKerell, A. D., Feig, M., and Brooks, C. L. (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comp. Chem.* **25**, 1400-1415
11. Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and MacKerell, A. D. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **31**, 671-690
12. Guvench, O., Greene, S. N., Kamath, G., Brady, J. W., Venable, R. M., Pastor, R. W., and MacKerell, A. D. (2008) Additive empirical force field for hexopyranose monosaccharides. *J. Comp. Chem.* **29**, 2543-2564
13. Guvench, O., Hatcher, E., Venable, R. M., Pastor, R. W., and MacKerell, A. D. (2009) CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Chem. Theory Comput.* **5**, 2353-2370
14. Guvench, O., Mallajosyula, S. S., Raman, E. P., Hatcher, E., Vanommeslaeghe, K., Foster, T. J., Jamison, F. W., and MacKerell, A. D. (2011) CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J. Chem. Theory Comput.* **7**, 3162-3180

15. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935
16. Durell, S. R., Brooks, B. R., and Bennaim, A. (1994) Solvent-induced forces between 2 hydrophilic groups. *J. Phys. Chem.* **98**, 2198-2202
17. Yu, W. B., He, X. B., Vanommeslaeghe, K., and MacKerell, A. D. (2012) Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of Computational Chemistry* **33**, 2451-2468
18. Mayne, C. G., Saam, J., Schulten, K., Tajkhorshid, E., and Gumbart, J. C. (2013) Rapid parameterization of small molecules using the force field toolkit. *J Comput Chem*
19. Deng, Y. Q., and Roux, B. (2006) Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2**, 1255-1273
20. Shirts, M. R., and Chodera, J. D. (2008) Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, 1-10
21. Houston, D. R., Recklies, A. D., Krupa, J. C., and van Aalten, D. M. F. (2003) Structure and ligand-induced conformational change of the 39-kDa glycoprotein from human articular chondrocytes. *Journal of Biological Chemistry* **278**, 30206-30212
22. Weis, W. I., and Drickamer, K. (1996) Structural basis of lectin-carbohydrate recognition. *Annual Review of Biochemistry* **65**, 441-473
23. Fusetti, F., Pijning, T., Kalk, K. H., Bos, E., and Dijkstra, B. W. (2003) Crystal structure and carbohydrate-binding properties of the human cartilage glycoprotein-39. *Journal of Biological Chemistry* **278**, 37753-37760
24. Mohanty, A. K., Singh, G., Paramasivam, M., Saravanan, K., Jabeen, T., Sharma, S., Yadav, S., Kaur, P., Kumar, P., Srinivasan, A., and Singh, T. P. (2003) Crystal structure of a novel regulatory 40-kDa mammary gland protein (MGP-40) secreted during involution. *Journal of Biological Chemistry* **278**, 14451-14460
25. Sun, Y. J., Chang, N. C., Hung, S. I., Chang, A. C., Chou, C. C., and Hsiao, C. D. (2001) The crystal structure of a novel mammalian lectin, Ym1, suggests a saccharide binding site. *The Journal of biological chemistry* **276**, 17507-17514
26. Zaheer-ul-Haq, Dalal, P., Aronson, N. N., and Madura, J. D. (2007) Family 18 chitolectins: comparison of MGP40 and HUMGP39. *Biochemical and Biophysical Research Communications* **359**, 221-226