A coarse-grained model of glycosaminoglycans for biomolecular simulations

Aishwary T. Shivgan[1, 2], Jan K. Marzinek[2], Alexander Krah[2], Paul Matsudaira[1], Chandra S. Verma[1,2,3,*], Peter J. Bond[1,2,*]

1. National University of Singapore, Department of Biological Sciences, 14 Science Drive 4, Singapore 117543
2. Bioinformatics Institute (A*STAR), 30 Biopolis Street, #07-01 Matrix, Singapore 138671
3. School of Biological sciences, Nanyang Technological University, 50 Nanyang Drive, Singapore 637551

[*] Corresponding authors:
Dr Peter J. Bond
Bioinformatics Institute (A*STAR)
30 Biopolis Str.
#07-01 Matrix
Singapore 138671
Email: peterjb@bii.a-star.edu.sg

Dr Chandra Verma
Bioinformatics Institute (A*STAR)
30 Biopolis Str.
#07-01 Matrix
Singapore 138671
Email: chandra@bii.a-star.edu.sg

Abstract

Proteoglycans contain glycosaminoglycans (GAGs), negatively charged linear polymers made of repeating disaccharide units of uronic acid and hexosamine units. They play vital roles in numerous physiological and pathological processes, particularly governing cellular communication and attachment. Depending on their sulphonation state, acetylation, and glycosidic linkages, GAGs belong to different families. The high molecular weight, heterogeneity, and flexibility of GAGs hampers their characterization at atomic resolution, but this may be circumvented via coarse-grained (CG) approaches. In this work, we report a CG model for a library of common GAG types in their isolated or proteoglycan-linked states compatible with the widely popular CG Martini forcefields (versions 2.2 and 3.0). The model reproduces conformational and thermodynamic properties for a wide variety of GAGs, as well as matching structural and binding data for selected proteoglycan test systems. The parameters developed here may thus be employed to study a range of GAG-containing biomolecular systems, benefitting from the efficiency and broad applicability of the Martini framework.

1 Introduction

Proteoglycans (PGs) are heavily glycosylated proteins. These high molecular weight biomolecules found in the extracellular matrix (ECM) play vital roles in governing cellular interactions, the mechanical and viscous properties of tissues, and numerous physiological and pathological processes[1,2]. PGs bind to water, ions, chemokines, growth factors and many other biomolecules[2]. These properties of PGs arise from the covalently attached linear anionic glycosaminoglycans (GAGs), composed of repeating disaccharide units of uronic acid (UA) and hexosamine (HEX) units. The UA units can either be β-D-glucuronic acid (Glc) or α-L-iduronic acid (IdoA), while the hexosamine units can be α-D- or β-D-glucosamine (GlcN) or N-acetyl-β-D-galactosamine (GalNAc). Depending on their sulphonation state, acetylation, and glycosidic linkages, these building blocks make up various GAG families. These encompass non-sulphated GAGs such as hyaluronic acid (HA) and heparin (HP), and sulphated forms including heparin sulphate (HS), chondroitin sulphate (CS), dermatan sulphate (DS) and keratan sulphate (KS) (

Figure 1). These GAGs are typically linked to a tetrasaccharide GlcA(β1→3)Gal(β1→3)Gal(β1→4)Xyl(β-Ser) sequence made of xylose (Xyl), two Galactose (Gal) units and Glucose (Glc), which then connects to the protein via an O-linked serine or threonine[3].

GAGs are implicated in various disease states including inflammatory conditions[4], cancers[5–7], corneal opacification[8,9], arthritis[10] as well as age-related diseases[11]. Furthermore, GAGs in the glycocalyx surrounding cells can serve as attachment factors or receptors for infectious pathogens[12]. For example, host HS may stabilize the so-called ‘up’ conformation of the SARS-CoV-2 spike protein in concert with its own viral N-glycans, thereby facilitating angiotensin-converting enzyme 2 (ACE2) receptor binding and subsequent COVID-19 infection[13]. Unsurprisingly, there is significant therapeutic interest in GAGs, but experimentally studying

their conformation, dynamics, and interactions at the molecular level is hampered by challenges associated with GAG synthesis[14] and further complicated by their polymeric nature and the flexibility of their glycosidic linkages[15]. As an alternative, computational approaches may be utilized to better understand their structure-function relationships. The large size of polymeric GAGs makes quantum mechanics (QM) based methods computationally challenging, but their simulation is more tractable using molecular dynamics (MD) simulations. In MD, the nature and the accuracy of the atomic interactions are defined by bonded and non-bonded interaction potentials collectively termed the force field (FF). Such FFs are typically optimized based on QM and experimental data[14,16]. Many all-atom (AA) FFs such as GLYCAM[17], CHARMM[16] and GROMOS[18] can correctly reproduce the conformational properties of GAG sequences and their protein binding properties[19,20]. AA studies have enabled e.g., improved understanding of anomeric effects or torsional preferences of sulphate groups, and characterization of critical inter- and intramolecular interactions that affect the flexibility and stability of GAGs[20,21]. Despite the high resolution accessible to AA FFs, their application to GAG sequences has typically been limited to, at best, dodecasaccharides, and in many cases, to tetra- or hexa-saccharides, due to the computational cost associated with obtaining sufficient sampling[22]. A solution to this is the development of simplified coarse-grained (CG) models, potentially enabling the study of conformational dynamics of higher-order GAG sequences as observed in nature (i.e. ~20-200 disaccharides long).

In the CG approach, sets of atoms are grouped together and replaced by ‘pseudo-atomic’ particles, resulting in a reduced number of degrees of freedom in the system, thereby allowing access to longer simulation and time- and length-scales at the expense of molecular detail. One of the first CG models for CS and HA was developed by Bath et al., which could predict the influence of pH and ionic strength on persistence length and end-to-end distances,

respectively[23]. The model was characterized by three beads representing a monosaccharide with an additional virtual particle representing its center of geometry and charge. In another CG model, Sattelle et al. implemented ring puckering in HP, CS and DS models to study their 3D shape, bioactivities and hydrodynamic properties[24]. In addition, the chain volume and its rigidity were shown to depend on glucuronic acid, the conformational properties of which could be modulated by the sulphonation state. An AMBER-compatible CG FF developed by Samsonov et al. used unique particle types based on several protein-GAG complexes[25]. In their model, the bonded interactions were optimized based on AA MD simulations, while the non-bonded parameters were optimized based on potential of mean force calculations. The resultant CG model reproduced many local and global properties from AA simulations, such as mean distributions of the root mean square deviation (RMSD), end-to-end distances, and radii of gyration (R_g). In a CG model by Kolesnikov et al., a polymeric field-theoretical approach could reproduce osmotic pressure data and many solution properties in the presence of ions[26].

Although CG models for GAGs alone exist, they are not necessarily compatible with other essential biomolecules such as proteins, lipids, RNA, or DNA, hampering the capacity to study GAGs in realistic, complex biological environments. Martini is currently the most widely used CG FF for biomolecular systems[27]. This FF has been designed to reproduce thermodynamic properties and partitioning data by using a few building blocks that mimic the chemical specificity of the underlying atoms[27]. In the Martini model, four heavy atoms are typically represented as a single CG bead (4:1 mapping). In addition, finer mappings, including 3:1 and 2:1, are utilized for representing rings or high-resolution features. A typical uncharged bead of CG Martini water represents four AA water molecules. A three-bead polarizable water model is also available, two of which are dipoles connected to the central bead[28]. Implicit solvation is also supported, called 'Dry' Martini, and currently is only compatible with lipid systems[29]. The modular building blocks, along with a small number of parameters, result in broad applicability

and transferability to other classes of biomolecules such as proteins, lipids, DNA, and N-glycans[30–34]. Since its introduction, the model has seen wide adoption among the scientific community for studying problems like lipid self-organization, membrane fusion, protein oligomerization, drug delivery, and many more, as summarized elsewhere[35]. The combination of the Martini FF with advances in computational resources and multiscale approaches have made possible simulations of large and complex biomolecular systems such as mitochondria[36], liposomes[37] and virus particles such as dengue[38] and influenza A[39]. Certain shortcomings in version 2.2 of the Martini FF have surfaced in recent years such as overly strong or 'sticky' interactions for proteins[40] and carbohydrates[34]. In an attempt to alleviate these problems, Martini v3.0 was introduced, where rebalancing of the non-bonded interactions was implemented[41], while new bead types and an expanded interaction matrix resulted in a wider chemical space.

In the present work, we have extended the Martini FFs (both v2.2 and v3.0) towards the simulation of various GAG molecules, including HP, HS, CS, DS, and HA, and their O-linked proteoglycan variants. The bonded parameters for GAGs, including linkers, were obtained from extensive AA simulation sampling. Multiple mapping schemes were tested for heparin and the most successful model was employed for other GAGs. Bead types for monosaccharides were validated by calculating partition coefficients and comparing them to predictions from various empirical methods. The resultant CG models were used to predict local and global GAG properties and were applied to some key case studies based on availability of relevant experimental data in the literature. Firstly, HP is essential for homo- and heterodimerization of the fibroblast growth factor receptor (FGFR), and GAG binding modes from simulations were compared with high-resolution structural studies and used to confirm its previously proposed capacity for binding longer HP ligands. Secondly, bikunin is a serine protease inhibitor used to treat acute pancreatitis, and its PG form contains a single O-linked CS chain that is essential

for function; simulations of the PG complex were carried out to characterize the dynamics of the GAG moiety, validated via hydrodynamic size estimates. The CG GAG library developed here represents a computationally efficient strategy for delineating the biological roles of GAG molecules, GAG-protein complexes, and PGs.

2 Methods

2.1 Simulation parameters

In AA simulations, the GROMOS54a7$_{GLYC}$[18] united-atom FF was used for the UA and HEX units involved in various GAG monosaccharides and Xyl-β-Ser O-glycosidic linkages representing the tetrasaccharide GAG linker. The GAG molecules were solvated using the SPC water model[42]. The systems were neutralised using Na^+ and Cl^- counterions. The simulations were performed at a temperature of 310 K using the velocity rescale thermostat[43] with a relaxation time of 0.1 ps. The pressure was maintained at 1 bar using the weak coupling Berendsen barostat with a relaxation time of 1 ps. A cutoff of 1.4 nm was used for electrostatic and van der Waals interactions. Long-range electrostatics were treated using the particle mesh Ewald[44] method with a cutoff of 1.2 nm.

In the case of CG simulations, both Martini v2.2[27,30,45] and v3.0[41] FFs were used. Systems were solvated using Martini water beads, and for systems using v2.2 FF an additional 10% antifreeze water beads were added. A temperature of 310 K and pressure of 1 bar were maintained using the velocity rescale thermostat[46] and Parrinello-Rahman barostat[47] with relaxation constants of 1 ps and 12 ps, respectively. The reaction-field method[48] was used for electrostatics with a cutoff of 1.1 nm. Both AA and CG systems were minimized for 5,000 steps, equilibrated for 50 ns and then subjected to productions runs (summarised in **Table S1Error! Reference source not found.**).

GROMACS 2018[49] was used for all the simulations and run on: i) an in-house Linux cluster using 20 CPUs (E5-2690v3) and 1 GPU (NVIDIA K40) and ii) the ASPIRE 1 supercomputer of the National Supercomputing Centre Singapore (https://www.nscc.sg) using 24 CPUs (E5-2690v3) and 1 GPU (NVIDIA K40).

2.2 Mapping schemes

Two mapping schemes were initially tested for the assessment of the stability of GAG chains (**Figure 2**). Linear ('map1') and ring-like ('map2') mapping schemes were tested for HP. The linear mapping scheme is similar to that previously developed for carbohydrates[30] and N-glycans[34]. The ring-like scheme has previously been used for glycolipids[33]. The underlying chemical nature of the grouped atoms and previously modelled glycans were considered when selecting the bead types. In both schemes, charged functional groups (i.e. sulphates and carboxylates) were treated with a single negatively charged bead. Initial bonded parameters were obtained from tetrasaccharide simulations.

The bonds between beads were fitted using:

$$V_{bond} = \frac{1}{2} k_{bond}(r - r_{bond})^2 \quad (1)$$

where r_{bond} and k_{bond} are the equilibrium distance and the force constant, respectively. The angles within the CG beads were treated using:

$$V_{angle} = \frac{1}{2} k_{angle}\left(\theta - \theta_{angle}\right)^2 \quad (2)$$

where θ_0 is the equilibrium angle and k_{angle} is the force constant. The dihedrals were described using:

$$V_{dihedral} = k_{dihedral}(1 + \cos(n * \theta - \theta_{dihedral})) \quad (3)$$

where $\theta_{dihedral}$ is the equilibrium angle between planes defined by the coordinates of the four consecutively bound atoms *i, j, k* and *j, k, l* respectively, $k_{dihedral}$ is the force constant, and *n* is

the multiplicity. The dihedral distributions had a single minimum and were fitted using a multiplicity of 1.

2.2.1 Validation of non-bonded parameters

The initial bead types were selected by chemical intuition and by analogy with previously parametrized carbohydrate and N-glycan molecules[30,34]. The UA and HEX units had not been parametrized before. The additional sulphate groups make it necessary to validate the choice of the Martini bead types. As Martini parameterization is based on the reproduction of experimental partitioning coefficients ($\log P_{OW}$), these were calculated for various disaccharide combinations commonly found in GAGs. These coefficients were obtained using thermodynamic integration (TI) calculations with 55 windows along with additional windows around the high curvature region. Softcore potentials by Beutler (implemented in Gromacs) were used for sampling the LJ potential at high scaling values[50]. The calculations were performed in 1-octanol and water phases. The difference in solvation free energies of GAG disaccharides in 1-octanol (ΔG_O) and water (ΔG_W) was used to calculate the partition coefficients using $\Delta\Delta G_{OW} = -2.3RT \log P_{OW}$. The water phase was simulated using one GAG disaccharide solvated in a box of 1,000 water molecules, while the organic phase consisted of 43 water and 519 1-octanol molecules, representing a 0.255 water-octanol molar fraction[51]. The coefficients were obtained for both v2.2 and v3.0 versions of the Martini FFs. The obtained simulation data was compared to the results of empirical ($\log P_{OW}$) prediction methods within the ALOGPs 2.1[52] program. The predicted coefficients from multiple methods such as ALOGPs, ALOGP, KOWWIN, ClogP, MLOGP, miLogP, AB/LogP. These methods differ from one another in terms of molecule fragmentation patterns and atomic contributions to the predicted $\log P_{OW}$. The choice of Martini beads for mapping the GAGs (based on the 'map1' scheme) is summarized in **Table 1**.

2.2.2 Protein-GAG studies

The AA crystal structures of FGFR1-FGF2[53] dimer (PDB:1FQ9) and bikunin[54] (PDB:1BIK) were converted to CG representation using the *martinize.py* script available at cgmartini.nl. An elastic network was applied for preserving higher-order protein structure. The network was applied to atoms within the lower and upper cutoff distances of 0.5 and 0.9 nm with a force constant of 500 kJ mol^{-1} nm^{-2}. These values ensured that the folds of each protein complex were retained, with the backbone RMSD of each system with respect to the experimental structure plateauing at ~5 Å by the end of each trajectory (**Figure S1)**.

In the case of a HP20 GAG pre-bound to the FGFR1-FGF2 complex, the GAG was manually aligned to the GAG HP fragment co-crystallized with the X-ray structure (PDB:1FQ9) and simulated for 500 ns using the Martini v2.2 FF with the protein-to-GAG non-bonded LJ interactions scaled by 0.9. This scaling of LJ interactions has previously been shown to be important for correctly representing the solution behaviour of glycans and their interactions with proteins[34]. Besides this, multiple unbiased self-assembly simulations were run, based on two sets of initial configurations of a simulation box containing FGFR1-FGF2 and two randomly placed HP_{20} molecules. Two replicates of each of the configurations were simulated for 1 μs which resulted in 4 independent simulations. In the case of bikunin, the GAG is attached via an O-linked modification at the Ser-10 position[55]. The protein structure of bikunin (PDB:1BIK) has been well studied[54] while the heterogeneous GAG sequences[56] attached to the protein have been characterized and found to range from a length of 27-39 saccharide units[57]. Therefore, either a 30-mer CS (CS_{30}) or a 40-mer CS (CS_{40}) chain was attached to the O-glycosylation site at Ser10[58]. Both systems were solvated in a cubic box and production runs were performed for 5 μs.

Results

2.3 Mapping scheme

Initially, AA tetrasaccharide simulations were used to seed bonded parameter estimates for simulations of longer GAG chains. In the case of HP_{10}, the 'map1' scheme (**Figure 2**) resulted in a very stiff chain with higher end-to-end distances compared to AA data, and reducing the associated dihedral force constants did not lead to significant improvement. On the other hand, the 'map2' scheme (**Figure 2**) resulted in collapse of the polymer into an aggregated state, and consequently, very low end-to-end distances compared to AA data. In the latter case, only extremely strong (10-fold increase) dihedral force constants improved the end-to-end distances (**Figure 3**). Overall, the linear mapping scheme or 'map1' resulted in better correlations with the AA data in these preliminary tests.

To further improve the dynamics of longer GAG chains, backbone angles and dihedrals were next optimized against AA simulations of longer decasaccharide chains, initially focussing on HP_{10} (**Figure 3**). The backbone dihedral angles ϕ and ψ for beads connecting UA-HEX-UA-HEX and HEX-UA-HEX-UA sugars were monitored every GAG chain (**Figure 4** and **Figure 5**). For example, in the case of HP, these angles were defined for G7-G3-G7-G3 and G3-G7-G3-G7 beads (**Figure 4**). This led to an improvement in end-to-end distances and dihedral conformations, and reproduction of the helical nature of the chains (**Movie S1, Figure 5 and Figure 6**), resulting in a much better overall correlation between AA and CG data (**Figure 3**). The 'map1' scheme (**Figure 2**) was eventually selected as the model of choice and as the basis for subsequent parameterization of all other GAG molecules. This was because the 'map2' scheme required four backbone dihedrals to represent each tetrasaccharide unit, compared to two for 'map1', and with higher force constants that often resulted in simulation instabilities arising from constraint errors. Furthermore, visual inspection of simulated chains using the 'map2' scheme revealed unwanted monosaccharide rotations. The final set of refined bonded parameters are detailed in **Table S2**.

2.4 Conformational properties of GAGs

The local conformational properties of CG 10mer GAG sequences using the selected linear 'map1' scheme were generally in good agreement with the AA data (**Figure 5**), as were end-to-end distances (**Figure 6**). In the case of CS and HA, the dihedral distributions were almost identical to their AA counterparts. The CG models of HP, HS and DS resulted in slightly narrower dihedral distributions, while correctly reproducing the AA equilibrium values. As shown for HP (**Figure S2**), lowering the dihedral force constants of the CG model resulted in deviation from the global AA minimum. In the case of KS, the AA distributions exhibited multiple local minima, which is problematic to achieve at CG resolution. Nevertheless, our final CG model enabled two of the three minima to be well sampled during simulations, while accurately reproducing the chain flexibility and end-to-end distances between the central bead of the first and last monosaccharide in KS_{10} (**Figure 6**).

2.5 Hydrodynamic radii

The hydrodynamic radius (R_h) describes the sphere of diffusion volume, helping to characterize polymer dynamics, and is typically similar to the radius of gyration (R_g)[59]. We compared R_g values calculated from CG simulations of S_{100}, DS_{80} and HP_{20} to available R_h values obtained from dynamic light scattering (DLS) experiments[60,61]. The R_g was measured for the last 500 ns of each 1 μs simulation and used to estimate R_h based on the relation $R_g = 1.17 * R_h$ derived for polymeric chains[62]. These computed R_h estimates for both the shorter and longer GAGs were in good agreement with experimental values (**Table 2**), consistent with the prior calibration of each GAG polymer in terms of both local structural/dynamic properties and end-to-end distances of up to 10mer sequences. The slight deviations observed for the longer CS_{100} and DS_{80} molecules may have been a result of simulating the fully sulphated forms, compared to the GAGs derived from natural sources used in experiments, which typically have variable sulphonation states that may affect chain flexibility and dynamics.

2.6 Validation of non-bonded parameters

The newly defined UA and HEX monosaccharides used across the various GAG sequences were validated by calculating partition coefficients ($P_{Oct \rightarrow Wat}$) for both Martini v2.2 and v3.0, based on a series of TI calculations as described in the Methods section. Solvation free energies for each GAG obtained in organic (ΔG_{Oct}) and water (ΔG_{Wat}) phases (**Table 3**) were used to obtain $\log P_{Oct \rightarrow Wat}$ coefficients (**Table S3**). As expected, differences in solvation free energies were observed between the two Martini FFs. All values obtained for both ΔG_{Oct} and ΔG_{Wat} were negative but were consistently greater in magnitude for v2.2 compared to v3.0 of the Martini FF, in agreement with the former being more 'sticky'[41,63]. The negative $\Delta\Delta G_{Oct \rightarrow Wat}$ values obtained for both FFs reflect the preference for GAGs to partition into the water phase, due to the presence of polar hydroxyls along with anionic carboxylate and sulfate groups. Martini FF v3.0 exhibited slightly larger $\Delta\Delta G_{Oct \rightarrow Wat}$ values, indicating a greater tendency to partition into the bulk water phase.

As experimental values are unavailable, likely due to the highly polar nature of GAGs, the partition coefficients calculated from solvation free energies were compared to estimates obtained from empirical prediction methods including atomic (AlogP), hybrid (XlogP3) and fragment-based (ClogP, AC_logP) approaches **(Figure 7 and Table S3Error! Reference source not found.**). The predicted values correlated slightly better in Martini v3.0 than v2.2, wherein the average unsigned errors (AUE) compared to the mean of the empirical predictions were 8.14 and 8.5 kJ mol^{-1}, respectively. The average of the predicted values from all these methods was generally slightly closer to the ones calculated from Martini v3.0 than to v2.2.

2.7 Aggregation studies

The highly polar and often charged nature of GAGs make them readily soluble in water. HP is water soluble with concentrations as high as 50 g L^{-1} [64]. Therefore, we performed aggregation studies of 35 HP tetrasaccharides corresponding to a significantly lower concentration of ~23

312 g L^{-1} (**Figure 8**). In Martini v2.2 simulations, the GAGs aggregated within the first few hundred
313 nanoseconds and remained bound to one another upon extending the simulations to 1 μs. This
314 suggested that the non-bonded interaction potentials are too attractive. Similar behaviour has
315 previously been shown for carbohydrates[65] and N-glycans[34] when using the Martini v2.2 FF.
316 To correct for the imbalance in the glycan-glycan nonbonded interactions, the well depths (ε)
317 of the LJ interaction potentials were scaled according to the following equation:

$$\epsilon_{new} = 2 + \lambda(\epsilon_{original} - 2) \qquad 0.1$$

318 where λ is the scaling factor and $\epsilon_{original}$ is the unscaled well depth of the LJ potential. The
319 same correction factor has been used before for glycan[34,65] and protein[40] parameters. Multiple
320 scaling factors, including 0.95, 0.9, 0.85 and 0.8 were tested. The radial distribution factions
321 $g(r)$ from CG simulations were compared to those obtained from AA simulations using the
322 CHARMM36 FF (**Figure 8**). A drastic contrast in the $g(r)$ values for unscaled Martini v2.2
323 simulations (λ=1.0) and the AA simulations was observed, suggesting excessive aggregation
324 propensity. Scaling the interactions by λ=0.95 resulted in about 50% lower aggregation. Further
325 scaling the interactions reduced the RDFs even more, and no aggregates were observed upon
326 visual inspection of the trajectory in the case of λ=0.9. Although CG simulations could not
327 perfectly reproduce the AA curve using λ=0.9 scaling or below, the differences between CG
328 and AA models were minimal. Therefore, when using Martini v2.2, a scaling factor of 0.9 is
329 recommended for GAG studies, similar to the value recommended for N-glycans[34], where
330 smaller glycans (up to 5 monosaccharides) needed the non-bonded interactions to be scaled by
331 0.9. In contrast, Martini v3.0 showed much lower aggregation propensity, as expected: the
332 resultant RDF was similar to that obtained for Martini v2.2 with a scaling factor of 0.9.

333 **2.8 GAG-protein studies**

334 **2.8.1 Heparin-binding to FGFR1-FGF2 complex**

FGFR1 is a dimeric receptor that is tightly regulated by a unique set of FGFs ligands[66,67], and comprises immunoglobulin-like domains in the extracellular region, a transmembrane helix, and cytoplasmic domains, including a kinase domain[68,69]. HP is essential for the homo- and heterodimerization of the receptor, and hence crucial in the proper functioning of the signalling process[70,71]. GAG binding experiments with FGFR showed dependence on HP chain length. Hexasaccharides and even disaccharides exhibit some binding activity against the receptor[72–74], while progressively increasing affinities have been reported for up to dodecasaccharides [16,75]. Crystallographic studies showed that FGFR GAG binding is facilitated by a positively charged 'canyon' which may be occupied by two HP_{10} chains[53] (**Figure S3Error! Reference source not found.**) with docking studies indicating possible occupation of the entire canyon by longer chains[71]. Therefore, we studied interactions of HP_{20} binding to the FGFR1-FGF2 complex via both biased docking approaches and unbiased self-assembly simulations (see Methods section for further details). In the pre-docked simulations, HP_{20} remained stably bound throughout the simulations (**Figure 9**). The saccharide units within the central cavity of D-II showed minimum fluctuations, as assessed visually, with terminal sugars showing higher flexibility (**Figure 9B**). The homodimeric symmetric interaction pattern of HP_{20} within the positively charged canyon was evident from contact probabilities averaged over the entire trajectory (**Figure 9C**), which was similar to that observed in the crystal structure.

As the CG representation allows access to higher timescales, unbiased self-assembly simulations of the receptor in the presence of two randomly placed HP_{20} ligands were next performed. In 3 out of 4 self-assembly simulations, HP_{20} ligands were able to find the binding site on FGFR1 within a few hundred nanoseconds (**Figures S4A-C**), with some HP_{20} saccharides binding to positively charged residues on FGF2 in the remaining replicate (**Figure S4D**). In one of the productive replicates, an HP_{20} ligand became bound to half of the FGFR1 cavity, similarly to the binding mode observed for HP_{10} in the crystal structure[66], while the

other HP_{20} ligand interacted partially with the remainder of the cavity as well as FGF2 (Error! Reference source not found.**B**). In the remaining productive replicates, HP_{20} sampled the whole of the binding cavity (**Figure S4**Error! Reference source not found.**A, C**), supporting previous hypotheses for the interaction of longer ligands with the FGFR1-FGF2 complex[71].

2.8.2 Bikunin-CS proteoglycan

Bikunin PG is a 16 kDa serine protease inhibitor belonging to the kunin family of inhibitors [76] and contains a single GAG attached via an O-linked modification at the Ser-10 position[55]. The resultant PG is a therapeutically important molecule that has been used for the treatment of acute pancreatitis[77]. The 3D structure of the GAG sequences in either the free[23,61,78–80] or bikunin-attached form[61] have been studied, and found to range from a length of 27 to 39 saccharide units[57] when attached to bikunin. Therefore, either a 30-mer CS (CS_{30}) or a 40-mer CS (CS_{40}) chain was used to generate an O-linked bikunin PG via a tetrasaccharide linker (LIN) (**Figure 4**). Subsequent simulations were used to explore the dynamics of these 'extremes' of CS chain length, validated against hydrodynamic radii estimated from previously performed DLS measurements.

The R_g values were monitored throughout the simulation. We found that starting from an extended conformation, the CS saccharides near the protein start interacting with the protein while the terminal ends point toward the solvent (**Figure 10A, B**). The same was observed while measuring the R_g values which were found to gradually reduce throughout the simulation, stabilizing at around ~3 and ~4 nm for CS_{30} and CS_{40}, respectively. The experimentally estimated R_g values ($R_g = 1.17 * R_h$) lie within the predicted R_g values observed for the two simulation extremes, confirming that the most prominent chain length indeed lies within the previously estimated chain length range of 27-39 CS units[57] (**Figure 10C**).

2.9 Discussion

Structural characterization of GAGs in PGs is hampered by the fact that their observable lengths in crystal structures are typically short (typically only up to 5 to 8 saccharides)[81]. GAG-protein analyses performed in the past have typically focused on shorter GAG sequences[82–86]. Although these results may be extrapolated to longer polymers, the need for methodologies allowing for directly studying longer sequences has motivated the development of CG GAG models. The CG approach allows for the sampling of extended timescales for larger GAG polymers and PG complexes. Previous CG approaches[23–26] have made advancements in understanding the behaviour of longer GAG chains, but lack wider compatibility with other classes of biomolecules.

Herein, we report a CG model for a library of common GAG types compatible with the widely popular CG Martini FFs (v2.2 and v3.0) for biomolecular simulations. CG models for HP, HS, CS, DS, KS and HA were carefully developed after assessing the stability of two alternative mapping schemes. A linear mapping scheme similar to previous ones for carbohydrates[30] and N-glycans[34] successfully reproduced the local and global properties of tetra- and decasaccharide structures according to AA data for all GAGs, with notable deviations only evident in the case of the flexible KS polymer whose multiple local minima in dihedral space were hard to fully replicate at CG resolution. The conformational properties of the GAGs were also further validated against available experimental hydrodynamic data.

The GAG parameters are still potentially limited by the extent to which anomeric linkages are distinguished in the di/tetrasaccharides. However, this effect is minimized with longer chains where the helical nature is dependent upon the backbone dihedrals and angles. In addition to the glycosidic linkages, the conformations of GAG chains are influenced by the monosaccharide ring conformations, which have been shown to affect local and global properties of glycans, including GAGs[25,87–89]. In our AA simulations (**Figure S5**), the UA and

HEX units populated the $^4C_1$ conformation as observed in other FFs as well as NMR studies[87–89]. $^4C_1$ (major) and $^2S_0$ (minor) ring conformers were observed for the GlcA sugar, similar to the lowest energy conformers observed in previously reported MD simulations[88]. The IduA sugar in HP and DS populated the $^1C_4$ conformer, analogous to experimental observations[87]. However, a slight preference towards the ^2S_O (37%) conformer was seen depending on the sulphonation state of IduA and its neighbouring hexosamine sugars. Hence, the AA FF used reproduces the experimentally observed puckering conformations, increasing the confidence in the obtained CG GAG models. Whilst puckering of individual units cannot be reproduced at the CG resolution, the global puckering dynamics should emerge in the angles and dihedrals along the GAG chains.

The non-bonded parameters governing the interactions of GAGs with water, ions and other biomolecules are equally important[4,90–92]. The Martini bead type choices were validated by comparing partition coefficients to empirical fragment-based prediction methods. Both v2.2 and v3.0 resulted in $\log P_{OW}$ values in the range of the predicted data, meaning that either of the FF versions would be suitable for simulations. On the other hand, in the case of systems containing multiple solvated GAG molecules, we found that the GAGs aggregated within a few hundred nanoseconds for v2.2 of the FF. Similar behaviour has been previously reported wherein the v2.2 FF turned out to be 'sticky' especially for carbohydrate containing molecules[33,34,65,93]. Similar to N-glycan[34] and carbohydrate[65] studies, a scaling approach proved beneficial for optimizing the nonbonded interactions. Based on the solution aggregate simulations, a scaling factor of 0.9 could be recommended for Martini v2.2, while no scaling should be necessary with Martini v3.0. Following the corrections to the nonbonded interactions, the model performed very well in predictions of protein-GAG interactions and dynamics. Long HS chains were shown to bind spontaneously to FGFR1-FGF2 at the proposed binding site, whilst O-linked CS chains on the bikunin PG could map out the hydrodynamic

space for the predicted GAG length. This confirms the model's applicability to biologically relevant systems.

3 Conclusions

In summary, we focused on extending v2.2 and v3.0 of the Martini FF to GAGs. The resultant model can accurately predict local and global properties such as end-to-end distances and hydrodynamic radii. Solvation free energies were in good agreement with prediction data, thereby justifying the choice of bead types. In solution, a modest correction factor of 0.9 was best able to reproduce the underlying atomic resolution $g(r)$ for Martini v2.2. Applying the parameters to test systems composed of protein-GAG and PG complexes successfully reproduced a range of structural and biophysical data. The model may thus be employed for studying GAG properties in solution or in complex with proteins, benefitting from the efficiency and broad applicability of the Martini CG framework.

Acknowledgements

The computational work for this article was partially performed on resources of the National Supercomputing Centre, Singapore (https://www.nscc.sg) and BII-ASTAR in house clusters. The research leading to these results was supported by funding from the Ministry of Education in Singapore (MOE AcRF Tier 3 Grant Number MOE2012-T3-1-008) and BII (A*STAR) core funds. PJB acknowledges funding from A*STAR (grant number FY21_CF_HTPO SEED_ID_BII_C211418001). JKM acknowledges AME Young Individual Research Grant (YIRG) number A2084c0160.

Figures & Tables

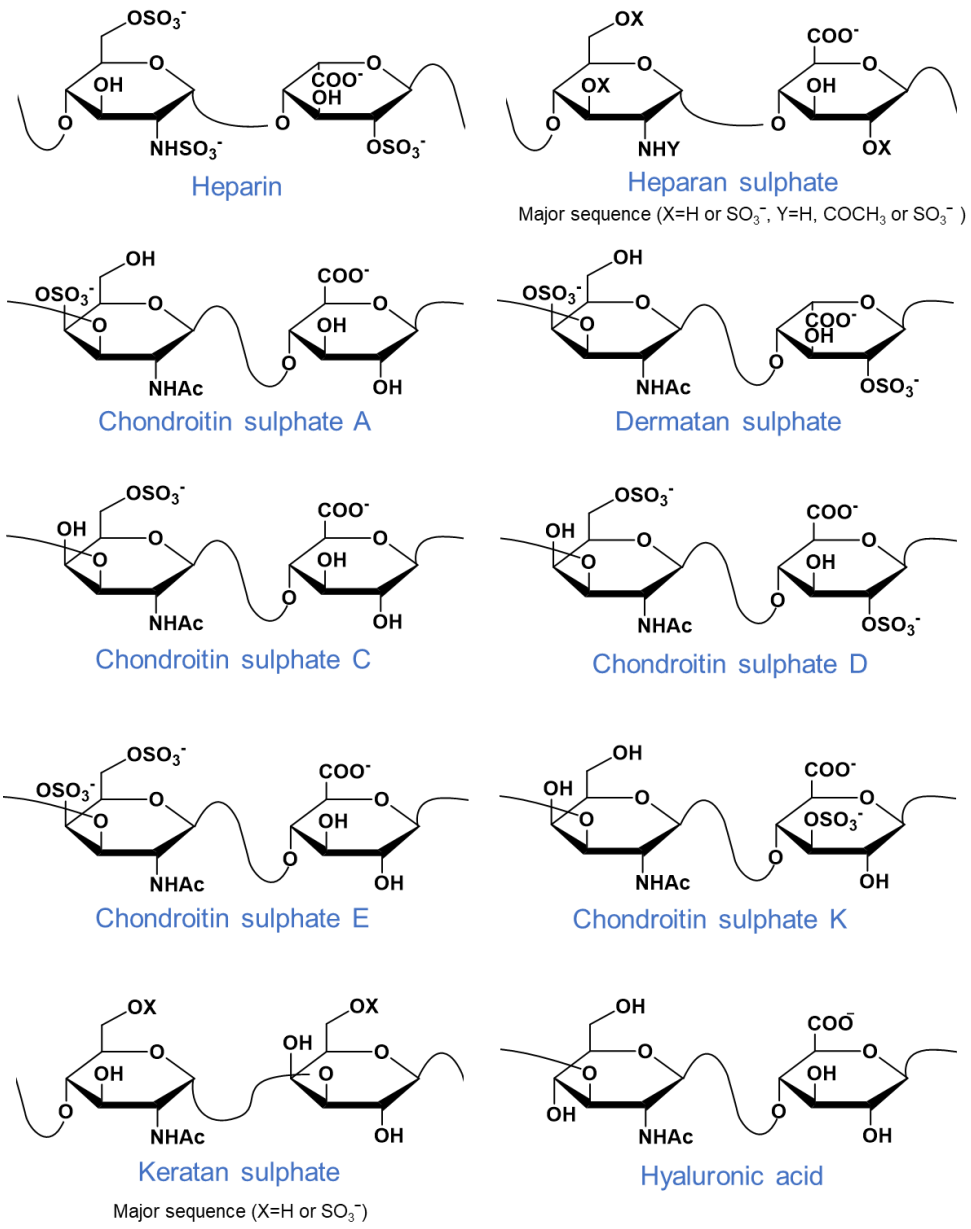

Figure 1. Classification of GAGs. The figure shows the major GAG structures for Heparin [IdoA2S(α1→4)**GlcNS6S(α1→4)IdoA2S(α1→4)**GlcNS6S], Heparan sulfate [GlcA(β1→4)**GlcNAc(α1→4)GlcA(β1→4)**GlcNAc], Chondroitin sulfate [GlcA(β1→3)**GalNAc4S(β1→4)GlcA(β1→3)**GalNAc4S], Dermatan sulfate [IdoA(α1→3)**GalNAc4S(β1→4)IdoA(α1→3)**GalNAc4S], Keratan sulfate [Gal6S(β1→4)**GlcNAc6S(β1→3)Gal6S(β1→4)**GlcNAc6S], Hyaluronic acid [GlcA(β1→3)**GlcNAc(β1→4)GlcA(β1→3)**GlcNAc]

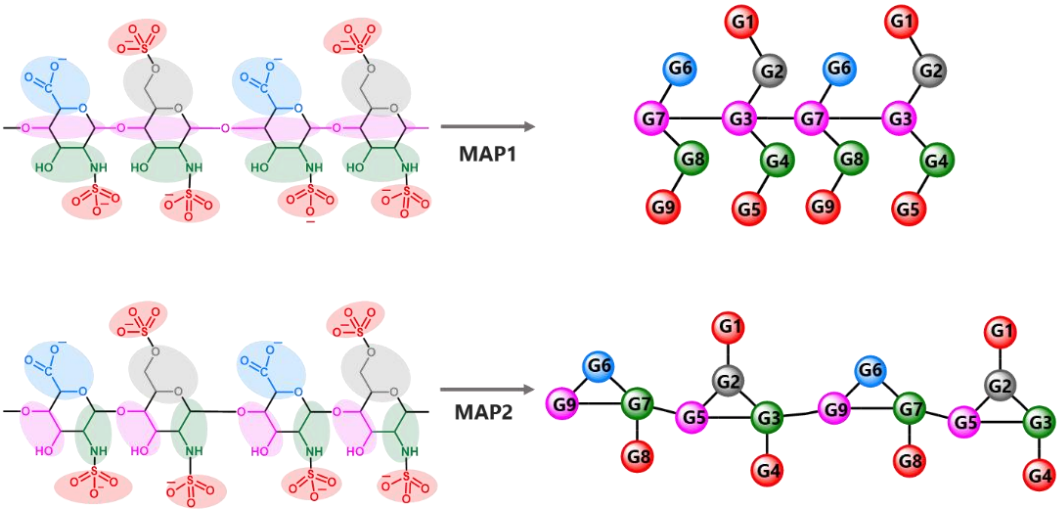

Figure 2. Mapping schemes for heparin. The first scheme makes a linear linkage for the monosaccharides, while the second scheme uses a ring-like topology represented by three beads. Sulphate and carboxylate groups are represented as separate beads in both schemes.

459

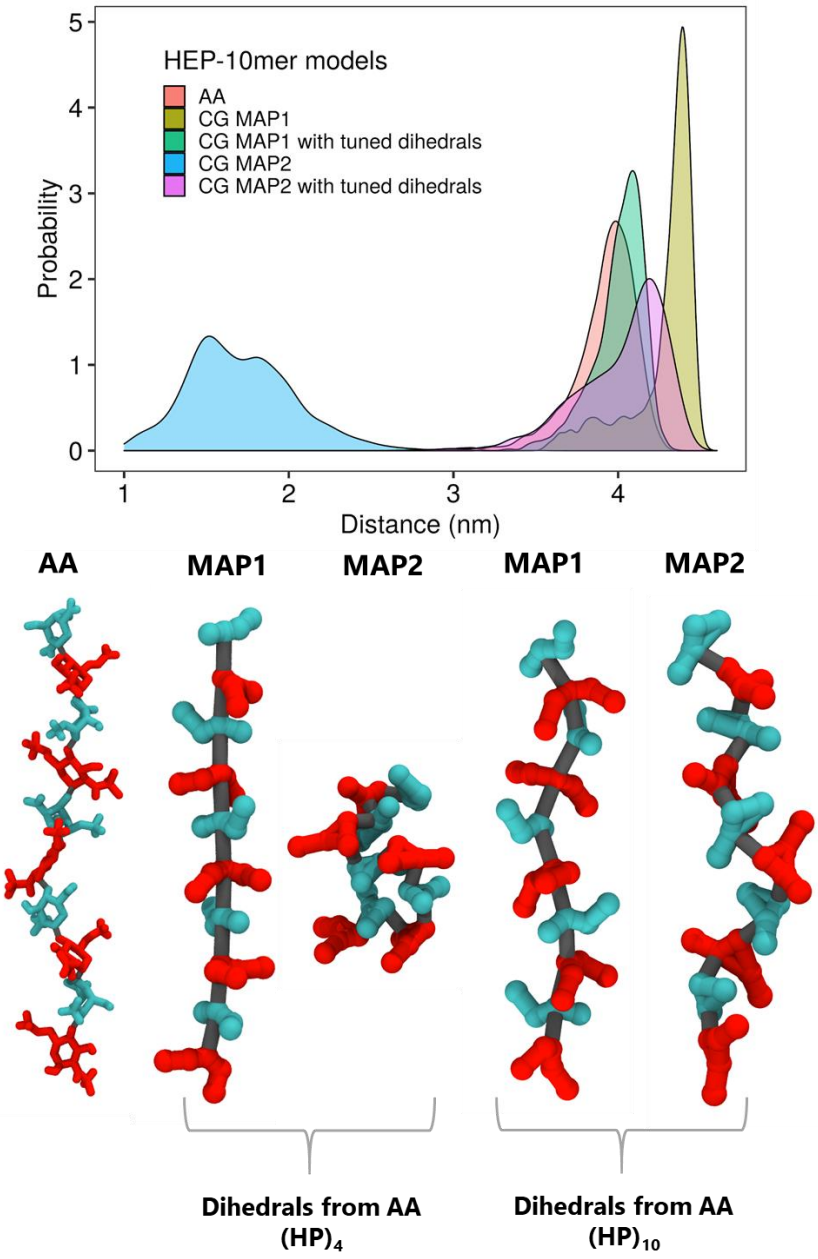

Figure 3. End to end distances of HP_{10} using two mapping schemes. (Top) Distribution of end to end distances. Distances were calculated from beads representing C1 and C4 atoms of the first and last monosaccharides. The AA trajectory was converted to a pseudo-CG trajectory to compare the end-to-end distances from the CG simulations. (Bottom) Final simulation snapshots of HP_{10} models. The hexosamine units are represented in red, while uronic acid units are colored in cyan.

460

461

462

463

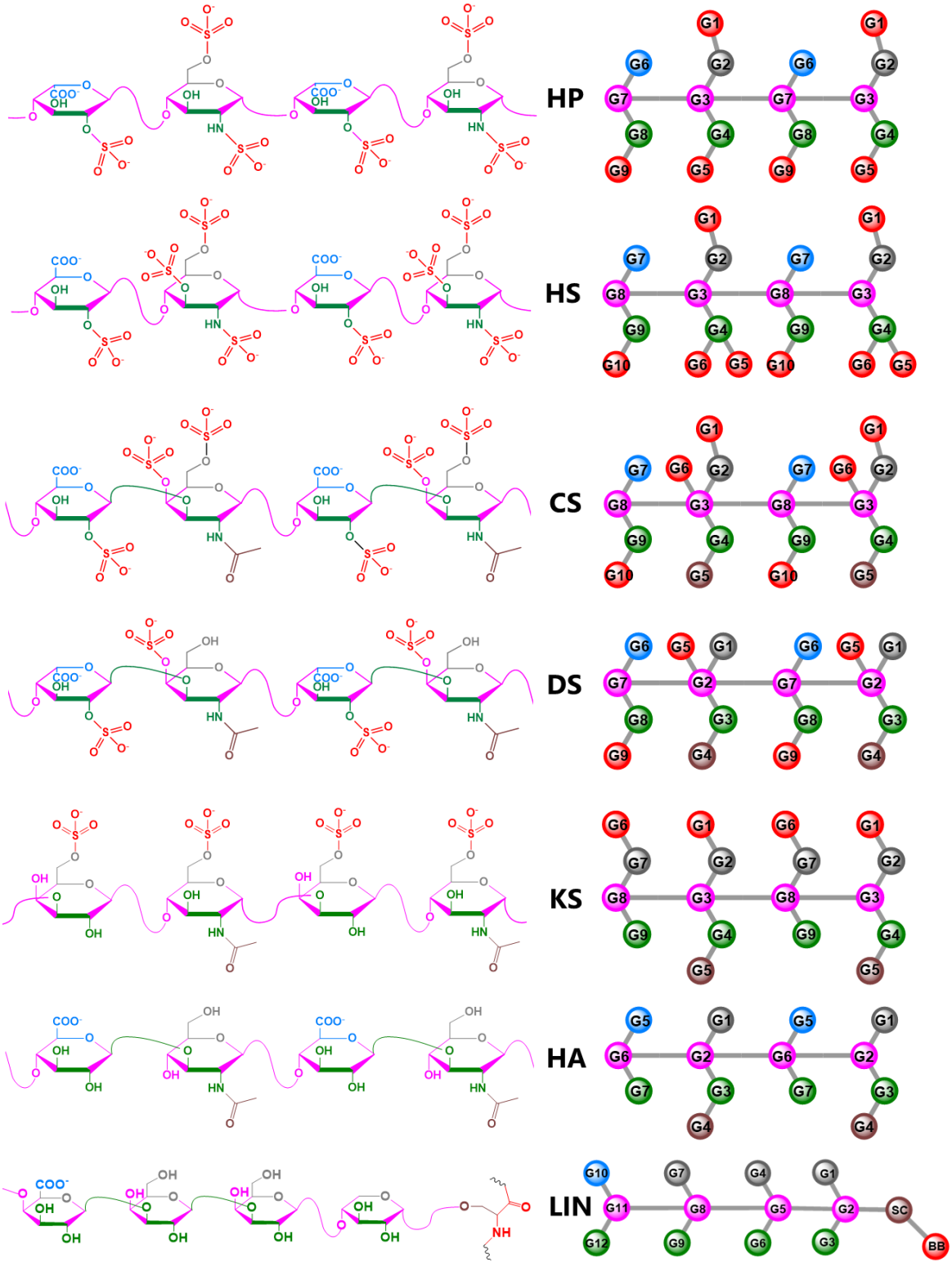

Figure 4. Mapping schemes for all GAG sequences. In addition to the common GAG sequences, the tetrasaccharide linker or **LIN** [GlcA(β1→3)Gal(β1→3)Gal(β1→4)Xyl(β-SER)] for the linkage between GAG and SER was parametrised.

464

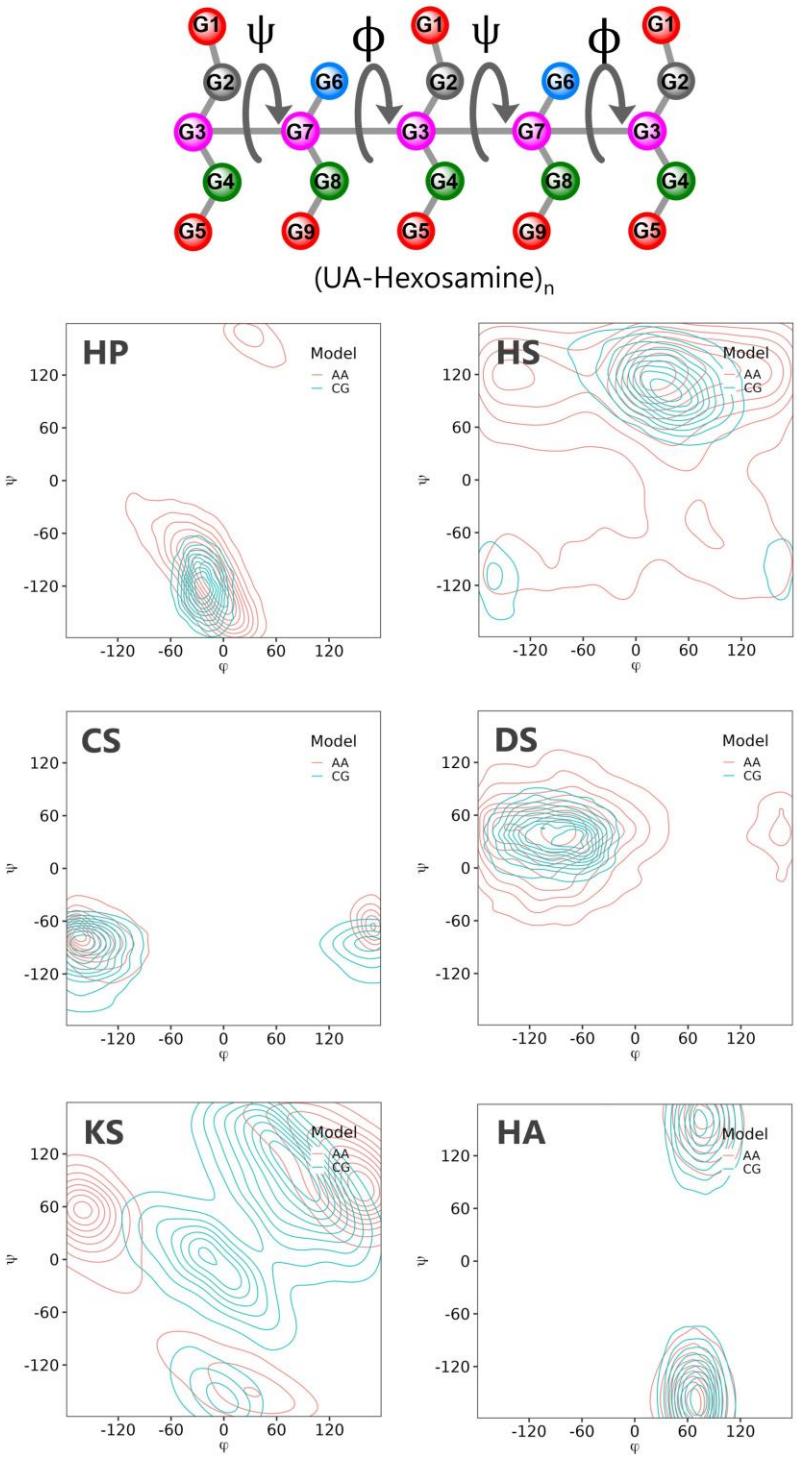

Figure 5. Backbone dihedral distributions of GAGs in atomistic and CG models. CG data was based on Martini v2.2. ϕ and ψ angles were defined as dihedrals between neighboring disaccharide units in the CG model. For example, in the case of heparin, ϕ is defined as the dihedral angle between the G3-G7-G3-G7 beads while ψ is defined as the angle between the G7-G3-G7-G3 beads. The angles were plotted as contour plots for HP, HS, CS, DS, KS and HA. The Martini bead definition is shown in **Table 2**.

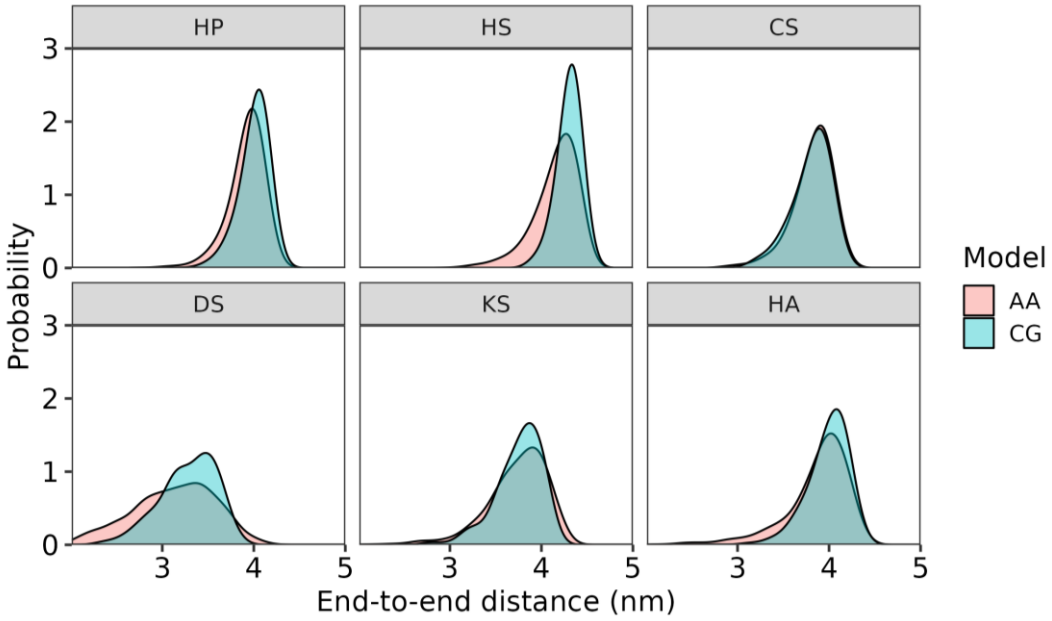

Figure 6. End to end distance distributions for decasaccharide GAG sequences. The distances were calculated from beads representing C1 and C4 atoms of the first and last monosaccharides. AA trajectories were converted to the pseudo-CG trajectories to compare with Martini v2.2 CG simulations.

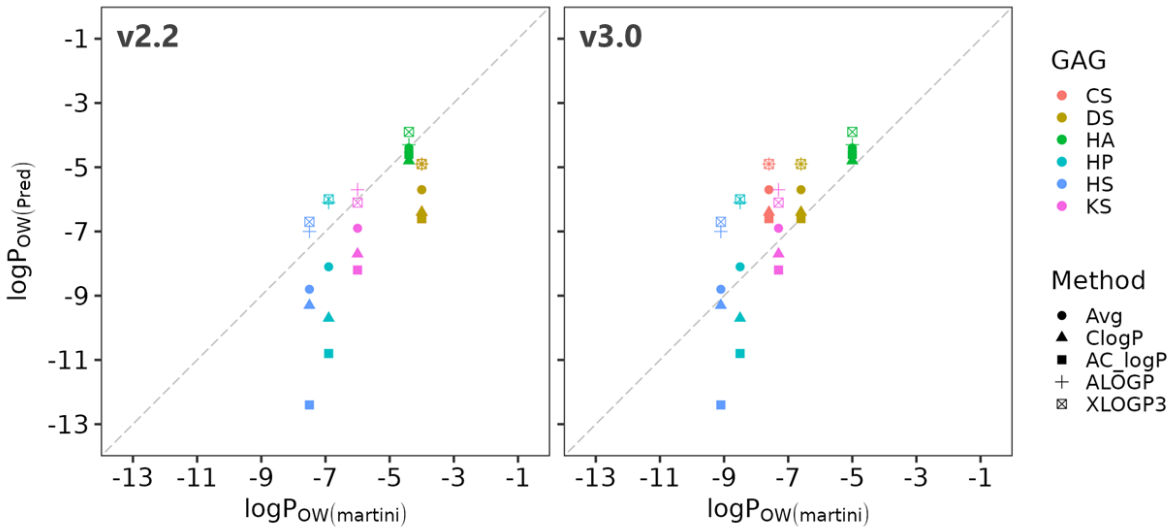

Figure 7. Partition coefficients (log P_{OW}) of different GAG sequences. The log P_{OW} values obtained from simulations were compared to those obtained from empirical prediction methods as labelled. Avg represents the mean of values from all the prediction methods.

471

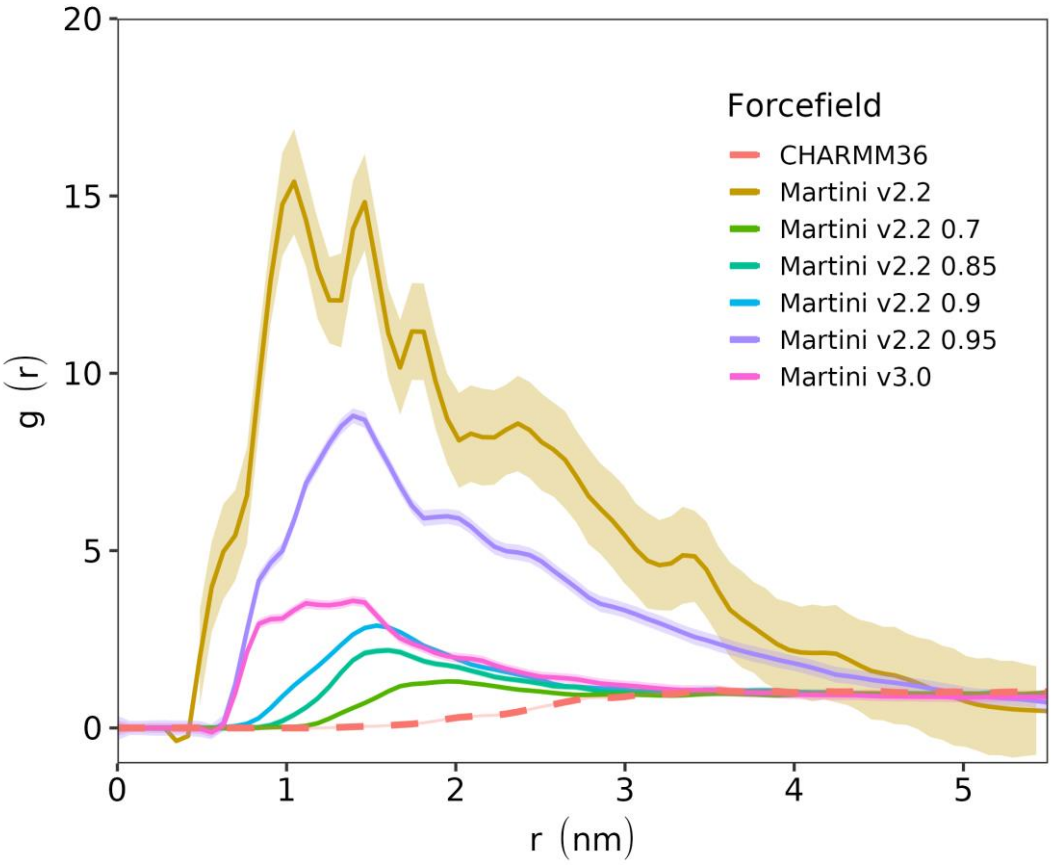

Figure 8. Aggregation propensity of heparin tetrasaccharides for different FFs and LJ scaling factors. The average RDFs for CG and AA models are shown. A range of FFs and LJ scaling factors were tested, as labeled. The RDFs were calculated from the final 200 ns of each trajectory.

472

473

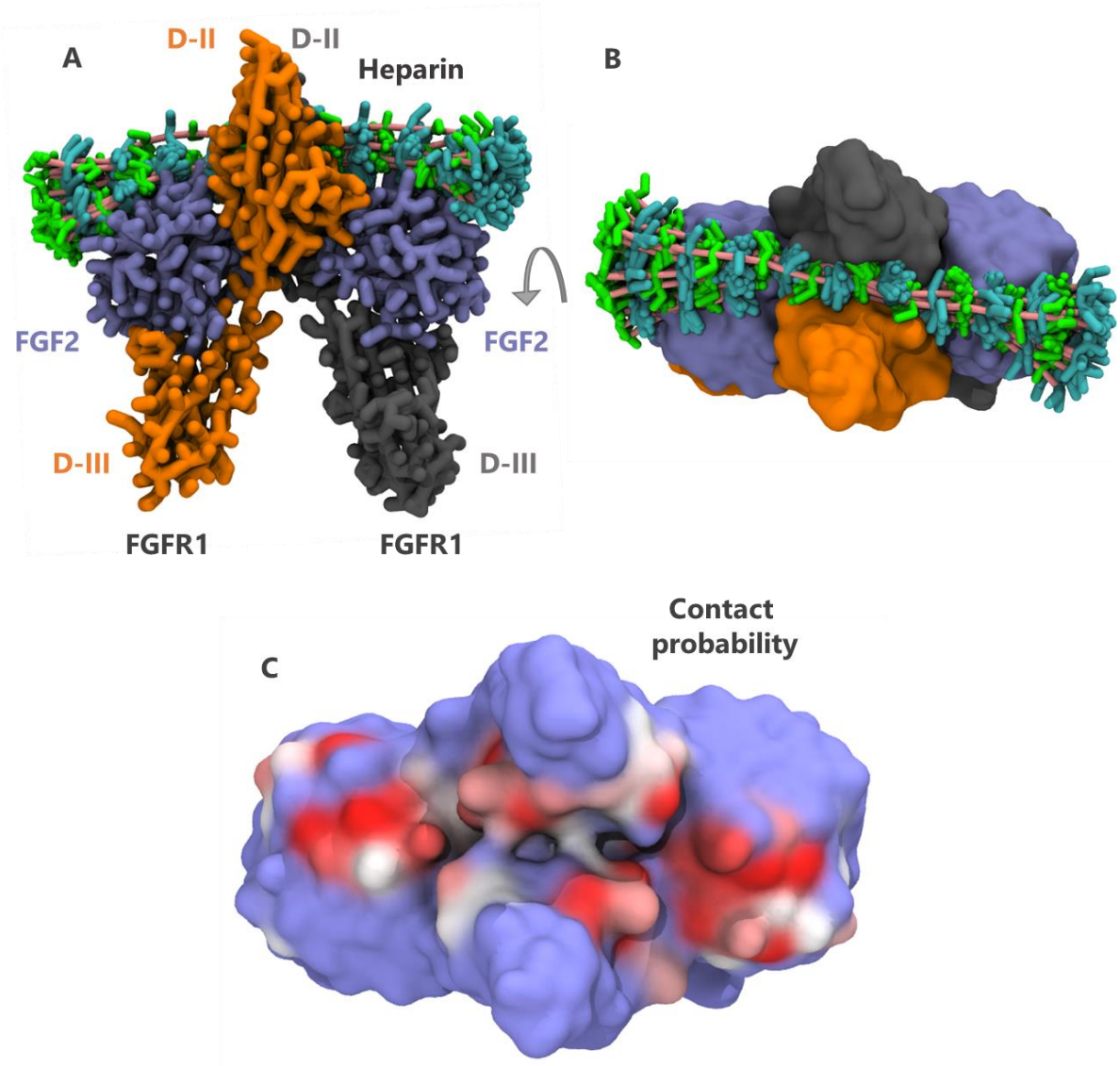

Figure 9. CG simulations of FGFR1-FGF2 complex with docked HP_{20}. An ensemble of conformations of HP_{20} is shown as 20 overlapping and equally spaced frames derived from a 500 ns trajectory, viewed from **(A)** the side, and **(B)** the top of the complex. The GlcNS and IdoA sugars of HP_{20} are shown in licorice representation with cyan and green colors, respectively, and its backbone is colored salmon pink. Protein is colored as in Figure 8A-B. **(C)** Contact probability of HP_{20} to FGFR1-FGF2 complex (shown in surface representation) within a cutoff of 6 Å, colored on a scale from blue (P=0) through grey to red (P=1). Simulations were performed using Martini v2.2 parameters with a LJ scaling factor of 0.9.

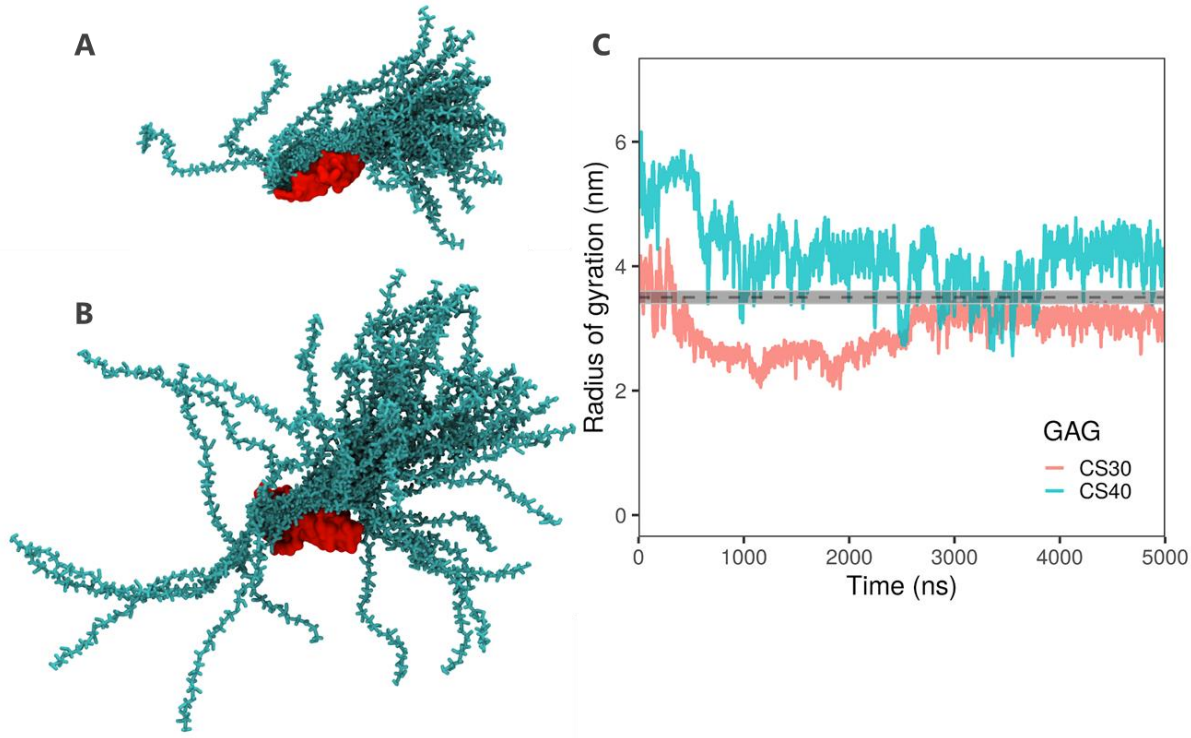

Figure 10. Dynamics of Bikunin PG bound by CS GAG chains. Ensemble of conformations of **(A)** CS_{30} and **(B)** CS_{40} attached to Bikunin. GAG chains are colored in cyan while bikunin is colored in red. Snapshots are overlaid from every 100 ns of the entire trajectory. **(C)** Radius of gyration for bikunin complex formed with CS_{30} (orange) or CS_{40} (cyan). The experimental value of R_g is shown as a dashed line with errors in grey. Simulations were performed using Martini v2.2 parameters with a LJ scaling factor of 0.9.

477

478

Table 1. Martini bead type selection for different CG GAGs as shown in Figure 4. The selections shown are for Martini v3.0. In the case of Martini v2.2, the Q1 bead was replaced with Qa bead. Mapping is based on 'map1' scheme (refer to **Figure 2**).

GAG	Martini bead type											
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
HP	Q1	P1	SP2	P2	Q1	Q1	SP2	P2	Q1			
HS	Q1	P1	SP2	P2	Q1	Q1	Q1	SP2	P2	Q1		
CS	Q1	P1	SP2	P2	SP1	Q1	Q1	SP2	P2	Q1		
DS	P1	SP2	P2	SP2	Q1	Q1	SP2	P2	Q1			
KS	Q1	P1	SP2	P4	SP2	Q1	P1	SP2	P4			
HA	P1	SP2	P2	SP1	Q1	SP2	P4					
LIN	SP1	SP2	P4	P1	SP2	P4	P1	SP2	P4	Q1	SP2	P4

Table 2. Hydrodynamic radii (R_h) of GAG polymers from CG simulation and experiments.

GAG	R_h (sim) / nm	R_h (DLS) / nm
CS_{100}	4.8 ± 0.3	~4.5[61]
DS_{80}	3.8 ± 0.2	~4.2[61]
HP_{20}	1.9 ± 0.03	1.9[60]

488 **Table 3. Solvation free energies calculated for GAG disaccharides with Martini**
489 **FFs v2.2 and v3.0.**

GAG	Martini FF	*ΔG (Water)* kJ mol^{-1}	*ΔG (Octanol)* kJ mol^{-1}	*ΔΔG (Oct→Wat)* kJ mol^{-1}
Heparin (HP)	v3.0	-183.0	-132.4	-50.7
	v2.2	-189.3	-148.1	-41.2
Heparan-Sulphate (HS)	v3.0	-200.6	-146.7	-53.9
	v2.2	-202.0	-157.3	-44.7
Chondroitin-Sulphate (CS)	v3.0	-183.0	-138.2	-44.9
	v2.2	-192.7	-169.1	-23.5
Dermatan-Sulphate (DS)	v3.0	-163.0	-123.6	-39.4
	v2.2	-173.7	-150.1	-23.6
Keratan-Sulphate (KS)	v3.0	-164.7	-121.5	-43.1
	v2.2	-186.1	-150.6	-35.5
Hyaluronan (HA)	v3.0	-120.94	-91.35	-29.6
	v2.2	-144.55	-118.76	-25.8

490

491

References

1. Casu, B. & Lindahl, U. Structure and biological interactions of heparin and heparan sulfate. *Adv Carbohydr Chem Biochem* **57**, 159–206 (2001).

2. Jackson, R. L., Busch, S. J. & Cardin, A. D. Glycosaminoglycans: Molecular properties, protein interactions, and role in physiological processes. *Physiological Reviews* vol. 71 481–539 Preprint at https://doi.org/10.1152/physrev.1991.71.2.481 (1991).

3. Lamberg, S. I. & Stoolmiller, A. C. Glycosaminoglycans. A biochemical and clinical review. *J Invest Dermatol* **63**, 433–449 (1974).

4. Gupta, R. C., Lall, R., Srivastava, A. & Sinha, A. Hyaluronic acid: Molecular mechanisms and therapeutic trajectory. *Front Vet Sci* **6**, 192 (2019).

5. X Z Wu. Sulfated oligosaccharides and tumor: Promoter or inhibitor? *Panminerva Med* **48**, 27–31 (2006).

6. Peraphan Pothacharoen *et al.* Raised Serum Chondroitin Sulfate Epitope Level in Ovarian Epithelial Cancer | The Journal of Biochemistry | Oxford Academic. *The Journal of Biochemistry* **140**, 517–524 (2006).

7. P. Kiewe *et al.* Increased chondroitin sulphate proteoglycan expression (B5 immunoreactivity) in metastases of uveal melanoma - Annals of Oncology. *ORIGINAL ARTICLES MELANOMA* **17**, 1830–1834 (2006).

8. Chakravarti, S. *et al.* Lumican regulates collagen fibril assembly: Skin fragility and corneal opacity in the absence of lumican. *Journal of Cell Biology* **141**, 1277–1286 (1998).

9. Quantock, A. J., Young, R. D. & Akama, T. O. Structural and biochemical aspects of keratan sulphate in the cornea. *Cellular and Molecular Life Sciences* **67**, 891–906 (2010).

10. Hayashi, M., Kadomatsu, K. & Ishiguro, N. Keratan sulfate suppresses cartilage damage and ameliorates inflammation in an experimental mice arthritis model. *Biochem Biophys Res Commun* **401**, 463–468 (2010).

11. Melrose, J. Keratan sulfate (KS)-proteoglycans and neuronal regulation in health and disease: the importance of KS-glycodynamics and interactive capability with neuroregulatory ligands. *J Neurochem* **149**, 170–194 (2019).

12. Cagno, V., Tseligka, E. D., Jones, S. T. & Tapparel, C. Heparan Sulfate Proteoglycans and Viral Attachment: True Receptors or Adaptation Bias? *Viruses 2019, Vol. 11, Page 596* **11**, 596 (2019).

13. Kearns, F. L. *et al.* Spike-heparan sulfate interactions in SARS-CoV-2 infection. *Curr Opin Struct Biol* **76**, 102439 (2022).

14. Mende, M. *et al.* Chemical Synthesis of Glycosaminoglycans. *Chem Rev* **116**, 8193–8255 (2016).

15. Zappe, A., Miller, R. L., Struwe, W. B. & Pagel, K. State-of-the-art glycosaminoglycan characterization. *Mass Spectrom Rev* **41**, 1040–1071 (2022).

16. Guvench, O. *et al.* 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 (2011).

17. Kirschner, K. N. *et al.* GLYCAM06: A generalizable biomolecular force field. carbohydrates. *J Comput Chem* **29**, 622–655 (2008).

18. Pol-Fachin, L., Rusu, V. H., Verli, H. & Lins, R. D. GROMOS 53A6 GLYC, an improved GROMOS force field for hexopyranose-based carbohydrates. *J Chem Theory Comput* **8**, 4681–4690 (2012).

19. Bayraktar, H., Akal, E., Sarper, O. & Varnali, T. Modeling glycosaminoglycans - Hyaluronan, chondroitin, chondroitin sulfate A, chondroitin sulfate C and keratan sulfate. *Journal of Molecular Structure: THEOCHEM* **683**, 121–132 (2004).

20. Verli, H. & Guimarães, J. A. Molecular dynamics simulation of a decasaccharide fragment of heparin in aqueous solution. *Carbohydr Res* **339**, 281–290 (2004).

21. Pol-Fachin, L. & Verli, H. Depiction of the forces participating in the 2-O-sulfo-α-l-iduronic acid conformational preference in heparin sequences in aqueous solutions. *Carbohydr Res* **343**, 1435–1445 (2008).

22. Nagarajan, B., Sankaranarayanan, N. V. & Desai, U. R. Perspective on computational simulations of glycosaminoglycans. *WIREs Computational Molecular Science* **9**, (2019).

23. Bathe, M., Rutledge, G. C., Grodzinsky, A. J. & Tidor, B. A coarse-grained molecular model for glycosaminoglycans: Application to chondroitin, chondroitin sulfate, and hyaluronic acid. *Biophys J* **88**, 3870–3887 (2005).

24. Sattelle, B. M., Shakeri, J. & Almond, A. Does microsecond sugar ring flexing encode 3D-shape and bioactivity in the heparanome? *Biomacromolecules* **14**, 1149–1159 (2013).

25. Samsonov, S. A., Bichmann, L. & Pisabarro, M. T. Coarse-grained model of Glycosaminoglycans. *J Chem Inf Model* **55**, 114–124 (2015).

558 26. Kolesnikov, A. L., Budkov, Y. A. & Nogovitsyn, E. A. Coarse-grained model of
559 glycosaminoglycans in aqueous salt solutions. A field-theoretical approach. *Journal of*
560 *Physical Chemistry B* **118**, 13037–13049 (2014).

561 27. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. The
562 MARTINI force field: Coarse grained model for biomolecular simulations. *Journal of*
563 *Physical Chemistry B* **111**, 7812–7824 (2007).

564 28. Yesylevskyy, S. O., Schäfer, L. V., Sengupta, D. & Marrink, S. J. Polarizable water
565 model for the coarse-grained MARTINI force field. *PLoS Comput Biol* **6**, 1–17 (2010).

566 29. Arnarez, C. *et al.* Dry martini, a coarse-grained force field for lipid membrane
567 simulations with implicit solvent. *J Chem Theory Comput* **11**, 260–275 (2015).

568 30. López, C. A. *et al.* Martini coarse-grained force field: Extension to carbohydrates. *J*
569 *Chem Theory Comput* **5**, 3195–3210 (2009).

570 31. Monticelli, L. *et al.* The MARTINI coarse-grained force field: Extension to proteins. *J*
571 *Chem Theory Comput* **4**, 819–834 (2008).

572 32. Uusitalo, J. J., Ingólfsson, H. I., Akhshi, P., Tieleman, D. P. & Marrink, S. J. Martini
573 Coarse-Grained Force Field: Extension to DNA. *J Chem Theory Comput* **11**, 3932–3945
574 (2015).

575 33. López, C. A., Sovova, Z., Van Eerden, F. J., De Vries, A. H. & Marrink, S. J. Martini
576 force field parameters for glycolipids. *J Chem Theory Comput* **9**, 1694–1708 (2013).

577 34. Shivgan, A. T. *et al.* Extending the Martini Coarse-Grained Force Field to N-Glycans. *J*
578 *Chem Inf Model* **60**, 3864–3883 (2020).

579 35. Marrink, S. J. & Tieleman, D. P. Perspective on the martini model. *Chem Soc Rev* **42**,
580 6801–6822 (2013).

581 36. Pezeshkian, W., König, M., Wassenaar, T. A. & Marrink, S. J. Backmapping
582 triangulated surfaces to coarse-grained membrane models. *Nat Commun* **11**, 2296
583 (2020).

584 37. Risselada, H. J. & Marrink, S. J. Curvature effects on lipid packing and dynamics in
585 liposomes revealed by coarse grained molecular dynamics simulations. *Physical*
586 *Chemistry Chemical Physics* **11**, 2056–2067 (2009).

587 38. Marzinek, J. K., Huber, R. G. & Bond, P. J. Multiscale modelling and simulation of
588 viruses. *Curr Opin Struct Biol* **61**, 146–152 (2020).

589 39. Reddy, T. *et al.* Nothing to Sneeze At: A Dynamic and Integrative Computational Model
590 of an Influenza A Virion. *Structure* **23**, 584–597 (2015).

591 40. Stark, A. C., Andrews, C. T. & Elcock, A. H. Toward optimized potential functions for
592 protein-protein interactions in aqueous solutions: Osmotic second virial coefficient
593 calculations using the MARTINI coarse-grained force field. *J Chem Theory Comput* **9**,
594 4176–4185 (2013).

595 41. Souza, P. C. T. *et al.* Martini 3: a general purpose force field for coarse-grained
596 molecular dynamics. *Nat Methods* **18**, 382–388 (2021).

597 42. Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., H. J. Interaction models for
598 water in relation to protein hydration. in *Intermolecular Forces* 331–342 (Springer,
599 Dordrecht, 1981). doi:10.1007/978-94-015-7658-1_21.

600 43. Hünenberger, P. H. Thermostat algorithms for molecular dynamics simulations.
601 *Advances in Polymer Science* **173**, 105–147 (2005).

602 44. Darden, T., Perera, L., Li, L. & Lee, P. New tricks for modelers from the crystallography
603 toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations.
604 *Structure* **7**, R55-60 (1999).

605 45. De Jong, D. H. *et al.* Improved parameters for the martini coarse-grained protein force
606 field. *J Chem Theory Comput* **9**, 687–697 (2013).

607 46. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling.
608 *Journal of Chemical Physics* **126**, 014101 (2007).

609 47. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and
610 54B7. *European Biophysics Journal* **40**, 843–856 (2011).

611 48. Tironi, I. G., Sperb, R., Smith, P. E. & Van Gunsteren, W. F. A generalized reaction
612 field method for molecular dynamics simulations. *J Chem Phys* **102**, 5451–5459 (1995).

613 49. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *J Comput Chem* **26**, 1701–
614 1718 (2005).

615 50. Beutler, T. C., Mark, A. E., van Schaik, R. C., Gerber, P. R. & van Gunsteren, W. F.
616 Avoiding singularities and numerical instabilities in free energy calculations based on
617 molecular simulations. *Chem Phys Lett* **222**, 529–539 (1994).

618 51. Best, S. A., Merz, K. M. & Reynolds, C. H. Free energy perturbation study of
619 octanol/water partition coefficients: Comparison with continuum GB/SA calculations.
620 *Journal of Physical Chemistry B* **103**, 714–726 (1999).

621 52. Kujawski, J., Bernard, M. K., Janusz, A. & Kuźma, W. Prediction of log P: ALOGPS
622 application in medicinal chemistry education. *J Chem Educ* **89**, 64–67 (2012).

623 53. Plotnikov, A. N., Hubbard, S. R., Schlessinger, J. & Mohammadi, M. Crystal structures
624 of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity.
625 *Cell* **101**, 413–424 (2000).

626 54. Xu, Y., Carr, P. D., Guss, J. M. & Ollis, D. L. The crystal structure of bikunin from the
627 inter-α-inhibitor complex: A serine protease inhibitor with two Kunitz domains. *J Mol*
628 *Biol* **276**, 955–966 (1998).

629 55. Enghild, J. J. *et al.* Organization of the inter-α-inhibitor heavy chains on the chondroitin
630 sulfate originating from Ser10 of bikunin: Posttranslational modification of IαI-derived
631 bikunin. *Biochemistry* **38**, 11804–11813 (1999).

632 56. Ly, M. *et al.* The proteoglycan bikunin has a defined sequence. *Nat Chem Biol* **7**, 827–
633 833 (2011).

634 57. Chi, L. *et al.* Structural analysis of bikunin glycosaminoglycan. *J Am Chem Soc* **130**,
635 2617–2625 (2008).

636 58. Kaczmarczyk, A., Thuveson, M. & Fries, E. Intracellular coupling of the heavy chain of
637 pre-α-inhibitor to chondroitin sulfate. *Journal of Biological Chemistry* **277**, 13578–
638 13582 (2002).

639 59. Gert R. Strobl. *The Physics of Polymers - Concepts for Understanding Their Structures*
640 *and Behavior*. (Springer, 2007).

641 60. Guo, X. *et al.* Determination of molecular weight of heparin by size exclusion
642 chromatography with universal calibration. *Anal Biochem* **312**, 33–39 (2003).

643 61. Sattelle, B. M., Shakeri, J., Cliff, M. J. & Almond, A. Proteoglycans and Their
644 Heterogeneous Glycosaminoglycans at the Atomic Scale. *Biomacromolecules* **16**, 951–
645 961 (2015).

62. Cohen, D. E., Thurston, G. M., Chamberlin, R. A., Benedek, G. B. & Carey, M. C. Laser light scattering evidence for a common wormlike growth structure of mixed micelles in bile salt- and straight-chain detergent- phosphatidylcholine aqueous systems: Relevance to the micellar structure of bile. *Biochemistry* **37**, 14798–14814 (1998).

63. Alessandri, R. *et al.* Pitfalls of the Martini Model. *J Chem Theory Comput* **15**, 5448–5460 (2019).

64. Heparin sodium salt from porcine intestinal mucosa . https://www.sigmaaldrich.com/catalog/product/sial/h4784?lang=en&region=SG.

65. Schmalhorst, P. S., Deluweit, F., Scherrers, R., Heisenberg, C. P. & Sikora, M. Overcoming the Limitations of the MARTINI Force Field in Simulations of Polysaccharides. *J Chem Theory Comput* **13**, 5039–5053 (2017).

66. Schlessinger, J. *et al.* Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol Cell* **6**, 743–750 (2000).

67. Ornitz, D. M. *et al.* Receptor specificity of the fibroblast growth factor family. *Journal of Biological Chemistry* **271**, 15292–15297 (1996).

68. Johnson, D. E. & Williams, L. T. Structural and functional diversity in the fgf receptor multigene family. *Adv Cancer Res* **60**, 1–41 (1992).

69. Jaye, M., Schlessinger, J. & Dionne, C. A. Fibroblast growth factor receptor tyrosine kinases: molecular analysis and signal transduction. *Biochimica et Biophysica Acta (BBA)* **1135**, 185–199 (1992).

70. Stauber, D. J., DiGabriele, A. D. & Hendrickson, W. A. Structural interactions of fibroblast growth factor receptor with its ligands. *Proc Natl Acad Sci U S A* **97**, 49–54 (2000).

669 71. Plotnikov, A. N., Schlessinger, J., Hubbard, S. R. & Mohammadi, M. Structural basis
670 for FGF receptor dimerization and activation. *Cell* **98**, 641–650 (1999).

671 72. Ornitz, D. M. *et al.* FGF binding and FGF receptor activation by synthetic heparan-
672 derived di- and trisaccharides. *Science (1979)* **268**, 432–436 (1995).

673 73. Zhou, F. Y. *et al.* Heparin-dependent fibroblast growth factor activities: Effects of
674 defined heparin oligosaccharides. *Eur J Cell Biol* **73**, 71–80 (1997).

675 74. Gambarini, A. G., Miyamoto, C. A., Lima, G. A., Nader, H. B. & Dietrich, C. P.
676 Mitogenic activity of acidic fibroblast growth factor is enhanced by highly sulfated
677 oligosaccharides derived from heparin and heparan sulfate. *Mol Cell Biochem* **124**, 121–
678 129 (1993).

679 75. Ornitz, D. M. *et al.* Heparin is required for cell-free binding of basic fibroblast growth
680 factor to a soluble receptor and for mitogenesis in whole cells. *Mol Cell Biol* **12**, 240–
681 247 (1992).

682 76. Pugia, M. J., Valdes, R. & Jortani, S. A. Bikunin (Urinary Trypsin Inhibitor): Structure,
683 Biological Relevance, And Measurement. in *Advances in Clinical Chemistry* vol. 44
684 223–245 (2007).

685 77. Michalski, C. *et al.* Preparation and Properties of a Therapeutic Inter-Alpha-Trypsin
686 Inhibitor Concentrate from Human Plasma. *Vox Sang* **67**, 329–336 (1994).

687 78. Sattelle, B. M., Shakeri, J., Roberts, I. S. & Almond, A. A 3D-structural model of
688 unsulfated chondroitin from high-field NMR: 4-sulfation has little effect on backbone
689 conformation. *Carbohydr Res* **345**, 291–302 (2010).

690 79. Almond, A. & Sheehan, J. K. Glycosaminoglycan conformation: do aqueous molecular
691 dynamics simulations agree with x-ray fiber diffraction? *Glycobiology* **10**, 329–338
692 (2000).

693 80. Kaufmann, J., Möhle, K., Hofmann, H. J. & Arnold, K. Molecular dynamics of a
694 tetrasaccharide subunit of chondroitin 4-sulfate in water. *Carbohydr Res* **318**, 1–9
695 (1999).

696 81. Mulloy, B. The specificity of interactions between proteins and sulfated
697 polysaccharides. *An Acad Bras Cienc* **77**, 651–664 (2005).

698 82. Pichert, A. *et al.* Characterization of the interaction of interleukin-8 with hyaluronan,
699 chondroitin sulfate, dermatan sulfate and their sulfated derivatives by spectroscopy and
700 molecular modeling. *Glycobiology* **22**, 134–145 (2012).

701 83. van der Smissen, A. *et al.* Growth promoting substrates for human dermal fibroblasts
702 provided by artificial extracellular matrices composed of collagen I and sulfated
703 glycosaminoglycans. *Biomaterials* **32**, 8938–8946 (2011).

704 84. Hintze, V. *et al.* Sulfated glycosaminoglycans exploit the conformational plasticity of
705 bone morphogenetic protein-2 (BMP-2) and alter the interaction profile with its receptor.
706 *Biomacromolecules* **15**, 3083–3092 (2014).

707 85. Gandhi, N. S., Coombe, D. R. & Mancera, R. L. Platelet endothelial cell adhesion
708 molecule 1 (PECAM-1) and its interactions with glycosaminoglycans: 1. Molecular
709 modeling studies. *Biochemistry* **47**, 4851–4862 (2008).

710 86. Gandhi, N. S. & Mancera, R. L. Free energy calculations of glycosaminoglycan–protein
711 interactions. *Glycobiology* **19**, 1103–1115 (2009).

87. Ferro, D. R. *et al.* Conformer populations of l-iduronic acid residues in glycosaminoglycan sequences. *Carbohydr Res* **195**, 157–167 (1990).

88. Alibay, I. & Bryce, R. A. Ring Puckering Landscapes of Glycosaminoglycan-Related Monosaccharides from Molecular Dynamics Simulations. *J Chem Inf Model* **59**, 4729–4741 (2019).

89. Singh, A. *et al.* Extension and validation of the GLYCAM force field parameters for modeling glycosaminoglycans. *Can J Chem* **94**, 927–935 (2016).

90. Du, H. Y., Ji, S. L., Song, H. F., Ye, Q. N. & Cao, J. C. The relationship between the structure of dermatan sulfate derivatives and their antithrombotic activities. *Thromb Res* **119**, 377–384 (2007).

91. Fransson, L. A., Carlstedt, I., Cöster, L. & Malmström, A. The functions of the heparan sulphate proteoglycans. *Ciba Found Symp* **124**, 125–142 (1986).

92. Roth, M., Papakonstantinou, E. & Karakiulakis, G. Biological Function of Glycosaminoglycans. in *Carbohydrate Chemistry, Biology and Medical Applications* 209–226 (Elsevier Ltd, 2008). doi:10.1016/B978-0-08-054816-6.00009-4.

93. Javanainen, M., Martinez-Seara, H. & Vattulainen, I. Excessive aggregation of membrane proteins in the Martini model. *PLoS One* **12**, e0187936 (2017).