1-4-2017

Inhibition of Mammalian Glycoprotein YKL-40 *IDENTIFICATION OF THE PHYSIOLOGICAL LIGAND*

Abhishek A. Kognole
University of Kentucky, abhishek.kognole@uky.edu

Christina M. Payne
University of Kentucky, christy.payne@uky.edu

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Inhibition of Mammalian Glycoprotein YKL-40 *IDENTIFICATION OF THE PHYSIOLOGICAL LIGAND*

Inhibition of Mammalian Glycoprotein YKL-40
*IDENTIFICATION OF THE PHYSIOLOGICAL LIGAND**[S]

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

From the Department of Chemical and Materials Engineering, University of Kentucky, Lexington, Kentucky 40506

Edited by Gerald W. Hart

YKL-40 is a mammalian glycoprotein associated with progression, severity, and prognosis of chronic inflammatory diseases and a multitude of cancers. Despite this well documented association, identification of the lectin's physiological ligand and, accordingly, biological function has proven experimentally difficult. YKL-40 has been shown to bind chito-oligosaccharides; however, the production of chitin by the human body has not yet been documented. Possible alternative ligands include proteoglycans, polysaccharides, and fibers like collagen, all of which makeup the extracellular matrix. It is likely that YKL-40 is interacting with these alternative polysaccharides or proteins within the body, extending its function to cell biological roles such as mediating cellular receptors and cell adhesion and migration. Here, we consider the feasibility of polysaccharides, including cello-oligosaccharides, hyaluronan, heparan sulfate, heparin, and chondroitin sulfate, and collagen-like peptides as physiological ligands for YKL-40. We use molecular dynamics simulations to resolve the molecular level recognition mechanisms and calculate the free energy of binding the hypothesized ligands to YKL-40, addressing thermodynamic preference relative to chito-oligosaccharides. Our results suggest that chitohexaose and hyaluronan preferentially bind to YKL-40 over collagen, and hyaluronan is likely the preferred physiological ligand, because the negatively charged hyaluronan shows enhanced affinity for YKL-40 over neutral chitohexaose. Collagen binds in two locations at the YKL-40 surface, potentially related to a role in fibrillar formation. Finally, heparin non-specifically binds at the YKL-40 surface, as predicted from structural studies. Overall, YKL-40 likely binds many natural ligands in vivo, but its concurrence with physical maladies may be related to associated increases in hyaluronan.

YKL-40, also known as chitinase 3-like 1, is a mammalian glycoprotein implicated as a biomarker associated with progression, severity, and prognosis of chronic inflammatory dis-

eases and a multitude of cancers (1–4). Many different types of cells including synovial, endothelial, epithelial, smooth muscle, and tumor cells produce YKL-40 in vivo, likely in response to environmental cues (5–8). Speculation as to biological function of YKL-40 varies from both inhibiting and antagonizing collagen fibril formation as a result of injury or disease (9), as well as conferring drug resistance and increasing cell migration leading to progression of cancer (3), and protection from chitin-containing pathogens (10). Although the association of YKL-40 with physical maladies is well documented, identification of the physiological ligand of this lectin, and thus biological function, remains elusive.

Mammalian YKL-40 is classified as a family 18 glycoside hydrolase based on high homology with this well conserved class of enzymes in the CAZy database (10–12). Although similar in structure to family 18 glycoside hydrolases, YKL-40 lacks catalytic activity as a result of substitution of the glutamic acid and aspartic acid motif typical of catalytically active family 18 hydrolases, rendering YKL-40 a lectin, a non-catalytic sugar-binding protein. Structural evidence suggests that YKL-40 exhibits at least two functional binding regions (10). The primary binding cleft has nine binding subsites lined with aromatic residues compatible with carbohydrate binding (Fig. 1). A putative heparin-binding site, located within a surface loop, has also been suggested (Fig. 1), although in vitro binding affinity studies have been unable to conclusively demonstrate this (11).

Binding affinity and structural studies reveal that chito-oligosaccharides are natural substrates (6, 10, 11). In line with family 18 glycoside hydrolases, YKL-40 uniquely binds short and long chito-oligomers, indicating preferential site selection based on affinity (10). Chitohexaose binding has also been purported to induce conformational changes in YKL-40 (10), although this has not been observed in all structural studies (11). Lectin binding niches are widely believed to be "preformed" to the preferred ligand, exhibiting little conformational change upon binding (13, 14). Despite apparent affinity, chitin is not a natural biopolymer within mammalian or bacterial cells, and the presence of chitin or chito-oligosaccharides in mammals is likely related to fungal infection (15). The noted up-regulation of YKL-40 in response to inflammation lends credence to the argument that YKL-40 functions as part of the innate immune response in recognition of self from non-self (6, 16); however, high expression levels of YKL-40 in carcinoma tissues suggest function beyond the innate immune response may also exist (17, 18). The extracellular matrix is comprised of a mesh of proteoglycans (protein-attached glycosaminogly-
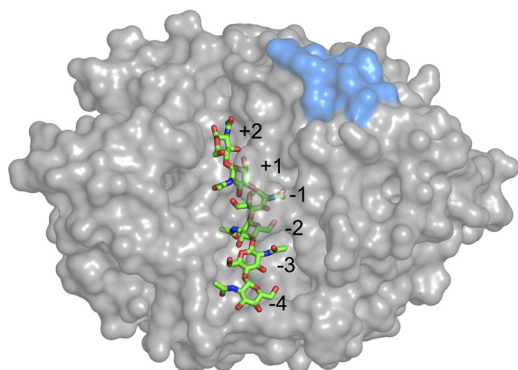
FIGURE 1. **Surface representation of YKL-40 showing the binding cleft with a bound hexamer of chitin.** Binding sites +2 through −4 are numbered. Sites −5, −6, and −7 have also been identified but are not shown. The putative heparin-binding site is shown in *blue*.

cans), polysaccharides, and fibers including collagen. An alternate theory to the pathogenic protection function is that a closely related polysaccharide, instead of chitin, plays the role of the physiological ligand in mediating cellular function (11).

Despite the structural similarity between chito-oligosaccharides and the proteoglycan carbohydrate monomers, little evidence of polysaccharide binding beyond the original structural studies exists (10, 11). In fact, we are aware of only one other study focusing on the molecular-level mechanism of carbohydrate binding in YKL-40 (19). From a bioinformatics and structural comparison of YKL-40 to a similar chi-lectin, mammary gland protein 40 (20), the authors propose an oligosaccharide binding mechanism that involves tryptophan-mediated gating of the primary carbohydrate binding site (Fig. 1) (19). However, in lieu of a dynamics-based investigation, little can be concluded about the binding mechanism of YKL-40 ligands other than chito-oligosaccharides, and conformational changes relative to binding are inaccessible. From protein purification techniques, namely heparin-Sepharose chromatography, we also know that YKL-40 reversibly binds heparin (7, 11, 21); however, affinity data for this interaction do not exist. Based on the interaction with heparin, it is reasonable to hypothesize heparan sulfate glycosaminoglycans, existing as part of the extracellular matrix construct, are potential physiological ligands. Visual inspection of the protein structure suggests that heparan sulfate fragments may be easier to accommodate within the carbohydrate-binding site than heparin itself (11). It follows that other structurally similar carbohydrate fragments would bind with similar affinity in a comparable mechanism.

The association of YKL-40 with ailments such as arthritis, fibrosis, and joint disease is suggestive of molecular-level interactions with connective tissue and thus collagen (22–26). Motivated by understanding the physiological role of YKL-40 in connective tissue remodeling and inflammation, Bigg *et al.* (9) investigated association of YKL-40 with collagen types I, II, and III using affinity chromatography to confirm binding to each type. The authors report YKL-40 specifically binds to all three collagen types. Additionally, the authors used surface plasmon resonance to confirm binding to collagen type I. Unfortunately, the reported affinity constants were inconsistent across experiments as a result of aggregation. Nevertheless, the work clearly indicates that YKL-40 is capable of binding collagen. This fur-

ther confounds the question of mechanism when considering physiological ligands, because YKL-40 is capable of binding both protein and carbohydrates.

Understanding the mechanism and affinity by which YKL-40 binds ligands is crucial to our understanding of its physiological function. This knowledge will serve as a foundation for future campaigns toward rational development of a potent antagonists enabling cell biological study and addressing YKL-40 as a therapeutic target. To accomplish this goal, we must describe the molecular level mechanisms governing the interaction of YKL-40 with both polysaccharide and collagen-like polypeptides and quantitatively evaluate affinity. In this study, we used classical molecular dynamics (MD)[2] simulations to differentiate modes of ligand recognition and specificity. Using free energy perturbation with replica exchange molecular dynamics (FEP/λ-REMD) and umbrella sampling MD, we quantitatively determined affinities overcoming the experimental difficulties encountered thus far. As physiological ligands, we considered several options within both the polysaccharide and proteinaceous classes. Below, we provide a brief description of each carbohydrate ligand considered, as well as justification for consideration of the collagen-like models.

Polysaccharides—We selected polysaccharides for this study based on their similarity to chito-oligomers, as well as natural occurrence in mammalian cell walls and/or the extracellular matrix. The chito-oligomer from structural studies was included as a control (10, 11). Chitin is a naturally occurring biopolymer comprised of repeating *N*-acetyl-D-glucosamine (GlcNAc) monomeric units connected by β-1,4-glycosidic linkages (Fig. 2). The central ring of the monomer, a six-membered pyranose, is common to a number of carbohydrates including glucose. Given the chemical similarity, as well as the general presence of glucose in mammalian cells as a form of energy, a hexameric cello-oligomer was also examined as a potential physiological ligand, despite its unlikely presence among mammalian glycosaminoglycans.

As described above, YKL-40 binds heparin, and thus, likely also binds heparan sulfate. Heparan sulfate, a less sulfated form of heparin, is a polysaccharide found in abundance in the extracellular matrix and at the cell surface (27). Heparan sulfate is constructed from a repeating disaccharide of β-D-glucuronic acid and *N*-acetyl-α-D-glucosamine (Fig. 2). Of all the glycosaminoglycans, heparan sulfate is the most structurally complex. At least 24 different combinations of the disaccharide monomer exist, with differences arising as a result of variation in both isomer and degree of side chain sulfation (28). Additionally, the heparan sulfate polysaccharide can exhibit both sulfated and unsulfated domains. Physiologically, the unsulfated disaccharide β-D-glucuronic acid (1, 4) *N*-acetyl-α-D-glucosamine is the most prevalent form of heparan sulfate (28). Focusing on the most relevant physiological ligands, we examined the fully sulfated form heparin and the unsulfated form heparan sulfate.
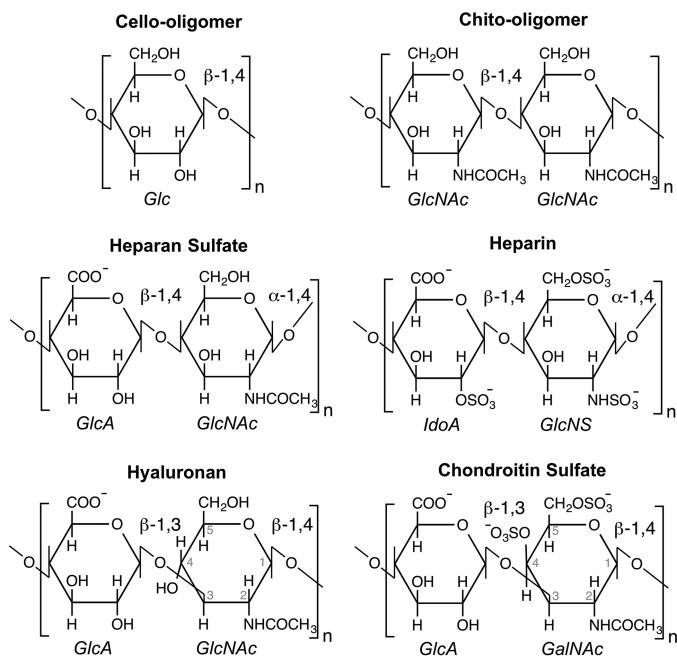

FIGURE 2. **Monomeric units of the polysaccharides considered as potential physiological ligands of YKL-40: cellohexaose, chitohexaose, heparan sulfate, heparin, hyaluronan, and chondroitin sulfate.** The chito-oligomer is a polymer of β-1,4-linked GlcNAc monomers. Heparan sulfate was modeled as a β-1,4, α-1,4-linked chain of GlcA and GlcNAc. Heparin was represented as the β-1,4, α-1,4-linked oligomer of GlcA and GlcNS. Hyaluronan and chondroitin sulfate are β-1,3, β-1,4-linked oligomers; the former consists of GlcA and GlcNAc, and the latter consists of GlcA and GalNAc. *GlcA*, β-D-glucuronic acid; *IdoA*, α-D-iduronic acid; *GlcNS*, *N*-sulfo-α-D-glucosamine; *GalNAc*, *N*-acetyl-β-D-galactosamine.

Hyaluronan is a particularly interesting glycosaminoglycan relative to this study, because chito-oligosaccharides are precursors to hyaluronan synthesis *in vivo* (29–31). The structural relationship of these two molecules is such that binding mechanisms may be similar at alternating binding sites. Hyaluronan is a polysaccharide of a repeating β-D-glucuronic acid and *N*-acetyl-β-D-glucosamine disaccharides connected by alternating β-1,3- and β-1,4-glycosidic linkages (Fig. 2). As with heparan sulfate, hyaluronan is also a glycosaminoglycan comprising the extracellular matrix. At extracellular pH, the carboxyl groups of glucuronic acid are fully ionized, giving the ligand an overall negative charge under typical physiological conditions (32).

Finally, we consider chondroitin sulfate, which is also a glycosaminoglycan prevalent in mammals. The primary structural units of chondroitin sulfate are a repeating β-D-glucuronic acid and *N*-acetyl-α-D-galactosamine disaccharides connected by alternating β-1,3- and β-1,4-glycosidic linkages (Fig. 2). As with heparan sulfate, chondroitin sulfate exists in variably sulfated types (33); we have selected the C4 and C6 sulfated variant of chondroitin sulfate polysaccharide as a model.

Collagen-like Peptide Models—Collagen has also been considered as a potential physiological ligand based on the noted affinity and participation of YKL-40 in collagen fibril formation (9). Collagen, unlike the other potential physiological ligands, is a macromolecular protein with a triple helix structure. There are at least 27 distinct types of human collagen, forming a variety of biological networks, all of which are constructed of a basic

Gly-Xaa-Yaa repeating amino acid sequence (34). Generally, the unspecified amino acids, Xaa and Yaa, are proline and hydroxyproline, respectively.

Model collagen peptides have been observed in two different symmetries: the original Rich and Crick model with 10/3 symmetry (10 units in 3 turns) and the 7/2 symmetry of a more tightly symmetrical triple helix (35–37). On the molecular scale, collagen type will have relatively little impact on binding to YKL-40. However, symmetry may have an impact on hydrogen bonding in the binding site, and thus, overall affinity, which will provide unique insight into physiological relevance. To date, model collagen peptides of a true 10/3 symmetry have not been reported. Rather, the peptides either have a 7/2 helical pitch or are somewhat "intermediate" in symmetry leading some to believe that the 7/2 symmetry is representative of the true collagen helical structure (38). However, it is not known how universally true this hypothesis is because the structures of model peptides capture just a small subsection of the larger macromolecular structure (39).

With a broad range of possible collagen architectures, we have selected four representative model collagen peptides whose structures are both available from crystallographic evidence and span the 10/3 and 7/2 symmetries to the greatest possible extent. The first collagen peptide considered is that of the basic collagen peptide model, PDB code 1CAG (40). The reported 1.9 Å resolution structure exhibits a single Gly to Ala substitution and 7/2 symmetry overall. Near the substitution site, the helix relaxes somewhat from 7/2 symmetry, although not so much as to change overall symmetry. The second collagen model peptide we consider is a variation of the 1CAG peptide, where we reverted the alanine substitution to its native glycine. Minimization of this structure returns the helix to full 7/2 symmetry; we refer to this peptide as "native 1CAG" here. The third model represents a segment from type III homotrimer collagen with approximate 10/3 symmetry in the middle part of the helix (PDB code 1BKV) (35). This middle part of the 1BKV model peptide, also referred as the T3–785 peptide, has an imino acid-poor sequence of GITGARGLA. Our fourth model, 1Q7D, is a triple helical collagen-like peptide sequence including a hexapeptide Gly-Phe-Hyp-Gly-Glu-Arg (GFOGER) motif in the middle (41); this motif is not sufficiently long to exhibit 10/3 symmetry, exhibiting, rather, an intermediate degree of 7/2 helical symmetry. This 1Q7D model is known to bind the integrin $\alpha 2 \beta 1$-I domain protein (42), and the GFOGER motif is found in the $\alpha 1$ chain of type I collagen.

Experimental Procedures

Molecular Dynamics Simulations

MD simulations were constructed starting from the chito-hexaose-bound YKL-40 structure (PDB code 1HJW) deposited by Houston *et al.* (10). The apo simulation simply removed the chito-oligomer from the primary binding cleft. The *N*-linked glycan captured in the YKL-40 structure, at Asn[60], was included in system preparation. Because crystal structures of YKL-40 bound to other polysaccharides are not available, we used the structural similarity of polysaccharides as the basis for modeling the remaining polysaccharides in this investigation. In the
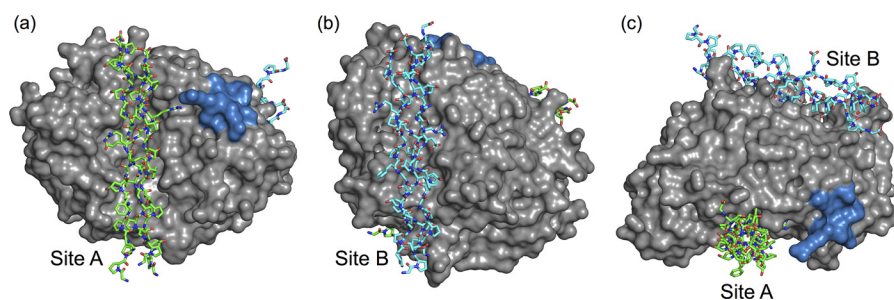
FIGURE 3. **Molecular shape complementarity docking calculations predict collagen-like peptides will bind to YKL-40 in two possible orientations.** *a*, the front view of YKL-40 (*gray surface*) with collagen docked in site A (*green stick*). *b*, the back view where collagen is docked in site B (*cyan stick*). *c*, top view of YKL-40 illustrating relative positions of binding sites. The putative heparin-binding subsite is shown in *blue surface* to aid in visualization of relative orientation of the protein·protein complexes. This particular figure shows the integrin-binding collagen peptide, 1Q7D (41), in the predicted binding sites along the surface of YKL-40; similar docking was carried out for other collagen models.

case of cellohexaose, hyaluronan, heparan sulfate, heparin, and chondroitin sulfate, we located the central ring atoms of the ligand backbone in the same location as that of the original chitohexaose. Appropriate pyranose side chains and glycosidic linkages (Fig. 2) were added using CHARMM internal coordinate tables to construct the remainder of the sugar residue (43). All polysaccharides were described using the CHARMM36 carbohydrate force field (44–46). The missing force-field parameters for *N*-sulfated glucosamine (supplemental Fig. S1) in heparin were developed using the force-field Toolkit plugin for VMD (47, 48). Details of this parameterization and output parameters (supplemental Table S2) are reported in the supplemental materials.

Construction of the collagen-bound YKL-40 models required docking calculations to appropriately position the ligand. The collagen peptides are significantly larger than any of the carbohydrate ligands; thus, it is unlikely that a collagen molecule occupies the primary YKL-40 binding site in the same manner as chito-oligomer. Standard affinity-based docking calculations, such as the ones performed in AutoDock, are not feasible for determination of an initial collagen-binding domain given the size and flexibility of the triple helix structures. Rather, the collagen peptides were docked on the basis of molecular shape complementarity using the online web server PatchDock Beta version 1.3 (49, 50). In the case of each of the four collagen-like model peptides, PatchDock predicted two potential occupancies along the surface of YKL-40, site A and site B. Binding site A corresponds to the primary carbohydrate-binding domain of YKL-40; however, the collagen ligand was not as deeply entrenched in the cleft as chitohexaose. Binding site B is located on the opposite side of YKL-40 from the primary binding cleft. Thus, for each collagen-like peptide, two MD simulations were constructed representing the two potential binding sites. Fig. 3 illustrates the results of the docking with predicted collagen binding sites A and B for the 1Q7D collagen-like model peptide.

The constructed protein-ligand systems were minimized in vacuum and subsequently solvated with water and sodium ions. Using CHARMM (43), the solvated systems were extensively minimized and heated to 300 K for 20 ps, which was followed by MD simulation for 100 ps in the *NPT* ensemble. The coordinates following density equilibration were used as a starting point for 250 ns of MD simulation in the *NVT* ensemble at

300 K using NAMD (51). Explicit procedural details are provided in the supplemental materials.

Free Energy Calculations

FEP/λ-REMD—Free energy perturbation with Hamiltonian replica exchange molecular dynamics (FEP/λ-REMD) was used to calculate the absolute free energy of binding the polysaccharide ligands to YKL-40 (52, 53). This protocol uses Hamiltonian replica exchange as a means of improving the Boltzmann sampling of free energy perturbation calculations. The parallel/parallel replica exchange MD algorithm in NAMD was implemented as recently described (51, 54). The free energy calculations performed using this approach were accomplished through two separate sets of free energy calculations following the thermodynamic cycle illustrated in supplemental Fig. S2. To obtain each binding free energy, ΔG, the bound carbohydrate ligand was first decoupled from the solvated protein·carbohydrate complex to determine ΔG_1. The second calculation entailed decoupling the solvated oligosaccharide from solution into vacuum to obtain ΔG_2. The difference between the two values, $\Delta G_2 - \Delta G_1$, gives the absolute free energy of binding the given ligand to YKL-40.

In each free energy calculation, five separate terms contribute to the potential energy of the system: the non-interacting ligand potential energy, repulsive and dispersive contributions to the Lennard-Jones potential, electrostatic contributions, and the restraining potential. In each calculation, the ligand was decoupled from the system by thermodynamic coupling parameters controlling the non-bonded interaction of the ligand with the environment. The parameters decoupled the ligand in a four-stage process, wherein the coupling parameters defined replicas that were exchanged along the length of the alchemical pathway. This decoupling has been described in detail previously (54) and is also reported in the supplemental materials. A total of 128 FEP replicas were used, and a conventional Metropolis Monte Carlo exchange criterion governed the swaps throughout the replica exchange process (53). The free energy of binding was determined from 20 consecutive, 0.1-ns simulations of each corresponding system, where the first 1 ns of data were discarded as equilibration. One standard deviation of the last 1 ns of simulation data were used to obtain an estimate of error. Additional details about these calculations are described in the supplemental materials.
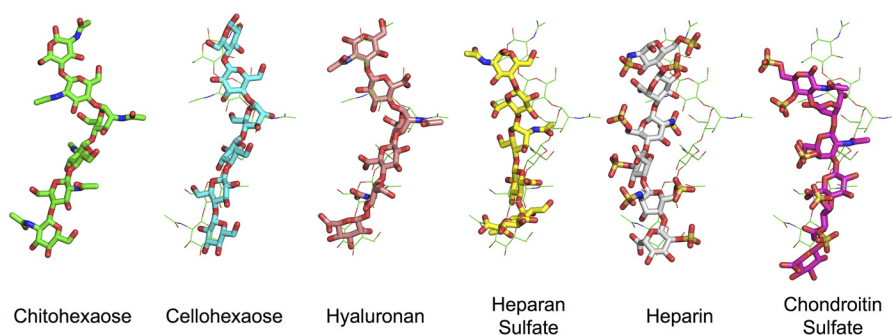
FIGURE 4. Relaxation of the polysaccharide ligands in the primary binding cleft of YKL-40. Each ligand is shown after a 100-ps equilibration in a thick stick representation. For comparison, the chito-oligomer, in its equilibrated conformation, is shown in *thin green lines* behind each oligosaccharide. The YKL-40 protein has been aligned such that each oligosaccharide is oriented in the same manner; although YKL-40 is not shown for visual clarity. Heparan sulfate, heparin, and chondroitin sulfate relax significantly from the initial distorted conformation.

Umbrella Sampling—Convergence challenges make FEP/λ-REMD inappropriate for determining the binding free energy of the much larger and more flexible collagen-like model peptides. Thus, umbrella sampling was used to determine the work required to detach the collagen ligands from the shallow clefts of YKL-40. Over the entire reaction coordinate, this value equates to binding affinity, enabling relative comparison of collagen peptide affinity. The MD umbrella sampling simulations used a native contacts-based reaction coordinate analogous to that defined by Sheinerman and Brooks (55) and as implemented in recent cellulose decrystallization studies (56, 57). Here, a native contact was defined as a YKL-40 protein residue within 12 Å of a collagen peptide residue; distance was defined by center of geometry of a given residue. The cutoff distance was selected to be larger than the non-bonded cutoff distance, ensuring that the collagen ligand was no longer interacting with YKL-40. Additionally, the water boxes of the collagen-YKL-40 systems were made bigger to accommodate the required separation distance.

The change in free energy was determined as a function of the reaction coordinate, ρ, formulated as the weighted sum of the states of the native contacts. The initial coordinates of the bound systems were selected from 250-ns equilibrated snapshots. The initial number of native contacts and their weights were calculated from these snapshots. An initial reaction coordinate of 0 (normalized) corresponds to this initial condition, and a final reaction coordinate of 1 corresponds to all of the native contacts being outside the 12-Å cutoff (*i.e.* the ligand is decoupled and freely sampling the bulk). The reaction coordinate was divided into 20 windows evenly spaced along the reaction coordinate, and each window was sampled for 5 ns, where the reaction coordinate was maintained at the specified value using a harmonic biasing force with the force constant of 41,840 kJ/mol. The potentials of mean force profiles were calculated using the weighted histogram analysis method (WHAM software), and error analysis was performed using bootstrapping. A detailed listing of all the simulations and free energy calculations performed as part of the objectives of this study is provided in supplemental Table S1.

Results and Discussion

Protein-Polysaccharide Binding in YKL-40—MD simulation suggests that of the six polysaccharide oligomers investigated,

only three bind in a stable fashion in the primary carbohydrate binding site of YKL-40. The three potential polysaccharide physiological ligands at this site include chitohexaose, cellohexaose, and hyaluronan. In the section that follows, we will describe the dynamics of chitohexaose, cellohexaose, and hyaluronan binding to YKL-40. The remaining three ligands—heparin, heparan sulfate, and chondroitin sulfate—were dislodged from the binding site over the course of MD simulations. The α-1,4-glycosidic linkages in heparin and heparan sulfate, instead of β-1,4, modifies the relative orientation of disaccharide monomers from that of the chito-oligosaccharide. The NMR solution structure of heparin (PDB code 1HPN) shows that the relaxed conformation is semi-helical (59), which cannot be feasibly accommodated in the conserved, narrow carbohydrate binding site of YKL-40. Heparan sulfate suffers from similar steric constraints posed by the relaxation driving force. The bulky sulfated side chains of heparin introduce further steric hindrance and, in the case of heparin and chondroitin sulfate, unfavorably strong electrostatic interactions resulting from negatively charged moieties inconveniently located along the cleft (*i.e.* without a co-located, oppositely charged protein residue) eject the ligands from the cleft.

In the cases of heparin, heparan sulfate, and chondroitin sulfate, the ligands quickly "relax" from the initial wide V-shaped conformation as they are expelled from the cleft by charge- and steric-based effects. Relaxation of the sugar from the initial binding pose is sufficient to initiate loss of critical non-bonded interactions along with a subsequent reduction in affinity (Fig. 4). Within 25 ns, heparin, heparan sulfate, and chondroitin sulfate were expelled from the cleft into bulk solution. Each of the three ligands capable of binding with the primary binding cleft maintained the −1 boat conformation over the entire simulation. Chitohexaose and cellohexaose remained in the binding cleft over the entire 250-ns MD simulation while maintaining the initial wide V shape. Hyaluronan developed a sharp V shape within a few nanoseconds and maintained this conformation within the binding cleft for the remainder of the simulation (supplemental Fig. S6); this is primarily due to variation in glycosidic linkage, where hyaluronan exhibits a β-1,3-linkage within the monomer instead of the β-1,4-linkage of cello- and chitohexaose. Also, comparison of the equilibrated chitohexaose- and hyaluronan-bound structures disabuses one of
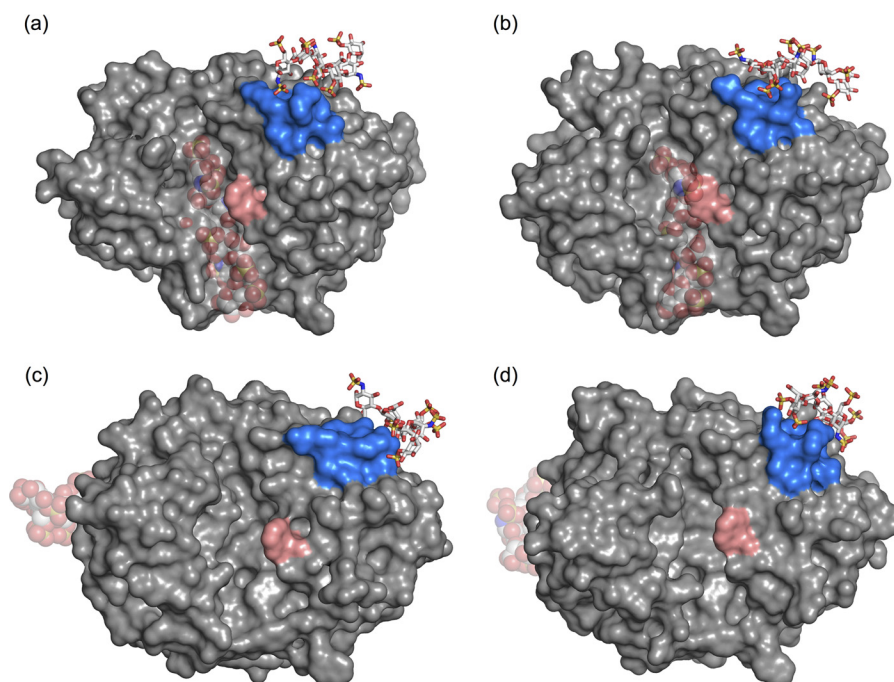
FIGURE 5. Snapshots from four independent MD simulations of heparin (*white stick*) binding to a putative heparin-binding site (*blue surface*) of YKL-40 (*gray surface*). The primary oligosaccharide binding site of YKL-40 is marked by an aromatic residue shown in salmon surface representation. *Transparent spheres* illustrate the initial simulation positions of heparin. In two cases, *a* and *b*, the heparin ligand was initially bound in the primary YKL-40 binding site. In both cases, the ligand was expelled from the primary binding site into solution and located the heparin-binding site through electrostatic interactions. Two additional simulations, *c* and *d*, were initialized with the heparin ligand free in solution. Again, the ligands identified the YKL-40 heparin-binding domain through charge-based interactions. The ligand did not specifically bind in a particular conformation. Rather, the ligand dynamically interacted with the heparin-binding domain.

the notion that similar binding mechanisms exist at alternate binding sites, because only the −1 site pyranose appears to maintain similar side chain orientation.

The native distorted conformation is characteristic of glycoside hydrolase pyranose binding behavior in the −1 site (Fig. 4) (60). In solution, polysaccharide pyranose moieties adopt the energetically favorable chair conformation (61); however, when bound to an enzyme, the active sites of catalytically active glycoside hydrolases distort the pyranose ring in the −1 binding subsite into a less energetically favorable conformation, such as a boat or skew conformation (62–65), priming the substrate for hydrolytic cleavage. Interestingly, the chitohexaose ligand bound in the primary binding site of YKL-40 exhibits a boat conformation despite not being catalytically active (10). This suggests that the sugar distortion in the −1 binding site contributes to ligand binding as well as catalysis, because there is no evolutionary requirement to overcome an activation energy barrier in catalytically inactive lectins. A recent study of a homologous chitinase suggested that −1 pyranose relaxation reduces binding affinity and affords the ligand more flexibility and entropic freedom (66), which is consistent with our findings from the 250-ns MD here.

Putative Heparin-binding Site—Despite the fact that the heparin oligomer could not be accommodated by the YKL-40 binding cleft, MD simulations do suggest that the oligomer interacts with the surface of YKL-40 at a putative heparin-binding site (Fig. 1). After ejection from the primary binding site, the oligomer spontaneously binds to the YKL-40 heparin-binding site (supplemental Movie S1). To address the significance of this unanticipated event, we performed three additional independent MD simulations of the YKL-40/heparin system: one with a new random number seed, although in the same configuration, and two additional simulations with the ligand randomly placed in solution (supplemental Movie S2). In each case, the heparin oligomers were capable of finding and binding to a group of charged residues at the surface of YKL-40 (Fig. 5); these were the basic residues of a putative heparin-binding site, GRRDKQH, at positions 143–149. Interestingly, this domain follows a consensus protein sequence, *X*BB*X*B*X* (where B is a basic residue and *X* is any non-basic amino acid), that is noted for its ability to recognize polyanions like heparin (67). In all four cases, heparin recognized the binding site within 25 ns of MD simulation (Fig. 5), occasionally visiting other moderately basic surface locations before localizing around the GRRDKQH motif. The strong electrostatic interaction arose from the dynamic formation of salt bridges between either the sulfate or the carboxyl groups of the heparin oligosaccharide and the side chains of the basic amino acids (supplemental Table S3). Lys[155] and Lys[193] also exhibit large, favorable interactions with heparin, interacting with the opposite end of the polysaccharide ligand as it binds at the surface. Coupled with experimental observation of heparin affinity, our MD simulations suggest a non-specific, surface-mediated binding interaction between YKL-40 and the extensively sulfated heparin oligomer (10, 11). Although the unsulfated variant, heparan sulfate, did not visit the heparin-binding site, chondroitin sulfate also attached to the putative heparin-binding site in a similar fashion to heparin. Given the chemical similarity of these glycosaminoglycans, *i.e.* highly sulfated and negatively charged, we anticipate the

TABLE 1
Energetic components of the free energy of ligand binding to YKL-40

System	ΔG_{repu}	ΔG_{disp}	ΔG_{elec}	ΔG_{rstr}	ΔG_{Tot}	ΔG_b
	kJ mol^{-1}	*kJ mol^{-1}*	*kJ mol^{-1}*	*kJ mol^{-1}*	*kJ mol^{-1}*	*kJ mol^{-1}*
YKL-40 + Cellohexaose	379.5 ± 3.8	−382.3 ± 2.1	−306.6 ± 2.2	−1.1	−310.5 ± 3.6	−12.6 ± 3.7
Cellohexaose	316.2 ± 1.8	−287.4 ± 1.3	−326.7 ± 1.6	0	−297.9 ± 2.8	
YKL-40 + Chitohexaose	535.9 ± 11.7	−538.9 ± 4.2	−407.7 ± 7.5	−5.5	−416.2 ± 11.2	−63.3 ± 12.6
Chitohexaose[a]	329.6 ± 4.5	−306.2 ± 3.0	−376.3 ± 3.4	0	−352.9 ± 5.9	
YKL-40 + Hyaluronan	438.1 ± 2.8	−439.5 ± 1.5	−1395.9 ± 4.2	−1.3	−1398.6 ± 4.3	−106.8 ± 4.6
Hyaluronan	333.1 ± 1.4	−306.0 ± 1.3	−1318.9 ± 1.5	0	−1291.8 ± 2.5	

[a] Hamre *et al.* (68).

XBBXBX motif may also routinely appear in chondroitin-binding proteins.

Polysaccharide Ligand Binding Affinity—Each of the three polysaccharides maintaining contact with the primary binding site of YKL-40, cellohexaose (or likely any glucose derivative), chitohexaose, and hyaluronan, are feasible ligands. However, free energy calculations suggest that hyaluronan may preferentially bind with YKL-40 if chitin is not present in the human body as a result of fungal infection. The absolute free energies of binding cellohexaose, chitohexaose, and hyaluronan to YKL-40 were −12.6 ± 3.7, −63.3 ± 12.6, and −106.8 ± 4.6 kJ/mol, respectively. Repulsive, dispersive, and electrostatic components of the free energy changes are tabulated in Table 1. Convergence assessment of the free energy calculations has been provided in supplemental Fig S3. We recently calculated the free energy of solvation for chitohexaose as part of a study on family 18 chitinases (68); this value has been used in our calculation of chitohexaose binding affinity to YKL-40 for computational efficiency. The methods used to calculate solvation free energy of chitohexaose were identical to those described here. Furthermore, the binding free energy of chitohexaose to YKL-40 is in good agreement with that of homologous family 18 chitinases, despite mutation of the catalytic motif in the lectin.

Chitohexaose and cellohexaose are both neutral ligands but display a significant difference in binding affinity to YKL-40. Electrostatic interactions appear to be one of the more significant contributors to the enhanced affinity of chitohexaose over cellohexaose (Table 1). For cellohexaose, the change in the electrostatic component of binding free energy was unfavorable (20.1 ± 2.7 kJ/mol), whereas the same component for chitohexaose was energetically favorable (−31.4 ± 8.2 kJ/mol). In the case of hyaluronan, electrostatic interactions play an even greater role in enhancing affinity of the ligand for YKL-40 (−77.0 ± 4.5 kJ/mol). This increasing electrostatic contribution is reflective of increasing number of electronegative atoms in the side chains of carbohydrates as we go from cellopentaose to chitohexaose to hyaluronan. We observe no significant differences in cellohexaose, chitohexaose, or hyaluronan binding to YKL-40 arising from Weeks-Chandler-Anderson (WCA) dispersion and repulsion (Table 1). This is largely a function of the molecular similarity of the pyranose rings comprising the monomeric units of three oligosaccharides (Fig. 2). The pyranose rings of carbohydrates bound in the active sites of glycoside hydrolases, and by extension the binding clefts of lectins, form carbohydrate-π stacking interactions with surrounding aromatic residues along the clefts (69). In YKL-40, these stacking interactions are formed in the −3 and −1 binding sites with residues Trp[31] and Trp[352], respectively. Naturally, any polysac-

charide ligand capable of binding in the YKL-40 binding cleft will likely exhibit a similarly favorable WCA binding free energy component. In the following section, we expand upon the molecular level interactions that contribute to polysaccharide binding affinity in YKL-40.

Based on these results, it is unlikely that a cello-oligomer would bind in the cleft of YKL-40 over a chito-oligomer, and thus, although there is potential for YKL-40 to bind a cello-oligomer or glucose, it would not be inhibitory. Hyaluronan, on the other hand, likely competes with chito-oligomers in binding, which is due in large part to the electrostatic favorability of the charged side chains of hyaluronan in the YKL-40 binding cleft. Clinical data support hyaluronan as a biomarker for cancer prognosis and inflammation (32, 70), the same events in which YKL-40 appears at elevated serum levels (5). To our knowledge, there are no studies evaluating the co-existence of YKL-40 and hyaluronan. The cell receptor protein CD44 has been implicated in hyaluronan binding interactions and is also involved in confounding scenarios, both aggravating and improving inflammation (71). Sequence alignment of YKL-40 with the hyaluronan-binding domain of human CD44 (72), using BLASTP 2.3.0 (73), shows no homology, further suggesting that this YKL-40-hyaluronan binding is different from previously known hyaluronan-binding proteins (74).

Polysaccharide Binding Dynamics—YKL-40 is highly homologous with carbohydrate-active enzymes found in glycoside hydrolase family 18 (12, 75). Despite lacking catalytic ability, the primary polysaccharide binding site of YKL-40 exhibits remarkable similarity to these family 18 chitinases. As such, one may reasonably expect that ligand binding within this family will demonstrate similar trends, regardless of evolutionary origin. Indeed, we observe that chitohexaose, cellohexaose, and hyaluronan binding in the primary binding site of YKL-40 follow a general pattern common to carbohydrate-active enzymes. Namely, that ligand binding interactions are mediated by carbohydrate-π stacking interactions with aromatic residues, and hydrogen bonding interactions are critical to overall ligand affinity and stability. We investigate these trends quantitatively through analysis of the MD simulation trajectories, including root mean square deviation (RMSD) of the protein, root mean square fluctuation (RMSF) of both the protein and the ligands over the course of the simulation, hydrogen bonding analysis, degree of solvation of the ligand, and interaction energy of the ligand with the protein.

Cellohexaose, chitohexaose, and hyaluronan binding in the primary YKL-40 binding site did not adversely affect protein dynamics. In each case, binding the polysaccharide ligand did not significantly disturb the protein backbone (*i.e.* protein fold),
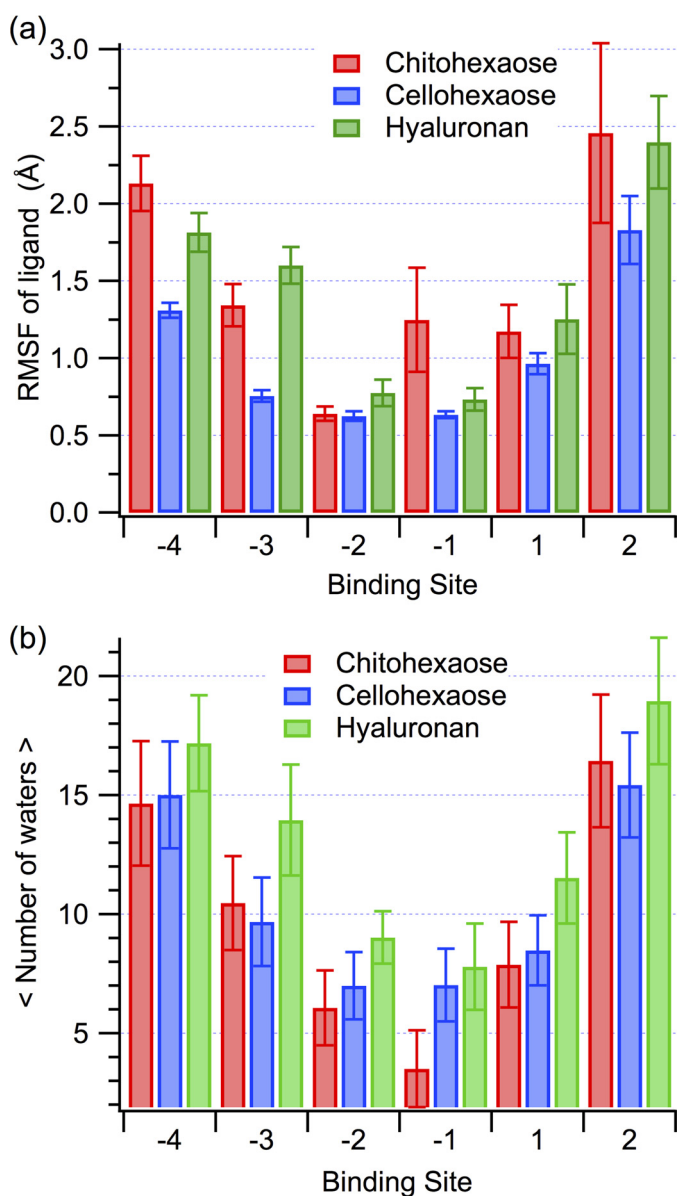
FIGURE 6. *a*, RMSF of the polysaccharide ligands on a per-binding-subsite basis. The *error bars* were calculated using block averaging over 2.5 ns. *b*, average number of water molecules within 3.5 Å of each ligand monomer. The *error bars* represent one standard deviation.

and the ligand remained relatively unperturbed over the course of the simulation. The relatively consistent RMSD of the protein backbones suggests that the simulations reached a local equilibrium (supplemental Fig. S4*a*). As with the RMSD calculation, the RMSF of the protein backbone suggests that the binding of polysaccharides does little to disturb the overall protein conformation (supplemental Fig. S4*b*). Additionally, we did not observe significant conformational changes in the carbohydrate-binding site of YKL-40 before or after ligand binding (supplemental Fig. S5). More details of protein dynamics and conformational changes have been provided in the supplemental text.

The RMSF of the ligand, averaged over 250 ns as a function of binding site, provides a measure of relative ligand stability. Error was estimated by block averaging over 2.5-ns blocks. Ligand stability over the course of the simulation suggests hya-

luronan is as stable, if not more so, as chitohexaose in the primary binding site (Fig. 6*a*); however, cellohexaose appears to be more stable relative to the two other ligands at the ends of the binding cleft. This latter finding is a function of the length of the side chains attached to the pyranose rings of each of the ligands. Of the three carbohydrates, the cello-oligomer has the shortest side chains, and thus, the ligand fluctuates less because it does not need to rearrange as significantly to induce binding. As shown above, this does not necessarily correspond to the most thermodynamically preferential ligand, and lower RMSF could also be interpreted hypothetically as loss of translational and conformational freedom, resulting in unfavorable entropic contribution.

The hydrogen-bonding partners of chitohexaose, cellohexaose, and hyaluronan are quite different, largely as a result of the conformational change of hyaluronan (supplemental Table S4). Defining a hydrogen bond as a polar atom within 3.4 Å and 60° of a donor, we identified the formation of donor-acceptor pairs and percent occupancy of these hydrogen bonds between the protein and each carbohydrate moiety over the course of the 250-ns MD simulations (supplemental Table S4). As described above, hyaluronan formed a sharp V shape in the polysaccharide binding cleft, which minimized steric hindrance and, in turn, modified accessible hydrogen bonding partners relative to chito- and cellohexaose. Hydrogen bonds at the +1 site, between glucuronic acid and Asp[207], Arg[263], and Tyr[141], stabilized the hyaluronan conformation (supplemental Table S4). In the −1 subsite, chitohexaose primarily hydrogen bonds with the side chain of Tyr[206] and the main chain of Trp[99]. In the cases of both cellohexaose and hyaluronan, the interaction with Tyr[206] was abolished and instead supplemented by Trp[99] alone. In the −2 subsite, the oxygen of the chitohexaose acetyl forms a long-lived hydrogen bond with the indole nitrogen of the buried Trp[352]; neither hyaluronan nor cellohexaose interact with the −2 site through this tryptophan. Rather, Trp[31], which stacks with the pyranose in the −3 subsite, acts as a hydrogen donor to the −2 subsite glucuronic acid side chain of hyaluronan. In the case of cellohexaose, the main chain of a solvent-exposed asparagine, Asn[100], almost exclusively mediates hydrogen bonding in the −2 site. The +2, −3, and −4 binding subsites exhibit less frequent hydrogen bonding between the ligand and the protein, and there is little consistency in bonding partners across ligands. Certainly, these variations will manifest in enthalpic contributions to ligand binding, because even a single hydrogen bond may account for 4–29 kJ/mol of binding free energy in biological systems (76, 77); such is likely the case for cellohexaose and chitohexaose binding to YKL-40, where the latter exhibits both greater hydrogen bonding capability and a more favorable binding free energy.

Key aromatic residues, Trp[31], Trp[99], and Trp[352], play a significant role in binding all three oligosaccharides. Notably, these tryptophans are conserved in other lectins, including mammary gland protein (MGP-40) and mammalian lectin Ym1 (20, 78). According to previous structural studies, these aromatic residues form hydrophobic stacking interactions with pyranose moieties at the −3, +1, and −1 binding subsites, respectively (11). As mentioned above, this carbohydrate-π stacking was observed across the three polysaccharide ligands
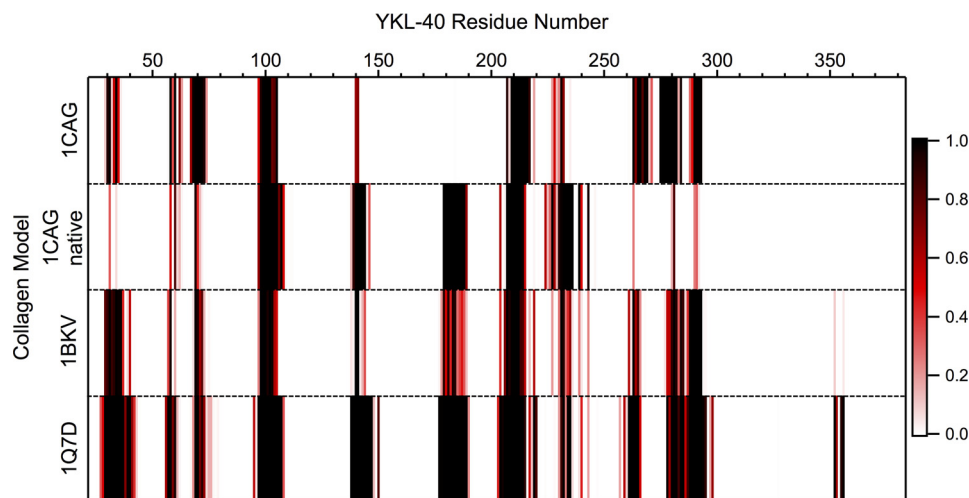
FIGURE 7. **Native contact analysis of each collagen-like peptide model binding to YKL-40 at site A.** The color scale represents the normalized frequency (*i.e.* fractional percentage of frames in which the contact was formed) of the respective YKL-40 residue as a native contact. A native contact was defined as any time a collagen residue was within 12 Å of a YKL-40 residue, where distance was defined by center of geometry of a given residue. Only frames from the last 100 ns of simulation, following a period of equilibration, were considered in this analysis.

as a result of the chemically similar carbohydrate "backbone" of pyranose rings. However, at the +1 site of hyaluronan, the stacking interaction with Trp[99] was not maintained. Instead, prominent hydrogen bonding forces the +1 pyranose ring in an orientation that is perpendicular to aromatic Trp[99] (supplemental Fig. S6). Nevertheless, the similarity in WCA contributions to binding free energy for all three polysaccharides suggests that this +1 site stacking interaction weakly contributes to the overall binding free energy.

The degree to which the binding cleft of YKL-40 was accessible to water molecules did not change significantly with the bound polysaccharide. The degree of ligand solvation was determined by calculating the average number of water molecules within 3.5 Å of a given pyranose ring over the course of the simulations (Fig. 6*b*); error is given as one standard deviation. Chitohexaose and cellohexaose display similar degrees of solvation across the length of the cleft. Hyaluronan allows a moderate increase in degree of solvation of the cleft by comparison to chitohexaose, across the −3 and +1 subsites, where its sharp V shape again contributes to variation in behavior. Given the similarity in solvent accessibility within the binding cleft regardless of ligand, it is unlikely that entropic contributions from solvation play a role in the observed differences in ligand binding free energy.

Protein-Protein Binding in YKL-40—Based on biochemical characterization, it is clear that YKL-40 functionally interacts with collagen. For example, Bigg *et al.* (9) uncovered the ability of YKL-40 to specifically bind types I, II, and III collagen fibers, and Iwata *et al.* (79) recently discovered that YKL-40 secreted by adipose tissue inhibits degradation of type I collagen by matrix metalloproteinase-1 and further stimulates the rate of type I collagen formation. However, a lack of structural evidence has precluded development of an understanding of the molecular nature of these interactions. Using molecular docking, MD simulation, and free energy calculations, we describe interactions of four collagen-like peptides with two putative protein-binding sites along the surface of YKL-40. The selection of model peptides, as well as multiple binding sites, encom-

passes as many potential binding modes as feasible to describe protein-protein binding dynamics and relative affinity of YKL-40 for collagen.

Ligand Binding Dynamics and Comparison of Model Collagen-like Peptides—Dynamics of the collagen-like peptide ligands varies with both binding site and the pitch of the triple helix. Root mean square deviation illustrates the relative stability of each collagen peptide in each of the two binding sites, A and B (supplemental Fig. S7). Although the molecular docking results in very close contact between collagen and YKL-40 (Fig. 4), such that collagen appears to be almost buried in the primary carbohydrate-binding site of YKL-40, minimization and MD simulation results in the slight rise and shift in the position of collagen for every model at binding site A. Each of the four ligands maintains association with the binding site A over the course of 250 ns, although with slightly different protein-protein contacts with YKL-40 (Fig. 7). Native 1CAG, 1BKV, and 1Q7D attained relative stability in a position not significantly different from the initial docked position, but the 1CAG peptide (supplemental Movie S3), with disrupted helical content resulting from the glycine to alanine mutation, required an adjustment in pitch before associating with YKL-40. This relative change in position is shown in the RMSD of the peptides during first 50–100 ns before stabilization (supplemental Fig. S7*a*). Binding site B accommodates helical pitches of 7/2 collagen peptides, because native 1CAG and 1Q7D associated with YKL-40 with very little change in orientation relative to the initial docked positions. The 1CAG ligand was expelled from binding site B, as was the somewhat imperfect 10/3-pitched 1BKV peptide. This suggests that YKL-40 may avoid physiological interactions with certain collagen fibril domains, especially those having imperfect helical pitches. The integrin-binding collagen-like peptide 1Q7D demonstrated the greatest stability among collagen peptides in both binding sites (supplemental Fig. S7) and formed more native contacts with YKL-40 than the other three collagen peptides at binding site A (Fig. 7). We anticipate that the GFOGER motif plays substantial role in mediating the interaction of this collagen peptide with YKL-40.

Examining the number of native contacts between each collagen peptide and binding site A of YKL-40 reveals several common interaction sites mediate collagen binding and helps narrow down key regions of interest (Fig. 7). YKL-40 residues 69–71, 98–108, 205–215, and 230–235 interact with all four collagen peptides and likely contribute to binding affinity, as we will discuss below. The region of YKL-40 between residues 179 and 189 associates with native 1CAG, 1BKV, and 1Q7D, but not with the original 1CAG, as this peptide with relaxed symmetry needed to adjust its position from docked conformation to stabilize the interactions. The 1Q7D model formed the greatest number of interactions with YKL-40 residues relative to the other three models. Similar native contact analysis for binding site B shows that even N-terminal and C-terminal residues of YKL-40 are involved in collagen binding at binding site B (supplemental Fig. S8). It shows that, unlike binding site A, there is little difference in the number of interactions of model 1Q7D and native 1CAG collagen peptide with the binding site B of YKL-40.

To better understand the interactions collagen makes with YKL-40, identified through the native contact analysis, we calculated electrostatic and van der Waals interaction energies of each YKL-40 residue with each collagen peptide over the 250-ns MD simulations (supplemental Table S5). Visual inspection of the simulations reveals aromatic residues in the binding sites, such as Trp[212] and Trp[99] in binding site A and Phe[49] in binding site B, were involved in aromatic-proline stacking interactions with the collagen triple helices. Such interactions are favorable, occurring because of both hydrophobic effects and interaction between the π aromatic face and the polarized C-H bonds (80). This is illustrated in the van der Waals component of the interaction energy, where at binding site A, Trp[69], Trp[71], Trp[99], Trp[212], and Phe[234] show substantial favorable interaction with collagen peptides, although the contribution varies with each collagen peptide (supplemental Table S5). Additionally, acidic and basic residues of the integrin-binding GFOGER motif from collagen-like peptide 1Q7D form ionic interactions with the counter-ionic amino acids of YKL-40, also known as salt bridges. Specifically, 1Q7D forms three salt bridges at binding site A and one salt bridge at binding site B (Fig. 8). At site A, Arg[105], Asp[207], and Arg[263] of YKL-40 interact with Glu(a11), Arg(c12), and Glu(c11) of 1Q7D, respectively, where the a, b, or c in the parentheses corresponds to one of the three strands of the collagen model. Notably, Glu(a11), Arg(c12), and Glu(c11) belong to the GFOGER integrin-binding motif. At site B, Lys[23] of YKL-40 forms a salt bridge with Glu(a11) of 1Q7D. As anticipated, the GFOGER motif played a substantial role in the interaction of this collagen peptide with YKL-40, but its role was different from that of the integrin binding mechanism, which further involves coordination of a metal ion (42). Nevertheless, salt bridges and hydrophobic contacts are very important in both cases, significantly contributing to the electrostatic component of the binding affinity of this collagen peptide relative to collagen peptides lacking acidic or basic amino acids (*e.g.* native 1CAG) (supplemental Table S5). The hydroxyl oxygens of hydroxyprolines from 1CAG and native 1CAG appear to be involved in ionic interactions with acidic YKL-40 residues (favorable electro-
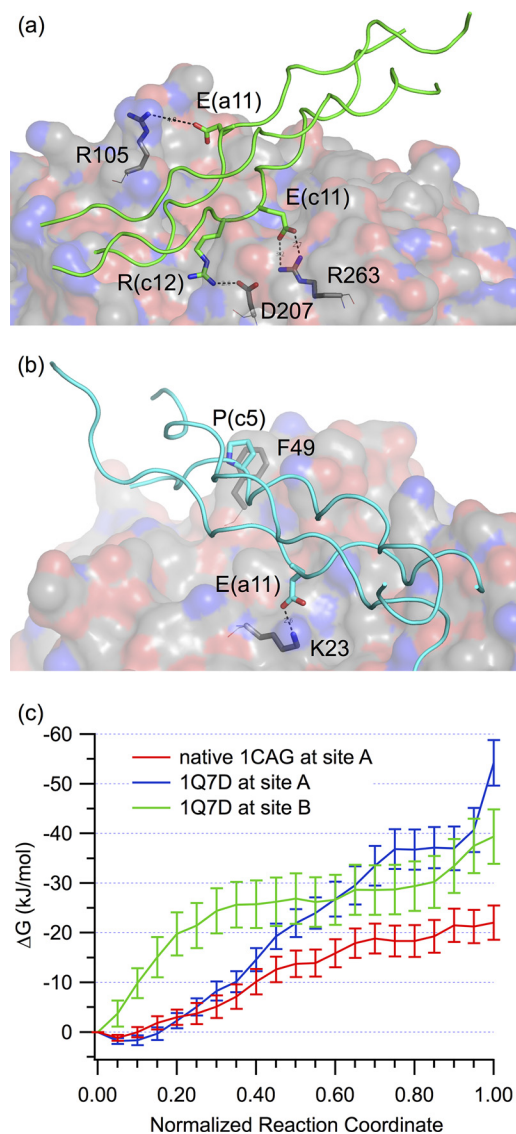
FIGURE 8. **Collagen binding with YKL-40.** *a*, salt bridges formed between the 1Q7D collagen peptide (*green cartoon*) and binding site A (*gray surface*). *b*, salt bridge interactions of the 1Q7D collagen peptide (*cyan cartoon*) with binding site B (*gray surface*). *c*, binding free energy obtained from umbrella sampling MD simulations of the YKL-40-collagen peptide systems, interpreted as negative of PMF to decouple the partners. The collagen peptides in question are 1Q7D (at both sites) and 1CAG (at site A only). The free energy is shown as a function of normalized reaction coordinate, where the reaction coordinate is fraction of native contacts.

static interaction energies; supplemental Table S5), although as a result of hydrogen bonding rather than salt bridge formation. We note that the interaction energies of YKL-40 residues with residues of each collagen model peptide are not conserved as a result of differences in the collagen sequences, particularly in the middle regions consisting of different imino-triplets.

From MD simulation, we observe substantial hydrogen bonding between the collagen peptides and YKL-40 across the length of each binding site, which contributes to overall stability and binding affinity. The hydrogen bonding analysis for the collagen-YKL-40 systems was performed as described above for the polysaccharide ligands; pairs exhibiting greater than 10% occupancy over the simulation are reported individually in supplemental Table S6. The YKL-40 residues responsible for

hydrogen bonding are not consistent across each collagen model (supplemental Table S6). In general, Glu[70], Trp[99], Asn[100], Tyr[141], Arg[145], Ser[179], Lys[182], Thr[184], Asp[207], Arg[213], Phe[218], Asp[232], and Arg[263] in binding site A form hydrogen bonds with the peptides. Similarly at site B, Tyr[22], Lys[23], Asn[87], Asn[89], Lys[91], Lys[377], and Asp[378] are typically involved in hydrogen bonding. The variation in hydrogen bonding pairs between YKL-40 and the collagen peptides is a natural extension of the varying amino acids along the repeating Gly-Xaa-Yaa sequence; there are many different potential donor and acceptor pairs in each case. Hydroxyproline residues play a crucial role both as donor and acceptor in most pairs, benefitting from the extra hydroxyl group relative to proline. Although all the collagen peptides maintain association with collagen-binding site A, the hydrogen bonding characteristics are slightly different for each, which will in turn lead to affinity differences. The relaxed helical pitch of the 1CAG peptide effectively disrupts hydrogen bonding, and affinity for the ligand is lost at binding site B. The 1BKV peptide model with 10/3 symmetry was also unable to remain associated with binding site B, suggesting that binding site B may be more sensitive to helical pitch and prefers 7/2 symmetrical helices.

Collagen-like Peptide Binding Affinity—The relative binding affinity of collagen-like peptides to YKL-40 was determined from umbrella sampling MD simulations. Here, we report the binding affinity of the 1Q7D collagen-like peptide, which is the integrin binding peptide with an overall 7/2 helical pitch (41, 42), at both sites A and B. We have also calculated binding affinity of native 1CAG at binding site A for comparison of binding affinities of two different collagen peptides having different residue substitutions and helical pitches. Unfortunately, in case of umbrella sampling for the native 1CAG peptide at site B, we were unable to obtain statistically reliable results, and thus, we will not discuss findings relative to the affinity of this model.

The umbrella sampling MD simulations of the 1Q7D collagen peptide at both sites A and B show that YKL-40 has similar affinity for 1Q7D at both sites; however, the native 1CAG collagen peptide appears to bind with a lower affinity than 1Q7D at binding site A (Fig. 8c). We note that the last umbrella sampling window in case 1Q7D at binding site A shows a sudden, sharp increase in the potential of mean force (PMF), which is an artifact of the use of native contacts as an umbrella sampling reaction coordinate. As the standard C termini of three strands of collagen helix are negatively charged, they are attracted to the nearby, highly positively charged surface of heparin-binding site. As a result, the final window of the PMF overestimates the work to remove the 1Q7D peptide from binding site A exclusively (Fig. 8). Removing this latter window from the calculation, the free energy of binding 1Q7D is -40.8 ± 4.7 kJ/mol in site A and -42.9 ± 5.0 kJ/mol in site B. The free energy of binding native 1CAG in site A is -22.0 ± 3.4 kJ/mol. The relatively low statistical uncertainty at each window along the potential of mean force suggests sampling of the system was sufficient, providing a meaningful estimate of binding affinity.

The potential of mean force determined from umbrella sampling MD simulations determines the amount of work required to pull the collagen ligand from the binding site. Because free energy is a state function, the difference between the beginning and end state is the binding affinity, regardless of the path taken, as reported above. The path can provide information as to barriers to unbinding; however, the collagen peptides are readily removed from the binding sites along a relatively smooth path. This suggests there is little conformational rearrangement required of YKL-40 in the release of the collagen ligand. The difference between affinity for 1Q7D and native 1CAG at binding site A is reflected in the total hydrogen bonding occupancy in those two cases (supplemental Table S6). Notably, the binding free energies of collagen to YKL-40 are approximately half that of the tighter binding polysaccharide ligands. This suggests that YKL-40 will bind both hyaluronan and chito-oligomers over collagen in the presence of all three. This does not rule out collagen as a physiological ligand but strongly supports hyaluronan as a preferred physiological ligand of YKL-40.

Conclusions—We constructed polysaccharide-bound YKL-40 models and collagen peptide-bound models to understand the molecular level interactions of the protein with potential physiological ligands. MD simulations, as well as free energy calculations, overwhelmingly suggest that polysaccharide ligands, in particular chito-oligomers and hyaluronan, are preferential physiological ligands of YKL-40. However, YKL-40 does appear to bind collagen peptides at two locations along the surface of the molecule. This demonstrates diversity of function of YKL-40 and may potentially be related to the promotion or inhibition of collagen fibril formation, because one could easily imagine that YKL-40 mediates interaction between the triple helix fibers, potentially disrupting the natural matrix. Further structural studies of YKL-40·collagen complexes are certainly warranted and would be invaluable in our understanding of YKL-40 protein-protein interactions.

The apparent difference in binding affinity of YKL-40 for the polysaccharide ligands and collagens is related to the ability of the smaller ligands to penetrate the primary binding cleft. These ligands are able to form longer-lived hydrogen bonds deeper in the hydrophobic interior of the protein. Additionally, electrostatic interactions play a key role in ligand recognition and affinity to YKL-40. Improper alignment of the large, charged side chains of heparan sulfate, heparin, and chondroitin sulfate with the residues lining the YKL-40 cleft prohibits these ligands from binding in the primary binding cleft. However, the smaller, negatively charged side chain of hyaluronan interacts favorably in the primary binding cleft and contributes significantly to the affinity of this molecule. Additionally, we confirmed the non-specific interaction of heparin with the putative heparin-binding domain suggested from previous structural studies. The charged side chains repeatedly and spontaneously interact with charged residues at this secondary surface-binding site.

Overall, our findings indicate that YKL-40 may interact with several ligands *in vivo*, including both polysaccharides and collagen. We suggest that hyaluronan is a preferential physiological ligand of YKL-40, which may explain the pervasive association of YKL-40 with the physical maladies in which hyaluronan has been associated (2, 5, 81, 82). These findings not only identify physiological ligands of YKL-40, they enable future efforts

to rationally guide design of YKL-40 inhibitors, the design of which is invaluable in the control of inflammation-based disorders and possibly several types of cancer.

References

1. Faibish, M., Francescone, R., Bentley, B., Yan, W., and Shao, R. (2011) A YKL-40-neutralizing antibody blocks tumor angiogenesis and progression: a potential therapeutic agent in cancers. *Mol. Cancer Ther.* **10,** 742–751

2. Francescone, R. A., Scully, S., Faibish, M., Taylor, S. L., Oh, D., Moral, L., Yan, W., Bentley, B., and Shao, R. (2011) Role of YKL-40 in the angiogenesis, radioresistance, and progression of glioblastoma. *J. Biol. Chem.* **286,** 15332–15343

3. Ku, B. M., Lee, Y. K., Ryu, J., Jeong, J. Y., Choi, J., Eun, K. M., Shin, H. Y., Kim, D. G., Hwang, E. M., Yoo, J. C., Park, J.-Y., Roh, G. S., Kim, H. J., Cho, G. J., Choi, W. S., *et al.* (2011) CHI3L1 (YKL-40) is expressed in human gliomas and regulates the invasion, growth and survival of glioma cells. *Int. J. Cancer* **128,** 1316–1326

4. Park, J.-A., Drazen, J. M., and Tschumperlin, D. J. (2010) The chitinase-like protein YKL-40 is secreted by airway epithelial cells at base line and in response to compressive mechanical stress. *J. Biol. Chem.* **285,** 29817–29825

5. Johansen, J. S. (2006) Studies on serum YKL-40 as a biomarker in diseases with inflammation, tissue remodelling, fibroses and cancer. *Danish Med. Bull.* **53,** 172–209

6. Mizoguchi, E. (2006) Chitinase 3-like-1 exacerbates intestinal inflammation by enhancing bacterial adhesion and invasion in colonic epithelial cells. *Gastroenterology* **130,** 398–411

7. Nyirkos, P., and Golds, E. E. (1990) Human synovial-cells secrete a 39-kDa protein similar to a bovine mammary protein expressed during the non-lactating period. *Biochem. J.* **269,** 265–268

8. Renkema, G. H., Boot, R. G., Strijland, A., Donker-Koopman, W. E., van den Berg, M., Muijsers, A. O., and Aerts, J. M. (1997) Synthesis, sorting, and processing into distinct isoforms of human macrophage chitotriosidase. *Eur. J. Biochem.* **244,** 279–285

9. Bigg, H. F., Wait, R., Rowan, A. D., and Cawston, T. E. (2006) The mammalian chitinase-like lectin, YKL-40, binds specifically to type I collagen and modulates the rate of type I collagen fibril formation. *J. Biol. Chem.* **281,** 21082–21095

10. Houston, D. R., Recklies, A. D., Krupa, J. C., and van Aalten, D. M. (2003) Structure and ligand-induced conformational change of the 39-kDa glycoprotein from human articular chondrocytes. *J. Biol. Chem.* **278,** 30206–30212

11. 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. *J. Biol. Chem.* **278,** 37753–37760

12. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42,** D490–D495

13. Lis, H., and Sharon, N. (1998) Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98,** 637–674

14. Weis, W. I., and Drickamer, K. (1996) Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65,** 441–473

15. Siaens, R., Eijsink, V. G., Dierckx, R., and Slegers, G. (2004) ^{123}I-Labeled chitinase as specific radioligand for *in vivo* detection of fungal infections in mice. *J. Nucl. Med.* **45,** 1209–1216

16. Drickamer, K., and Taylor, M. E. (1993) Biology of animal lectins. *Annu. Rev. Cell Biol.* **9,** 237–264

17. Lal, A., Lash, A. E., Altschul, S. F., Velculescu, V., Zhang, L., McLendon, R. E., Marra, M. A., Prange, C., Morin, P. J., Polyak, K., Papadopoulos, N., Vogelstein, B., Kinzler, K. W., Strausberg, R. L., and Riggins, G. J. (1999) A public database for gene expression in human cancers. *Cancer Res.* **59,** 5403–5407

18. Lau, S. H., Sham, J. S., Xie, D., Tzang, C. H., Tang, D., Ma, N., Hu, L., Wang, Y., Wen, J. M., Xiao, G., Zhang, W. M., Lau, G. K., Yang, M., and Guan, X. Y. (2006) Clusterin plays an important role in hepatocellular carcinoma metastasis. *Oncogene* **25,** 1242–1250

19. Zaheer-ul-Haq, Dalal, P., Aronson, N. N., Jr., and Madura, J. D. (2007) Family 18 chitolectins: comparison of MGP40 and HUMGP39. *Biochem. Biophys. Res. Commun.* **359,** 221–226

20. 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. *J. Biol. Chem.* **278,** 14451–14460

21. De Ceuninck, F., Pastoureau, P., Bouet, F., Bonnet, J., and Vanhoutte, P. M. (1998) Purification of guinea pig YKL40 and modulation of its secretion by cultured articular chondrocytes. *J. Cell. Biochem.* **69,** 414–424

22. Connor, J. R., Dodds, R. A., Emery, J. G., Kirkpatrick, R. B., Rosenberg, M., and Gowen, M. (2000) Human cartilage glycoprotein 39 (HC gp-39) mRNA expression in adult and fetal chondrocytes, osteoblasts and osteocytes by in-situ hybridization. *Osteoarthritis Cartilage* **8,** 87–95

23. Hakala, B. E., White, C., and Recklies, A. D. (1993) Human cartilage gp-39, a major secretory product of articular chondrocytes and synovial-cells, is a mammalian member of a chitinase protein family. *J. Biol. Chem.* **268,** 25803–25810

24. Harvey, S., Weisman, M., O'Dell, J., Scott, T., Krusemeier, M., Visor, J., and Swindlehurst, C. (1998) Chondrex: new marker of joint disease. *Clin. Chem.* **44,** 509–516

25. Johansen, J. S., Hvolris, J., Hansen, M., Backer, V., Lorenzen, I., and Price, P. A. (1996) Serum YKL-40 levels in health children and adults: comparison with serum and synovial fluid levels of YKL-40 in patients with osteoarthritis or trauma of the knee joint. *Br. J. Rheumatol.* **35,** 553–559

26. Johansen, J. S., Jensen, H. S., and Price, P. A. (1993) A new biochemical marker for joint injury: analysis of YKL-40 in serum and synovial-fluid. *Br. J. Rheumatol.* **32,** 949–955

27. Iozzo, R. V. (1998) Matrix proteoglycans: from molecular design to cellular function. *Annu. Rev. Biochem.* **67,** 609–652

28. Rabenstein, D. L. (2002) Heparin and heparan sulfate: structure and function. *Nat. Prod. Rep.* **19,** 312–331

29. Meyer, M. F., and Kreil, G. (1996) Cells expressing the DG42 gene from early *Xenopus* embryos synthesize hyaluronan. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 4543–4547

30. Semino, C. E., Specht, C. A., Raimondi, A., and Robbins, P. W. (1996) Homologs of the *Xenopus* developmental gene DG42 are present in zebrafish and mouse and are involved in the synthesis of Nod-like chitin oligosaccharides during early embryogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 4548–4553

31. Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3,** 97–130

32. Fraser, J. R., Laurent, T. C., and Laurent, U. B. (1997) Hyaluronan: its nature, distribution, functions and turnover. *J. Intern. Med.* **242,** 27–33

33. Malavaki, C., Mizumoto, S., Karamanos, N., and Sugahara, K. (2008) Recent advances in the structural study of functional chondroitin sulfate and dermatan sulfate in health and disease. *Connect. Tissue Res.* **49,** 133–139

34. Brodsky, B., and Persikov, A. V. (2005) Molecular structure of the collagen triple helix. *Adv. Protein Chem.* **70,** 301–339

35. Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B., and Berman, H. M. (1999) Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* **6,** 454–457

36. Okuyama, K., Miyama, K., Mizuno, K., and Bächinger, H. P. (2012) Crystal structure of (Gly-Pro-Hyp)9: implications for the collagen molecular model. *Biopolymers* **97,** 607–616

37. Rich, A., and Crick, F. H. (1961) The molecular structure of collagen. *J. Mol. Biol.* **3,** 483–506

38. Okuyama, K., Xu, X., Iguchi, M., and Noguchi, K. (2006) Revision of collagen molecular structure. *Biopolymers* **84,** 181–191

39. Shoulders, M. D., and Raines, R. T. (2009) Collagen structure and stability. *Annu. Rev. Biochem.* **78,** 929–958

40. Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) Crystal-structure and molecular-structure of a collagen-like peptide at 1.9-angstrom resolution. *Science* **266,** 75–81

41. Emsley, J., Knight, C. G., Farndale, R. W., and Barnes, M. J. (2004) Structure of the integrin α2β1-binding collagen peptide. *J. Mol. Biol.* **335,** 1019–1028

42. Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J., and Liddington, R. C. (2000) Structural basis of collagen recognition by integrin α2β1. *Cell* **101,** 47–56

43. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., 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., *et al.* (2009) CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30,** 1545–1614

44. Guvench, O., Greene, S. N., Kamath, G., Brady, J. W., Venable, R. M., Pastor, R. W., and Mackerell, A. D., Jr. (2008) Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **29,** 2543–2564

45. Guvench, O., Hatcher, E. R., 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

46. Guvench, O., Mallajosyula, S. S., Raman, E. P., Hatcher, E., Vanommeslaeghe, K., Foster, T. J., Jamison, F. W., 2nd, and Mackerell, A. D., Jr. (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

47. 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.* **34,** 2757–2770

48. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38

49. Duhovny, D., Nussinov, R., and Wolfson, H. J. (2002) Efficient unbound docking of rigid molecules. In Gusfield et al., Ed. Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy. *Lect. Notes Comput. Sc.* **2452,** 185–200, Springer Verlag

50. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33,** W363-W367

51. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802

52. Jiang, W., Hodoscek, M., and Roux, B. (2009) Computation of absolute hydration and binding free energy with free energy perturbation distributed replica-exchange molecular dynamics. *J. Chem. Theory Comput.* **5,** 2583–2588

53. Jiang, W., and Roux, B. (2010) Free energy perturbation hamilton replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *J. Chem. Theory Comput.* **6,** 2559–2565

54. Payne, C. M., Jiang, W., Shirts, M. R., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2013) Glycoside hydrolase processivity Is directly related to oligosaccharide binding free energy. *J. Am. Chem. Soc.* **135,** 18831–18839

55. Beckham, G. T., Matthews, J. F., Peters, B., Bomble, Y. J., Himmel, M. E., and Crowley, M. F. (2011) Molecular-level origins of biomass recalcitrance: decrystallization free energies for four common cellulose polymorphs. *J. Phys. Chem. B* **115,** 4118–4127

56. Payne, C. M., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2011) Decrystallization of oligosaccharides from the cellulose I β surface with molecular simulation. *J. Phys. Chem. Lett.* **2,** 1546–1550

57. Sheinerman, F. B., and Brooks, C. L., 3rd (1998) Calculations on folding of segment B1 of streptococcal protein G. *J. Mol. Biol.* **278,** 439–456

58. Deleted in proof

59. Mulloy, B., Forster, M. J., Jones, C., and Davies, D. B. (1993) N.m.r., and molecular-modelling studies of the solution conformation of heparin. *Biochem. J.* **293,** 849–858

60. Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Ståhlberg, J., and Beckham, G. T. (2015) Fungal cellulases. *Chem. Rev.* **115,** 1308–1448

61. Angyal, S. J. (1969) The composition and conformation of sugars in solution. *Angew. Chem. Int. Ed. Engl.* **8,** 157–166

62. Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* **3,** 853–859

63. Davies, G. J., Planas, A., and Rovira, C. (2012) Conformational analyses of the reaction coordinate of glycosidases. *Acc. Chem. Res.* **45,** 308–316

64. Sinnott, M. L. (1990) Catalytic mechanisms of enzymatic glycosyl transfer. *Chem. Rev.* **90,** 1171–1202

65. Vocadlo, D. J., and Davies, G. J. (2008) Mechanistic insights into glycosidase chemistry. *Curr. Opin. Chem. Biol.* **12,** 539–555

66. Hamre, A. G., Jana, S., Reppert, N. K., Payne, C. M., and Sørlie, M. (2015) Processivity, substrate positioning, and binding: The role of polar residues in a family 18 glycoside hydrolase. *Biochemistry* **54,** 7292–7306

67. Cardin, A. D., and Weintraub, H. J. (1989) Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* **9,** 21–32

68. Hamre, A. G., Jana, S., Holen, M. M., Mathiesen, G., Väljamäe, P., Payne, C. M., and Sørlie, M. (2015) Thermodynamic relationships with processivity in *Serratia marcescens* family 18 chitinases. *J. Phys. Chem. B* **119,** 9601–9613

69. Asensio, J. L., Ardá, A., Cañada, F. J., and Jiménez-Barbero, J. (2013) Carbohydrate-aromatic interactions. *Acc. Chem. Res.* **46,** 946–954

70. Josefsson, A., Adamo, H., Hammarsten, P., Granfors, T., Stattin, P., Egevad, L., Laurent, A. E., Wikström, P., and Bergh, A. (2011) Prostate cancer increases hyaluronan in surrounding nonmalignant stroma, and this response is associated with tumor growth and an unfavorable outcome. *Am. J. Pathol.* **179,** 1961–1968

71. Aruffo, A., Stamenkovic, I., Melnick, M., Underhill, C. B., and Seed, B. (1990) CD44 is the principal cell surface receptor for hyaluronate. *Cell* **61,** 1303–1313

72. Liu, L.-K., and Finzel, B. (2014) High-resolution crystal structures of alternate forms of the human CD44 hyaluronan-binding domain reveal a site for protein interaction. *Acta Crystallogr. F Struct. Biol. Commun.* **70,** 1155–1161

73. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402

74. Toole, B. P. (1990) Hyaluronan and its binding proteins, the hyaladherins. *Curr. Opin Cell Biol.* **2,** 839–844

75. Bleau, G., Massicotte, F., Merlen, Y., and Boisvert, C. (1999) Mammalian chitinase-like proteins. *Exs.* **87,** 211–221

76. Bronowska, A. K. (2011) Thermodynamics of ligand protein interactions implications for molecular design in *Thermodynamics-Interaction Studies: Solids, Liquids and Gases*, pp. 1–48, InTech, Rijeka, Croatia

77. Emsley, J. (1980) Very strong hydrogen bonding. *Chem. Soc. Rev.* **9,** 91–124

78. 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. *J. Biol. Chem.* **276,** 17507–17514

79. Iwata, T., Kuwajima, M., Sukeno, A., Ishimaru, N., Hayashi, Y., Wabitsch, M., Mizusawa, N., Itakura, M., and Yoshimoto, K. (2009) YKL-40 secreted from adipose tissue inhibits degradation of type I collagen. *Biochem. Biophys. Res. Commun.* **388,** 511–516

80. Zondlo, N. J. (2013) Aromatic–proline interactions: electronically tunable CH/π interactions. *Acc. Chem. Res.* **46,** 1039–1049

81. Matou-Nasri, S., Gaffney, J., Kumar, S., and Slevin, M. (2009) Oligosaccharides of hyaluronan induce angiogenesis through distinct CD44 and RHAMM-mediated signalling pathways involving Cdc2 and γ-adducin. *Int. J. Oncol* **35,** 761–773

82. Itano, N., Zhuo, L., and Kimata, K. (2008) Impact of the hyaluronan-rich tumor microenvironment on cancer initiation and progression. *Cancer Sci.* **99,** 1720–1725

83. Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G. D., Roskies, R., Scott, J. R., and Wilkins-Diehr, N. (2014) XSEDE: accelerating scientific discovery. *Comput. Sci. Eng.* **16,** 62–74

Inhibition of Mammalian Glycoprotein YKL-40: IDENTIFICATION OF THE PHYSIOLOGICAL LIGAND

Abhishek A. Kognole and Christina M. Payne

Access the most updated version of this article at doi: 10.1074/jbc.M116.764985

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2017/01/04/M116.764985.DC1

This article cites 81 references, 18 of which can be accessed free at
http://www.jbc.org/content/292/7/2624.full.html#ref-list-1

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 6th 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

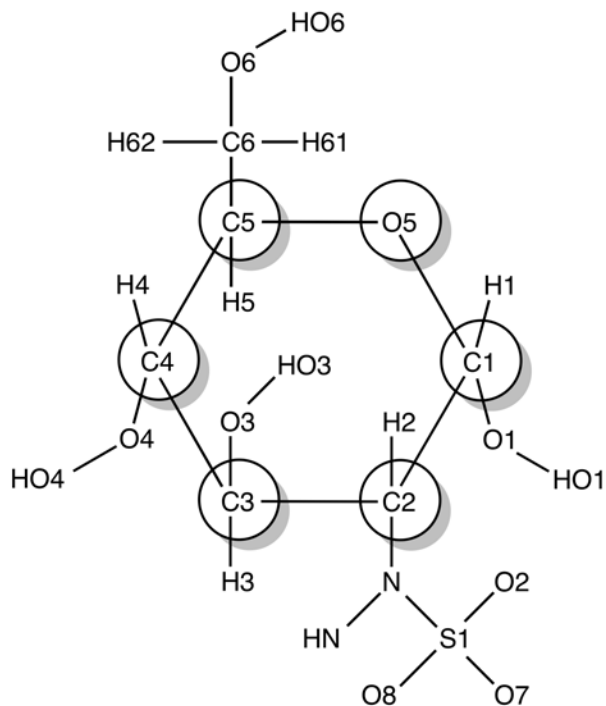 *this O4 is from the glycosidic linkage this residue will be involved in.

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.

$$\text{YKL40}\bullet\text{Ligand}_{(solv)} \xrightarrow{\Delta G_1} \text{YKL40}_{(solv)} + \text{Ligand}_{(vac)}$$

$$\text{Ligand}_{(solv)} \xrightarrow{\Delta G_2} \text{Ligand}_{(vac)}$$

$$\text{YKL40}_{(solv)} + \text{Ligand}_{(solv)} \xrightarrow{} \text{YKL40}\bullet\text{Ligand}_{(solv)}$$
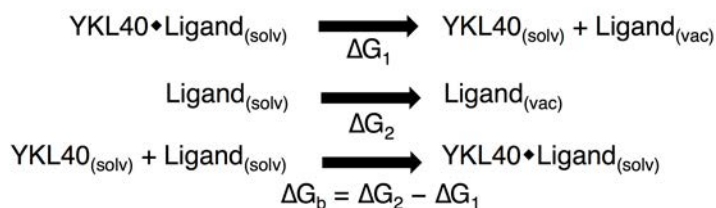
$$\Delta G_b = \Delta G_2 - \Delta G_1$$

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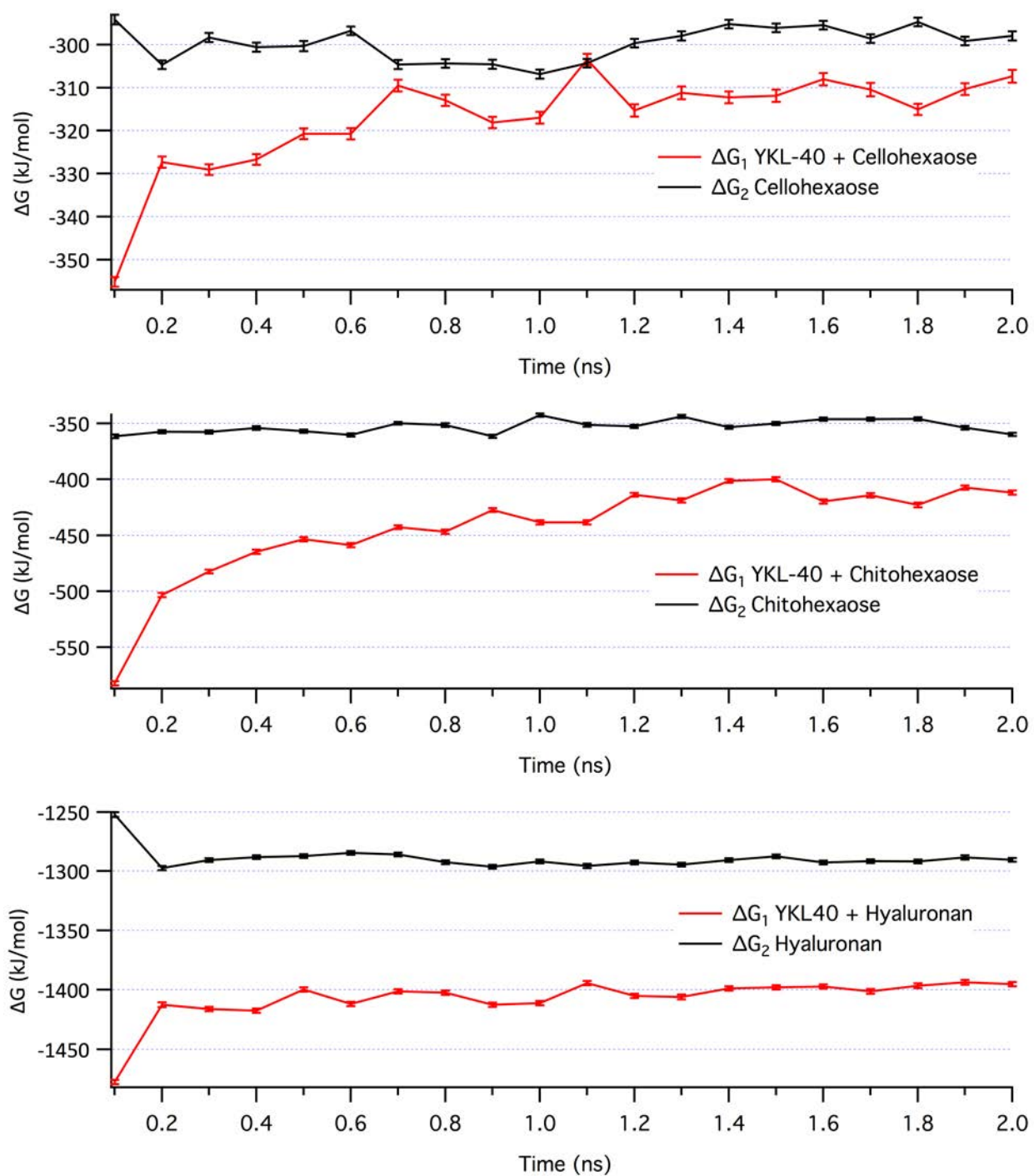
Figure S3. Convergence of ΔG over 20 consecutive 0.1-ns free energy perturbation calculations using the FEP/λ-REMD method.

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).
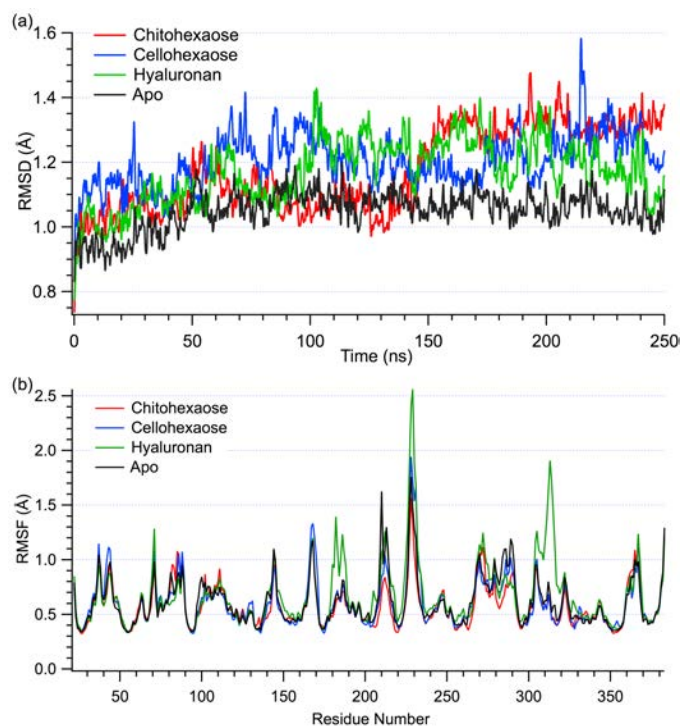

Figure S4. (a) Root-mean-square deviation over 250-ns MD simulations and (b) root-mean-square fluctuation of YKL-40 without a ligand (apo) and bound to chitohexaose, cellohexaose, and hyaluronan. Binding of chitohexaose, cellohexaose, and hyaluronan do not significantly alter the dynamics of YKL-40.

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.
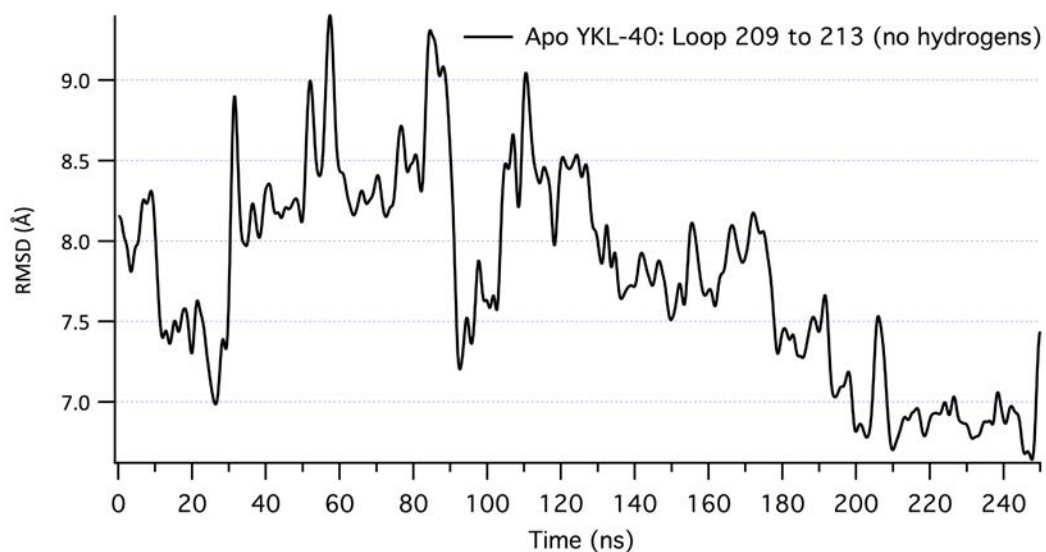
Figure S5. Root mean square deviation of loop of residues 209 to 213 from the unusual configuration in apo YKL-40 crystal structure during 250-ns MD simulation of apo YKL-40 prepared by removing the bound ligand from holo crystal structure.
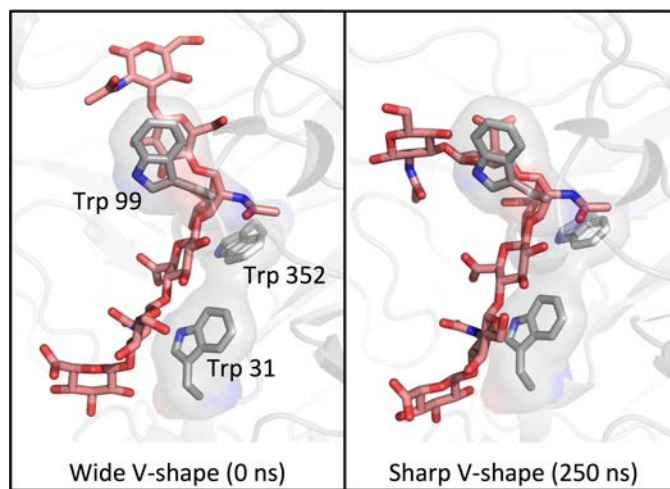
Figure S6. Hyaluronan in YKL-40 binding site at 0 ns (left) and at 250 ns (right) illustrating difference between V-shape conformations of hyaluronan.
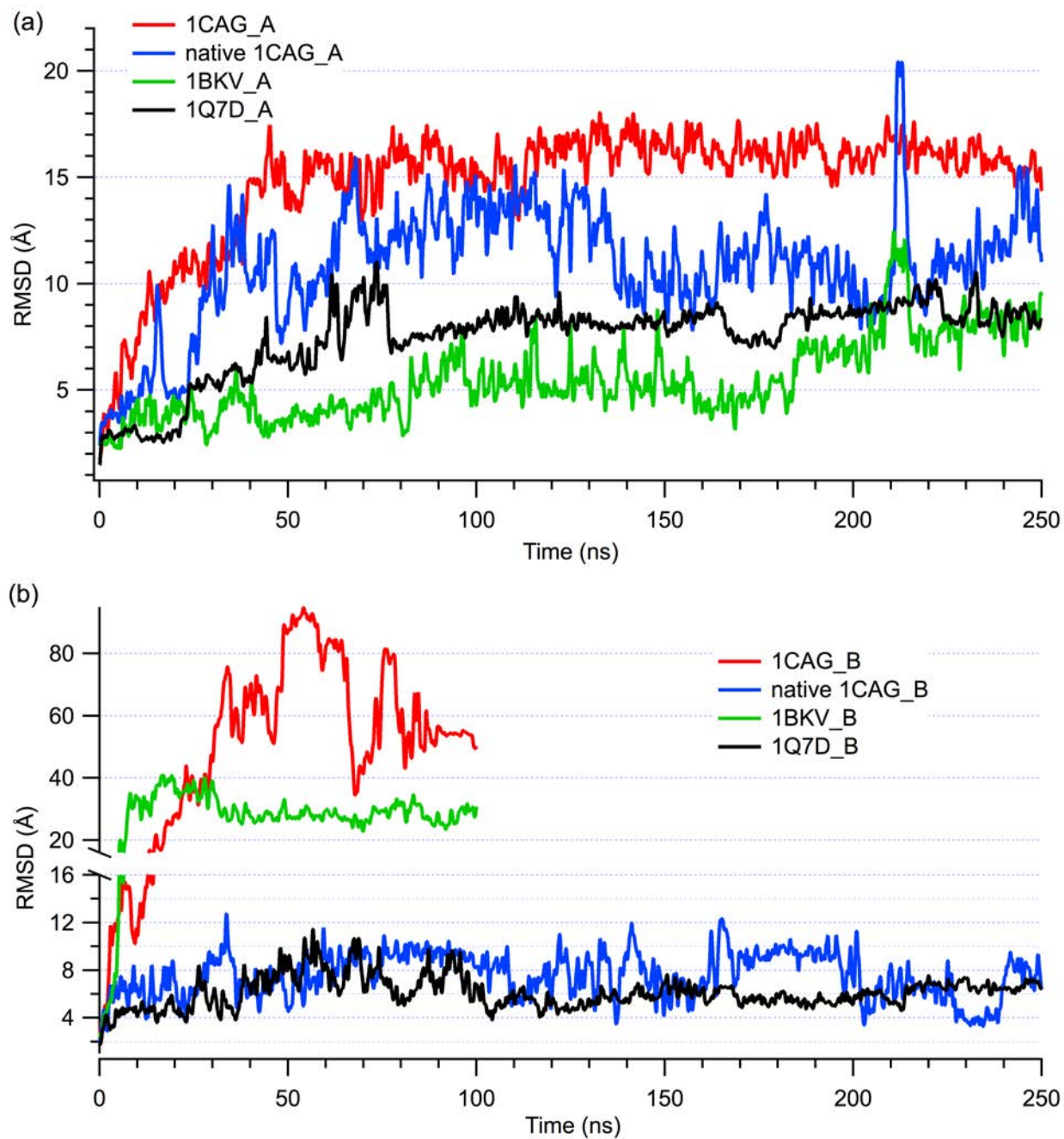
Figure S7. Root-mean-square deviation of collagen-like peptides over the course of 250-ns MD simulations at (a) collagen binding site A and (b) collagen binding site B. Each of the four collagen model peptides are shown.
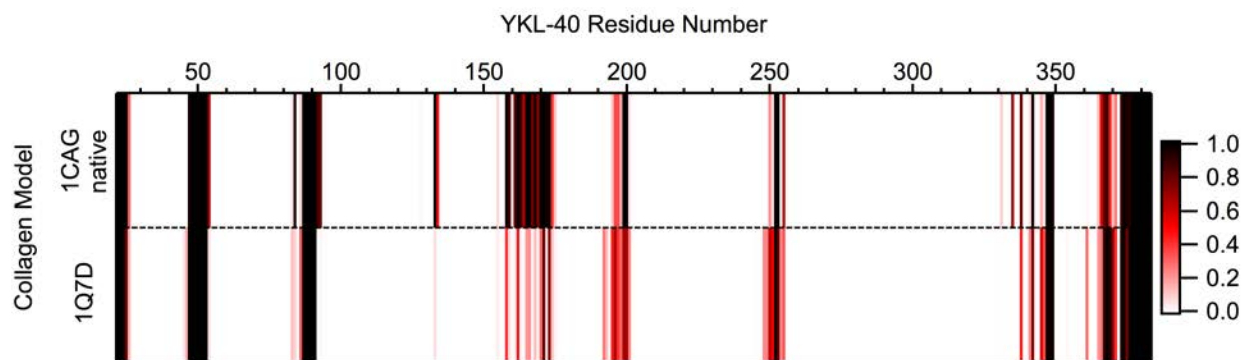
Figure S8. Native contact analysis of each collagen-like peptide model binding to YKL-40 at site B. The color scale represents the normalized frequency (i.e., fractional percentage of frames in which the contact was formed) of the respective YKL-40 residue as a native contact. A native contact was defined 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		Residue #	VdW-Avg	Elec-Avg	Total Avg
1Q7D - binding site A	ARG263	0.330	-23.566	-23.239	**Native 1CAG - binding site A**	ASP232	0.000	-27.814	-27.814
	THR184	-1.129	-16.381	-17.514		TRP99	-13.592	-7.553	-21.144
	LYS182	-0.723	-16.695	-17.418		TRP212	-15.666	0.351	-15.314
	TRP212	-13.667	-1.957	-15.624		VAL183	-9.899	-1.618	-11.517
	ASP207	-0.732	-14.353	-15.084		PHE234	-10.225	-0.084	-10.309
	TYR141	-4.437	-9.677	-14.118		ASN100	-5.290	-2.534	-7.825
	GLU70	0.042	-13.370	-13.328		GLU290	-0.694	-4.943	-5.637
	GLU290	-3.383	-8.109	-11.492		THR184	-2.835	-2.459	-5.294
	ARG145	0.138	-10.292	-10.154		GLN104	-2.196	-2.275	-4.466
	TYR34	-7.783	-0.815	-8.598		TYR141	-3.801	-0.385	-4.186
	ASN100	-6.821	-1.644	-8.464		ASP207	-0.661	-3.220	-3.881
	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		Residue #	VdW-Avg	Elec-Avg	Total Avg
1CAG - binding site A	GLU70	-0.882	-22.512	-23.394	**1BKV - binding site A**	ASP207	-4.057	-59.811	-63.868
	TRP99	-13.081	-4.040	-17.121		PHE208	-5.294	-14.210	-19.501
	GLU290	-2.409	-9.656	-12.061		ALA180	-1.957	-17.447	-19.404
	ASN100	-8.243	-2.530	-10.773		TYR141	-5.445	-12.680	-18.125
	TRP69	-6.357	-3.191	-9.548		TRP99	-13.031	-4.563	-17.594
	TRP71	-8.975	-0.443	-9.418		HIS209	-7.373	-7.419	-14.792
	ALA211	-5.796	-2.802	-8.598		TRP212	-8.247	-4.015	-12.257
	TRP212	-6.620	-1.255	-7.871		LYS182	-1.577	-9.890	-11.467
	ASP207	-0.309	-5.591	-5.901		SER179	-1.752	-9.656	-11.408
	TYR34	-5.474	0.744	-4.730		GLU290	-4.985	-5.805	-10.790
	TRP31	-3.358	-0.506	-3.864		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		Residue #	VdW-Avg	Elec-Avg	Total Avg
1Q7D - binding site B	LYS23	-0.795	-86.229	-87.023	**Native 1CAG - binding site B**	ASN89	-11.965	-26.439	-38.399
	TYR22	-4.788	-46.654	-51.443		LYS377	-7.122	-26.133	-33.255
	LYS91	0.046	-38.608	-38.562		ASP378	-5.119	-15.812	-20.927
	PHE49	-15.130	-1.878	-17.004		ALA381	-9.368	-3.538	-12.906
	ASP367	-0.259	-16.097	-16.360		GLN166	-6.411	-4.199	-10.606
	LYS377	-5.361	-9.142	-14.507		THR52	-8.163	-2.308	-10.472
	THR52	-3.178	-10.120	-13.295		GLN171	-6.382	-3.015	-9.397
	ASP47	-0.949	-8.310	-9.263		PHE49	-7.365	-0.912	-8.280
	LYS253	-1.171	-6.821	-7.992		TYR22	-4.412	-2.639	-7.051
	ASN89	-6.917	-0.719	-7.632		LYS91	-3.726	-2.881	-6.612
	ASP378	-0.941	-5.717	-6.658		LEU50	-2.831	-2.865	-5.696
	ALA381	-5.228	0.322	-4.910		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%	HYP17-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