JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

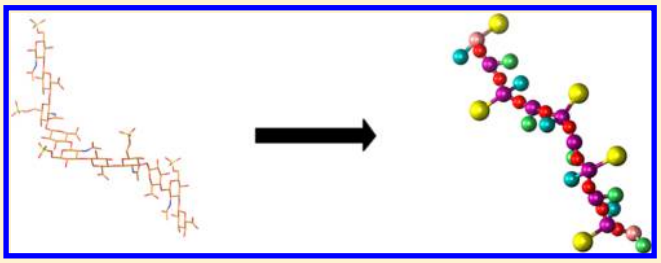
Coarse-Grained Model of Glycosaminoglycans

Sergey A. Samsonov,* Leon Bichmann, and M. Teresa Pisabarro*

Structural Bioinformatics, BIOTEC TU Dresden, Tatzberg 47-51, D-01307 Dresden, Germany

Supporting Information

ABSTRACT: Glycosaminoglycans (GAGs) represent a class of anionic periodic linear polysaccharides, which mediate cell communication processes by interactions with their protein targets in the extracellular matrix. Due to their high flexibility, charged nature, periodicity, and polymeric nature, GAGs are challenging systems for computational approaches. To deal with the length challenge, coarse-grained (CG) modeling could be a promising approach. In this work, we develop AMBER-compatible CG parameters for GAGs using all-atomic (AA) molecular dynamics (MD) simulations in explicit solvent and the Boltzmann conversion approach. We compare both global and local properties of GAGs obtained in the simulations with AA and CG approaches, and we conclude that our CG model is appropriate for the MD approach of long GAG molecules at long time scales.



INTRODUCTION

Glycosaminoglycans (GAGs) are a class of periodic negatively charged linear polysaccharides, which consist of repetitive disaccharide units made up of GlcNAc/GalNAc and GlcU/IdoU monosaccharides bound by 1−3 or 1−4 glycosidic linkages.[1] GAGs can be sulfated in different positions of their disaccharide units, which contributes to their chemical, structural, and functional variety.[2−4] Most GAGs (except for hyaluronic acid) are covalently bound to proteins in the extracellular matrix forming proteoglycans.[5] Long GAG chains of proteoglycans interact with their protein targets such as chemokines and growth factors[6−9] and, therefore, mediate cell signaling processes of anticoagulation, hemostasis, cell proliferation, tumor development, and inflammation.[10,11] Both natural and artificial GAGs are promising targets for design of new biomaterials for the use in the field of regenerative medicine.[12−14] GAG-containing as saccharide ones, are still very challenging from a computational point of view because of their high flexibility,[15] abundance of solvent-mediated interactions,[16,17] and their charged nature, which requires proper treatment of electrostatics.[1,18,19] For heparin, in addition, the rings of IdoU2S can adopt two pucker conformations ($^1C_4$ and ^2S_O) with comparable probability,[20−23] which represents a combinatorial task when modeling long molecules. This aspect is beyond the scope of classical molecular dynamics (MD) time scales and currently used force fields.[24,25] Nonetheless, both experimental and computational studies show that even a single IdoU2S ring conformational change can significantly contribute to the specificity of GAG−protein complex formation both from structural and thermodynamical points of view.[26,27] Most of the GAG−protein complexes for which experimental structures are available typically contain GAG molecules of short length (i.e., 5−8 dp, where dp is the degree of polymerization defined as the number of GAG monosaccharide units).[28] There are several successful studies recently carried out for the analysis of GAG−protein interactions focused on short GAGs,[29−35] whose results could be extrapolated to long polymers. However, for large systems including GAGs such as collagen-based matrices, where interactions with GAGs could be crucial for their mechanical properties,[36−38] there is still a need for development of methodologies, which would allow studying longer GAGs by considering their high flexibility at longer time scales. In these lines, to overcome the listed challenges, an implementation of the coarse-grained (CG) approach seems to be promising. In the CG approach, a decrease in the number of degrees of freedom in the system is achieved by construction of pseudoatoms, which are fewer than the physically existing atoms they represent but are nevertheless able to reproduce their properties manifested in the all-atomic (AA) approach.[39,40] At the moment, a number of knowledge-based and physics-based CG approaches are available.[41] Most of them are constructed to deal with soluble and membrane proteins,[42,43] but some others also deal with nucleic acids,[44] lipids,[45] and carbohydrates.[44,46] Some of these approaches especially focus on the electrostatic interactions, which in general are supposed to be most sophisticated for CG modeling of biomacromolar interactions.[44,47,48] To our knowledge, until now, there is only one GAG-specific CG force field available, which is particularly developed for few GAGs (desulfated chondroitin sulfate, chondroitin sulfates 4 and 6, and hyaluronic acid).[49] In this approach, three physical atoms (oxygen and carbons participating in glycosidic linkages) and two virtual atoms (geometry and charge centers of monosaccharides) have been implemented for modeling bonded and nonbonded interactions, respectively. The only flexible parameters in this model are glycosidic dihedral angles. The authors succeeded to simulate GAGs up to a 500 dp length and obtained a good agreement between the simulations and experimental data in

Received: November 5, 2014
Published: December 9, 2014

terms of conformational sampling, titration procedure, and pH effect. In our work, we aim to develop a different approach to deal with GAGs at the CG level. We construct pseudoatoms, which correspond to the functional groups in GAGs, obtain CG force field parameters using AA simulations and a Boltzmann conversion approach, and apply a simple charge model and steered molecular dynamics for the development of electrostatics and Lennard−Jones parameters, respectively. Then, we introduce parameters compatible with the AMBER MD package to simulate dermatan sulfate, heparin, desulfated heparan sulfate, and differently sulfated variants of hyaluronic acid and chondroitin sulfate. Many of the latter, although being artificial, represent high interest for the development of new biomaterials for regenerative medicine.[13] Analysis and comparison of the global and local properties obtained by AA and CG approaches show that the proposed set of parameters is appropriate for the description of GAGs and can be further used for studies of GAG−protein interactions by MD-based approaches. The results obtained in this study represent a methodologically new computational approach to study GAG-containing molecular systems and contribute to the knowledge of GAG conformational properties.

MATERIALS AND METHODS

PDB-Based GAG−Protein Interactions Analysis. In order to choose pseudoatoms for the CG model of GAGs, we first estimated the impact of different GAG functional groups in GAG−protein interations for a GAG−protein data set containing 11 complexes from the PDB: 1GMN (2.30 Å), 1HM2 (2.00 Å), 1LOH (2.00 Å), 1OFM (1.80 Å), 2D8L (1.70 Å), 2NWG (2.07 Å), 3ANK (2.02 Å), 3E7J (2.10 Å), 3INA (1.90 Å), 3OGX (2.80 Å), and 3OJV (2.60 Å). This data set was selected so that it is representative in terms of GAG types (contains heparin, hyaluronic acid, chondroitin sulfate, and dermatan sulfate) and GAG length range (from 2 to 8 dp) and nonredundant in terms of the protein counterpart in GAG−protein complexes. Water molecules were deleted from the complexes, which were further minimized using default parameters (implicit solvent, distant-dependent solvation model, cutoff of 10 Å, RMS gradient of 0.01 kcal mol^{-1} $Å^{-2}$ as the criteria for convergence) with the AMBER force field in the Molecular Operating Environment (MOE).[50] The "Ligand Interactions" functionality in MOE was used for the calculation of per-atom-contributions to the complex formation in terms of energies corresponding to three types of interactions: H-bond, salt bridge, and hydrogen−π interaction. To analyze the impact of each functional group, the individual impacts of all corresponding atoms from the group were summed up.

GAGs. GAG 10 dp structures were modeled based on the structures available in the PDB for hyaluronic acid (HA, PDB ID: 2BVK, NMR), chondroitin sulfate-4 (CS4, PDB ID: 1CS4, fiber diffraction), and heparin (HE, PDB ID: 1HPN, NMR). For HA- and CS-sulfated derivatives, the sulfate groups were modeled in MOE, where the sulfate charges were obtained from the literature.[52] In total, we modeled and analyzed the structure of the following dp10 GAGs: HA, HA4, HA6, HA46, HA2′, HA3′, HA2′3′, desulfated CS (deCS), CS4, CS6, CS46, CS2′, CS3′, CS2′3′, dermatan sulfate (DS), desulfated heparan sulfate (deHS), HE^{chair}, and HE^{boat}. In addition, HE dp6, dp8, dp12, dp14, dp16, dp18, dp24, dp30, dp32, dp36, dp48, and dp68 were modeled based on the HE dp12 NMR structure (PDB ID: 1HPN). Through the manuscript, sulfated derivatives of HA and CS are abbreviated by addition of a number if sulfation occurs in Glc/GalNAc moiety and by addition of a number with a prime if sulfation occurs in GlcU moiety. For HE, we also distinguish HE^{chair} and HE^{boat} conformations, which correspond to the cases where all IdoU2S rings are in either $^1C_4$ or ^2S_O conformations, respectively.

AA MD Simulations. AA MD simulations of the above listed GAGs were carried out using AMBER 12.[51] GAGs were solvated in a truncated octahedron TIP3P water periodic box with the minimal distance of 15 Å from the solute atoms to the box boundary and neutralized by Na^+ counterions. MD simulations were preceded by two energy minimization steps: 5×10^2 cycles of steepest descent and 10^3 cycles of conjugate gradient with 10 kcal mol^{-1} $Å^{-2}$ harmonic restraints on protein atoms; 3×10^3 cycles of steepest descent and 3×10^3 cycles of conjugate gradient without constraints; 10 ps of heating of the system from 0 to 300 K; and 50 ps of MD equilibration at 300 K and 10^6 Pa in isothermal isobaric ensemble (NPT). The criterion used for the equilibration was the convergence of GAG RMSD to the values similar to the ones in the following MD production run and the convergence of water density. From this step on, the monosaccharide units ring conformations (all rings in $^4C_1$ conformation except for HE Ido2S rings, which were modeled in $^1C_4$ and $^2S_0$ conformations for HE^{chair}, HE^{boat}, respectively) were maintained using 10 kcal mol^{-1} $Å^{-2}$ harmonic NMR restraints. A total of 100 ns of productive MD runs were carried out in periodic boundary conditions in NPT ensemble with Langevin temperature coupling with collision frequency parameter $\gamma = 1$ ps^{-1} and Berendsen pressure coupling with a time constant of 1.0 ps. The SHAKE algorithm, a 2 fs time integration step, an 8 Å cutoff for nonbonded interactions, and the Particle Mesh Ewald method were used. MD trajectories were recorded each 10 ps. ff99SB force field parameters for proteins and GLYCAM06g for GAGs were used. For implicit solvent AA simulations, igb = 5,[53] counterions, and $T = 300$ K were used, while the first step of the minimization described for explicit solvent simulations was skipped.

Development of CG Bonded Parameters. All the parameters for the CG model were obtained compatible to the form of standard AMBER force field parameters (frcmod) file (see http://ambermd.org/formats.html#frcmod for details).

Bonded Parameters: Bonds, Angles, Dihedral Angles. To obtain bonded parameters, the distributions of virtual bonds, angles, and dihedral angles between pseudoatoms were extracted from the AA simulations. Then, a Boltzmann conversion approach[54] was used for the calculation of the equilibrium values and harmonic constants for bonds and angles. In cases in which this approach failed to reproduce the parameters from AA simulations, high artificial constants of 100 kcal mol^{-1} $Å^{-2}$ and 500 kcal mol^{-1} rad^{-2} for bonds and angles were used, respectively. In case of dihedral angles, AMBER employs the parameters of the periodic torsional potential (phase, amplitude, periodicity). Because of the absence of symmetry for the chosen pseudoatoms, the periodicity parameter was set to 1, which corresponded to the only one maximum/minimum of the potential per 360°. The position of the potential maximum (phase) + 180° corresponded to the energy minimum, calculated by a Boltzmann conversion approach. The amplitude was calculated from the obtained one-dimensional energetic landscape as the difference between the global minimum and the highest energetical barrier between the global and local minima.

Nonbonded Parameters: Charges, Lennard−Jones Parameters. The integer point charges were empirically assigned to the pseudoatoms. Pseudoatoms corresponding to the sulfate and carboxyl groups were assigned to −1 and the rest to 0.

dx.doi.org/10.1021/ci500669w | J. Chem. Inf. Model. 2015, 55, 114−124

To obtain Lennard−Jones parameters, we used a protocol for potential of mean force (PMF), similar to the one previously described in the work of Gu et al.[43] First, AA groups corresponding to the pseudoatoms were capped with a hydrogen, parametrized with the gaff force field using the antechamber module of AMBER, and all the partial atomic charges were assigned to 0 in order to entirely ignore electrostatic interactions in the system. Corresponding AMBER residue libraries were created. Then, two molecules representing the same Lennard−Jones probe were initially put within the nonbonded interaction cutoff and minimized *in vacuo* (1.5×10^3 cycles of steepest descent and 10^3 cycles of conjugate gradient without constraints). The distance between the centers of mass of the two molecules was set up to be 0.5 Å lower than in the energy minimum state, and a steered molecular dynamics approach with a force constant of 100 kcal mol^{-1} $Å^{-2}$ was applied to move the probes apart up to 15.0 Å distance in 1000 steps. The dependence of the van der Waals energy on the distance between centers of mass of two probes was obtained and corrected for the van der Waals energy at a distance of 15.0 Å, at which there were no more intermolecular interactions. Analysis of the obtained curves carried out by their approximation with a 6−12 potential yielded Lennard−Jones parameters: energy well depth and van der Waals radius. Application of this approach for single atoms succeeded to quantitatively reproduce ff99SB parameters (data not shown).

Coarse-Grained MD Simulations. Prior to CG simulations, in-house scripts were used in order to calculate the initial coordinates of the pseudoatoms corresponding to the GAG structures used for AA simulations and to rename them accordingly to CG residues nomenclature. For each pseudoatom, the coordinates were calculated based on the mean coordinates of all corresponding non-hydrogen atoms from the AA representation without taking into account their atomic weights upon the averaging. The PDB file containing pseudoatoms was then read by the Leap module of AMBER, and pseudoatom types, charges, and topology were assigned based on the data contained in an AMBER-compatible residue library. A total of 100 ns of MD simulations in explicit and implicit solvent were carried out using the same procedures as for the AA case (as described above) without any restraints for the production run. In addition, MD simulations of 1 μs were carried out for HA46 both in explicit and implicit solvent in order to check for possible long-time scale artifacts.

Analysis of Global Properties. The following global properties of the simulated molecules were analyzed:

- *RMSD variance*: Root-mean-square deviation to the initial structure for an MD productive run was analyzed in terms of its standard deviation from the mean for heavy atoms and all pseudoatoms for AA and CG simulations.
- *Radius of gyration*.
- *Length of the molecule* was calculated as the distance between the centers of mass of two terminal sugar rings and two "ring" pseudoatoms for AA and CG simulations.

For these global properties, Pearson correlation coefficients of mean values (r_{mean}) and variances (r_{SD}) between different type of simulations were calculated.

Analysis of Local Properties. The ptraj module of AMBER was used to obtain the distributions of bonds, angles, and dihedral angles per type for both AA and CG simulations. For bonds and angles, mean values and standard deviations were compared. For dihedral angles, in addition, the obtained distributions were analyzed visually because the CG model was constructed to yield *a priori* only one maximum in the distribution, which can result in ambiguous differences when compared to the distributions obtained from AA simulations only in terms of mean and variance.

Data analysis and its graphical representation were carried out with the R package.[55]

RESULTS AND DISCUSSION

Analysis of GAG−Protein Interactions. In order to verify our choice of the pseudoatoms used in this CG model, we analyzed which GAG functional groups are energetically most important for GAG−protein interactions based on the PDB data. From 11 complexes and 367 interactions analyzed in total, about half were hydrogen bonds, whereas the other half corresponded to salt bridges (Table 1). Although it was reported

Table 1. Interactions of GAG Atoms Per Type in Studied GAG−Protein Complexes[a]

interation type, amount (fraction)	interaction per atom type (fraction)	ΔG, kcal/mol
hydrogen bond: 206 (0.56)	all types:	−4.6 ± 5.5
	100 (0.49) − SO_4^-	−5.7 ± 4.8
	46 (0.22) − COO^-	−5.2 ± 8.0
	45 (0.22) − OH + O_5 + O_{gl}	−2.0 ± 3.5
	9 (0.04) − O_{2N}	−4.3 ± 1.3
	3 (0.01) − C_1	−0.7 ± 0.3
	3 (0.01) − N_2	−4.0 ± 4.1
salt bridge: 150 (0.41)	all types:	−4.8 ± 2.5
	97 (0.65) − SO_4^-	−4.8 ± 2.5
	53 (0.35) − COO^-	−5.0 ± 2.6
hydrogen−π: 11 (0.03)	11 (1): H_{Cring}	−0.8 ± 0.2
all interactions: 367 (1.0)	197 (0.53) − SO_4^-	−5.2 ± 3.8
	99 (0.27) − COO^-	−5.1 ± 5.8
	45 (0.13) − OH + O_5 + O_{gl}	−2.0 ± 3.5
	9 (0.02) − O_{2N}	−4.3 ± 1.3
	3 (0.008) − C_1	−0.7 ± 0.3
	3 (0.008) − N_2	−4.0 ± 4.1
	3 (0.008) − H_{Cring}	−0.8 ± 0.2

[a] O_5 − ring oxygen atom; O_{gl} − oxygen atom from glycosidic linkage; N_2 − nitrogen atom bound to C_2; O_{2N} − oxygen atom from NAc-group; and H_{Cring} − nonpolar hydrogen atoms bound to ring carbon atoms. ΔG: free energy per atom−atom interaction.

dx.doi.org/10.1021/ci500669w | J. Chem. Inf. Model. 2015, 55, 114−124

that hydrogen−π interactions can play an important role for saccharides in general,[56] it is not the case for highly charged GAG molecules (the fraction of these interactions is only 0.03). Per atom type, the atoms of SO_4^- and COO^- groups contribute most favorably to binding. The impact of these interactions are comparable with the impact of N_2 and O_{2N} but are 2 orders of magnitude more abundant (197, 99 and 3, 9 interactions in total, respectively). Hydroxyl groups, O_5 and O_{gl}, though being also abundant (45 interactions in total), contribute about twice less favorably per interaction. When summarizing all energetic contributions per functional groups, SO_4^- and COO^- together contribute about 0.9 to all interactions, while the next hightest impact is 0.05 from hydroxyl groups, O_5 and O_{gl} together (Table 2).

Table 2. Free Energy Contributions of GAG Functional Groups and Atoms in the Studied 11 GAG−Protein Complexes[a]

functional groups/atoms	$\Delta G_{\sum \text{contributions per functional group}}/\Delta G_{\sum \text{all contributions}}$
SO_4^-	0.6070
COO^-	0.3013
$OH + O_5 + O_{gl}$	0.0549
O_{2N}	0.0200
N_2	0.0071
H_{Cring}	0.0040

[a]O_5 − ring oxygen atom; O_{gl} − oxygen atom from glycosidic linkage; N_2 − nitrogen atom bound to C_2; O_{2N} − oxygen atom from NAc-group; and H_{Cring} − nonpolar hydrogen atoms bound to ring carbon atoms.

On the basis of these findings, we conclude that electrostatic interactions are highly predominant among GAG−protein interactions, and therefore, the most important goal for a CG model of GAGs is to be able to properly describe and reproduce the behavior of charged groups.

Choice of Pseudoatoms and Construction of CG Residue Libraries. We introduce 28 pseudoatom types in total (Table 3). Five of them (OY, OV, OX, OA) are identical to the AA types from GLYCAM06 and represent the oxygen atoms of four glycosidic linkages: O_3 of GlcNAc and GalNAc, O_4 of GlcNS and GlcU/IdoU. This was done in order to properly account for GAG molecule flexibility rather than for considering these atoms for intra- and intermolecular interactions. Then, we define eight pseudoatom types (IY, TY, IV, TV, IH, TH, IA, TA) corresponding to the center of mass of the sugar rings including their internal (starting with I) and terminal variants (starting with T). In the two-letters code, the second letters Y, V, H, and A in the pseudoatom types correspond specifically to GlcNAc, GalNAc, GlcNS, and GlcU/IdoU rings, respectively. FC and AC pseudoatoms correspond to CH_2OH and NAc groups. All the pseudoatoms listed above are considered to be neutral. The rest of pseudoatoms are charged with a net charge of −1 and represent the COO^- group of GlcU/IdoU (KY/KW, respectively) and nine sulfate groups, eight of which are O-sulfates and one is an N-sulfate. Eleven AA molecule libraries for Lennard−Jones parameter calculations for CG pseudoatoms were created, and the parameters obtained by PMF were introduced into the CG model (Table 4). The parameters for all eight O-sulfates and for KY/KW were obtained using two unique probes. In the CG

Table 3. CG Pseudoatom Types[a]

CG atom	corresponding AA atoms	description	mass (a.u.)	charge	AA atoms defining center of mass
IY	$C_5O_2H_6$	internal pyranose ring in GlcNAc	97.10	0	C_1,C_2,C_3,C_4,C_5,O_5
IV	$C_5O_2H_6$	internal pyranose ring in GalNAc	97.10	0	C_1,C_2,C_3,C_4,C_5,O_5
IZ	$C_5O_2H_6$	internal pyranose ring in GlcNAc	97.10	0	C_1,C_2,C_3,C_4,C_5,O_5
TY	$C_5O_3H_7$	terminal pyranose ring in GlcNAc	115.11	0	C_1,C_2,C_3,C_4,C_5,O_5
TV	$C_5O_3H_7$	terminal pyranose ring in GalNAc	115.11	0	C_1,C_2,C_3,C_4,C_5,O_5
TZ	$C_5O_3H_7$	terminal pyranose ring in GlcNAc (deHS)	115.11	0	C_1,C_2,C_3,C_4,C_5,O_5
IH	$C_5O_2H_7$	internal pyranose ring in GlcNS	99.11	0	C_1,C_2,C_3,C_4,C_5,O_5
TH	$C_5O_2H_8$	terminal pyranose ring in GlcNS	100.12	0	C_1,C_2,C_3,C_4,C_5,O_5
IA	$C_5O_3H_7$	internal pyranose ring in GlcU/IdoU	115.11	0	C_1,C_2,C_3,C_4,C_5,O_5
TA	$C_5O_4H_8$	terminal pyranose ring in GlcU/IdoU	132.11	0	C_1,C_2,C_3,C_4,C_5,O_5
FC	CH_2OH	C_6-group GlcNAc/GalNAc	31.03	0	C_6,O_6
AC	$NHCOCH_3$	NAc group	58.06	0	N_2,C_{2N},O_{2N},C_{ME}
SN	$NHSO_3$	N-sulfate group	95.08	−1	S_1,O_{1S},O_{2S},O_{3S}
SS	SO_3	6-sulfate group (HA6/CS6)	80.06	−1	S_1,O_{1S},O_{2S},O_{3S}
SK	SO_3	4-sulfate group GlcNAc (HA4)	80.06	−1	S_1,O_{1S},O_{2S},O_{3S}
SQ	SO_3	4-sulfate group GalNAc (CS4)	80.06	−1	S_1,O_{1S},O_{2S},O_{3S}
SP	SO_3	4-sulfate group GlcNAc (CS46)	80.06	−1	S_1,O_{1S},O_{2S},O_{3S}
ST	SO_3	3-sulfate group GlcU/IdoU	80.06	−1	S_1,O_{1S},O_{2S},O_{3S}
SD	SO_3	2-sulfate group GlcU/IdoU	80.06	−1	S_1,O_{1S},O_{2S},O_{3S}
SL	SO_3	4-sulfate group GlcNAc (HA46)	80.06	−1	S_1,O_{1S},O_{2S},O_{3S}
SF	SO_3	6-sulfate group GlcNAc (HA46/CS46)	80.06	−1	S_1,O_{4S},O_{5S},O_{6S}
KY	CO_2	carboxyl group in GlcU	44.01	−1	C_6,O_{6A},O_{6B}
KW	CO_2	carboxyl group in IdoU	44.01	−1	C_6,O_{6A},O_{6B}
OY	O	glycosidic oxygen: GlcNAc	16.00	0	O_3
OV	O	glycosidic oxygen: GalNAc	16.00	0	O_3
OA	O	glycosidic oxygen: GlcU/IdoU	16.00	0	O_4
OX	O	glycosidic oxygen: GlcNS	16.00	0	O_4
OZ	O	glycosidic oxygen: GlcNAc (deHS)	16.00	0	O_4

[a]In total, 48 residues corresponding to three variants of GlcNAc, GlcNAc (deHS), GalNAc, GlcNAc4S, GlcNAc6, GlcNAc46S, GalNAc4S, GalNAc6S, GalNAc46S, GlcNS6S, GlcU, IdoU, GlcU2S, GlcU3S, GlcU23S, IdoU2S, IdoU3S, and IdoU23S were created.

dx.doi.org/10.1021/ci500669w | J. Chem. Inf. Model. 2015, 55, 114−124

Table 4. Lennard–Jones Parameters for CG Pseudoatoms in AMBER Consistent Format[a]

CG atom	AA molecule for PFM simulation	R_{vdw} (Å)	EDEP (kcal/mol)
IY/IV/IZ	H–$C_5O_2H_6$	1.9605	3.4048
TY/TV/TZ	H–$C_5O_3H_7$	1.9153	4.5279
IH	H–$C_5O_2H_7$	2.1680	3.5100
TH	H–$C_5O_2H_8$	2.1300	4.3570
IA	H–$C_5O_3H_7$	1.9270	4.6340
TA	H–$C_5O_4H_8$	1.9748	5.2856
FC	H–CH_2OH	1.7830	0.6188
AC	H–NHCO–CH_3	1.7500	2.0323
SN	H–NH–SO_3^-	2.1085	2.5276
SS	H–SO_3^-	1.6900	2.3137
SK	H–SO_3^-	1.6900	2.3137
SQ	H–SO_3^-	1.6900	2.3137
ST	H–SO_3^-	1.6900	2.3137
SD	H–SO_3^-	1.6900	2.3137
SL	H–SO_3^-	1.6900	2.3137
SP	H–SO_3^-	1.6900	2.3137
SF	H–SO_3^-	1.6900	2.3137
KY	H–C–O_2^-	1.6515	1.0201
KW	H–C–O_2^-	1.6515	1.0201
OY/OV/OA/OX/OZ	from Glycam06 AA force field	1.6837	0.1700

[a]H– corresponds to the additional hydrogen, which was used for PMF calculations but does not belong to any of the pseudoatoms. For terminal ring pseudoatoms, the values represent means from terminal rings at reducing and at nonreducing ends. EDEP: energy well depth.

model, every saccharide residue can be of three types: internal (with two connective atoms), terminal at nonreducing end, and terminal at reducing end (with one connective atom). Three libraries of three letters code with the nomenclature using the last letters "I" for internal, "N" for GLYCAM06 internal residue + ROH, and "R" for GLYCAM06 residues starting with "0" were created in AMBER-compatible format (Table S1, Supporting Information).

Special Case of Heparin. Development of the CG model for HE is principally different than for other GAGs due to the fact that IdoU2S can adopt both $^1C_4$ and ^2S_O ring conformations with comparable probabilities. Therefore, prior to the implementation of HE parameters in our CG model, we compared local and global properties of HE^{chair} and HE^{boat}. Whereas global properties do not differ significantly, as expected from previous experimental data[57](Table S2, Supporting Information), some bonds, angles, and dihedrals demonstrate essentially different behavior in MD simulations. In particular, there are two out of six bonds ([IA,TA]-OX and [IA,TA]-SD) that have mean differences higher than 0.1 Å (Table S3, Supporting Information). There are five angles out of forteen, for which the mean differences between HE^{chair} and HE^{boat} are larger than 10° (Table S4, Supporting Information). Therefore, for these two bonds and all angles, we developed such parameters that an equilibrium bond/angle length is an average of the corresponding means obtained in the simulations of HE^{chair} and HE^{boat}, while the force constant is chosen to yield such a parameter's variance in the GC model so that it contains the variances from both simulations of these different conformations (in particular, the new variance is equal to the root-mean-squared sum of squared variances). Ideally, such an approach represents an approximation, which would allows combining two AA HE models into one CG HE model featuring higher flexibility. For dihedral angles, most of the probability density distribution maxima for HE^{chair} and HE^{boat} were similar, but the shapes of the density distributions were essentially different (the data are presented later in the following subsection). Therefore, we combined the data on dihedral angles using a similar strategy as it was done for bonds/angles. We took an average of maxima as the maximum for the CG model, whereas we used the lowest of the energetic barriers obtained from the AA MD simulations to integrate them into the CG model.

Comparison of Global and Local Properties Obtained by AA and CG Approaches. Following the procedures described above, we obtained CG parameters for 28 bonds, 58 angles, and 50 dihedral angles for internal CG atoms. These parameters were also extended to the terminal CG atoms (Tables 5, 6, 7). Analysis of the results shows that the use of several of the so obtained bond and angle parameters in the CG model was not sufficient to reproduce well the values from AA simulations due to the approximations in the model.

Table 5. Covalent Bond Parameters for CG Pseudoatoms in AMBER Consistent format

[a]CG1	[b]CG2	[c]RK (kcal mol^{-1} Å^{-2})	[d]REQ (Å)	[e]AA MD
[IY,TY]	OY	121.6	2.936	HA
[IV,TV]	OV	137.0	2.960	deCS
[IY,TY]	OA	141.8	2.811	HA
[IV,TV]	OA	138.9	2.813	deCS
[IA,TA]	OY/OV	137.6	2.808	HA, deCS
[IA,TA]	OA	118.5	2.888	HA, deCS
[IY,TY]	AC	111.3	4.003	HA
[IV,TV]	AC	115.2	4.002	deCS
[IA,TA]	KY	110.7	3.366	HA, deCS
[IA,TA]	KW	67.2	2.940	DS
[IY,TY]	SS	1.9	4.983	HA6
[IY,TY]	SF	*100.0*	4.983	HA6
[IV,TV]	SS	3.7	5.211	CS6[f]
[IV,TV]	SF	*100.0*	5.211	CS6[f]
[IY,TY]	FC	17.9	3.229	HA[f]
[IV,TV]	FC	45.1	3.334	deCS
[IY,TY]	SK	16.6	4.343	HA4[f]
[IY,TY]	SL	*100.0*	4.343	HA46
[IV,TV]	SQ	49.0	4.054	CS4
[IV,TV]	SP	*100.0*	4.049	CS46
[IA,TA]	SD	*100.0*	4.321	HA2[f], CS2[f]
[IA,TA]	ST	*100.0*	4.310	HA3[f], CS3[f]
[IH,TH]	OX	119.2	2.897	HE^{chair}
[IH,TH]	OA	60.8	2.341	HE^{chair}
[IA,TA]	OX	30.0	2.492	$HE^{chair,boat}$
[IH,TH]	SN	45.4	4.123	HE^{chair}
[IH,TH]	SS	*100.0*	5.037	HE^{chair}
[IA,TA]	SD	9.9	4.236	$HE^{chair,boat[f]}$
[IZ,TZ]	OZ	118.0	2.895	deHS
[IZ,TZ]	OA	62.4	2.350	deHS
[IZ,TZ]	AC	80.6	3.998	deHS
[IZ,TZ]	FC	19.6	3.207	deHS[f]
OZ	[IA,TA]	138.4	2.809	deHS

[a]CG atom 1. [b]CG atom 2 in the covalent bond. [c]Force constant. [d]Equilibrium bond length. [e]AA MD simulation(s), from which the parameter was obtained. [f]Parameter probability density distribution from this AA simulation had more than 1 maximum. Squared parentheses stand for logical connective "or". Artificially high force constants are in italics.

For example, for the [IY,TY]–SF bond and the SL–[IY,TY]–SF angle, the force constant parameters obtained by the

dx.doi.org/10.1021/ci500669w | J. Chem. Inf. Model. 2015, 55, 114–124

Table 6. Angle Parameters for CG Pseudoatoms in AMBER Consistent format

[a]CG1	[b]CG2	[c]CG3	[d]TK (kcal mol^{-1}rad^{-2})	[e]TEQ (deg)	[f]AA MD
[IY,TY]	OY	[IA,TA]	52.5	140.8	HA
[IV,TV]	OV	[IA,TA]	38.8	137.7	deCS
[IY,TY]	OA	[IA,TA]	24.8	139.6	HA
[IV,TV]	OA	[IA,TA]	28.0	140.1	deCS
OA	IY	OY	196.2	121.1	HA
OA	IV	OV	185.7	121.4	deCS
OY/OV	IA	OA	77.6	173.6	HA, deCS
OY	[IY,TY]	[SS,SF]	5.7	118.1	HA6[g]
OV	[IV,TV]	[SS,SF]	4.4	115.6	CS6[g]
OA	[IY,TY]	[SS,SF]	7.7	104.9	HA6
OA	[IV,TV]	[SS,SF]	4.6	121.0	CS6[g]
OY	[IY,TY]	[SK,SL]	34.8	61.1	HA4
OA	[IY,TY]	[SK,SL]	17.4	163.1	HA4
OV	[IV,TV]	[SQ,SP]	66.2	64.8	CS4
OA	[IV,TV]	[SQ,SP]	23.8	130.1	CS4
OY	[IY,TY]	AC	45.2	58.7	HA
OV	[IV,TV]	AC	41.2	59.4	deCS
OA	[IY,TY]	AC	43.7	64.0	HA
OA	[IV,TV]	AC	39.8	63.5	deCS
OY	[IY,TY]	FC	18.5	123.2	HA[g]
OV	[IV,TV]	FC	23.6	127.1	deCS
OA	[IY,TY]	FC	18.0	111.6	HA[g]
OA	[IV,TV]	FC	27.0	109.4	deCS
OY/OV	[IA,TA]	KY	165.5	117.1	HA, deCS
OA	[IA,TA]	KY	163.5	60.6	HA, deCS
OV	[IA,TA]	KW	54.6	114.4	DS
OA	[IA,TA]	KW	127.4	65.2	DS
OY/OV	[IA,TA]	SD	31.3	66.5	HA2′, CS2′
OA	[IA,TA]	SD	28.7	116.9	HA2′, CS2′
OY/OV	[IA,TA]	ST	19.6	114.2	HA3′[g], CS3′[g]
OA	[IA,TA]	ST	32.7	71.6	HA3′, CS3′
SL	[IY,TY]	SF	*500.0*	77.7	HA46
SP	[IV,TV]	SF	*500.0*	81.0	CS46
AC	[IY,TY]	[SS,SF]	4.8	144.9	HA6
AC	[IV,TV]	[SS,SF]	13.0	159.6	CS6
SK	[IY,TY]	FC	18.6	67.3	HA4[g]
SQ	[IV,TV]	FC	21.1	74.6	CS4[g]
AC	[IY,TY]	FC	13.7	160.6	HA
AC	[IV,TV]	FC	27.4	167.5	deCS
[SK,SL]	[IY,TY]	AC	31.2	119.8	HA4
[SQ,SP]	[IV,TV]	AC	27.6	107.2	CS4
KY	[IA,TA]	SD	21.1	164.0	HA2′, CS2′
KY	[IA,TA]	ST	*500.0*	120.2	HA3′, CS3′
SD	[IA,TA]	ST	*500.0*	77.1	HA2′3′[g], CS2′3′[g]
OX	[IA,TA]	SD	7.7	82.9	HE[chair,boat]
OA	[IA,TA]	SD	8.9	107.5	HE[chair,boat]
[IH,TH]	OX	[IA,TA]	11.2	150.4	HE[chair,boat]
[IH,TH]	OA	[IA,TA]	9.5	162.5	HE[chair,boat]
OX	[IH,TH]	OA	16.8	124.6	HE[chair,boat]
OX	[IA,TA]	OA	42.3	166.8	HE[chair,boat]
OX	[IA,TA]	KW	38.8	108.5	HE[chair,boat]
OA	[IA,TA]	KW	66.3	61.7	HE[chair,boat]
OX	[IH,TH]	SN	3.8	126.6	HE[chair,boat]
OA	[IH,TH]	SN	12.3	71.2	HE[chair,boat]
OX	[IH,TH]	SS	4.7	74.3	HE[chair,boat]
OA	[IH,TH]	SS	5.3	124.2	HE[chair,boat]
SN	[IH,TH]	SS	*500.0*	142.6	HE[chair,boat]
KW	[IA,TA]	SD	*500.0*	152.3	HE[chair,boat]
[IZ,TZ]	OZ	[IA,TA]	20.5	140.5	deHS
[IZ,TZ]	OA	[IA,TA]	18.6	153.9	deHS
OA	[IZ,TZ]	OZ	29.6	126.3	deHS

dx.doi.org/10.1021/ci500669w | J. Chem. Inf. Model. 2015, 55, 114−124

Table 6. continued

[a]CG1	[b]CG2	[c]CG3	[d]TK (kcal mol^{-1}rad^{-2})	[e]TEQ (deg)	[f]AA MD
OZ	[IA,TA]	OA	76.3	173.3	deHS
OZ	[IZ,TZ]	AC	31.0	125.5	deHS
OZ	[IZ,TZ]	FC	56.1	67.9	deHS[g]
OA	[IZ,TZ]	AC	30.8	71.6	deHS
OA	[IZ,TZ]	FC	8.7	113.4	deHS[g]
AC	[IZ,TZ]	FC	19.6	157.3	deHS
OZ	[IA,TA]	KY	172.1	117.0	deHS

[a]CG atom 1. [b]CG atom 2. [c]CG atom 3 in the angle. [d]Force constant. [e]Equilibrium angle value. [f]AA MD simulation(s), from which the parameter was obtained. [g]Parameter probability density distribution from this AA simulation had more than 1 maximum. Squared parentheses stand for logical connective "or". Artificially high force constants are in italics.

Boltzmann conversion are too loose because the CG model does not take into account any particular accessibility of the *gg/gt/tg* dihedral angle conformations corresponding to the C_5-C_6 bond (as the AA approach explicitly does) but rather uses the whole distribution containing of these weighted conformations. Therefore, we introduced seven and six artificially high force constants for bonds and angles, respectively, so that the means for these parameters obtained in CG and AA MD simulations agree well. This leads to the fact that the obtained variances for these parameters are essentially lower for the CG model than in AA simulations.

Comparison of global GAG properties (RMSD, $R_{gyration}$, length) in terms of means and variances obtained in AA and CG simulations shows a fair correlation for RMSD variances in the case of the implicit solvent CG simulations (r_{SD} = 0.472), and weak correlations for the mean of $R_{gyration}$ (r_{mean} = 0.215) and the mean of length (r_{mean} = 0.236) for explicit solvent, the variance of $R_{gyration}$ (r_{SD} = 0.303) in explicit solvent, and the mean of $R_{gyration}$ (r_{mean} = 0.220) in implicit solvent (Table 8). RMSD and $R_{gyration}$ are very well reproduced by both explicit and implicit solvent CG simulations; the differences are about 0.1 Å for RMSD variance and below 0.02 Å for the mean of $R_{gyration}$. RMSD, $R_{gyration}$, and length variances are higher for AA simulations, which is to be expected due to the fact that there are more atoms in the AA representation, and therefore, the net amount of fluctuations is higher for AA simulations (Tables S5−S7, Supporting Information). The mean of the GAG length is very well comparable for AA and explicit solvent CG simulations, whereas for implicit solvent CG simulations, there is a clear overestimation of ∼10% of the GAG length, which probably reflects the more rigid behavior of the GAG in the CG implicit solvent model (Table S7, Supporting Information). In addition, we compared $R_{gyration}$ and length for the 13 HE molecules with different degrees of polymerization for AA, $CG_{explicit}$, and $CG_{implicit}$ MD simulations with the linear HE model based on the HE dp12 NMR structure (PDB ID: 1HNP) and the available experimental data from X-ray scattering-based modeling[58,59] (Tables S8 and S9, Supporting Information). The data from all simulations agree very well with the linear HE model in terms of $R_{gyration}$ and in terms of length for HE shorted than dp32. For longer HE molecules, CG MD simulations reveal bending from linear structures, which is reflected by lower length values than for the corresponding linear HE models. When comparing with the experimental data from X-ray scattering, both $R_{gyration}$ and length from the MD simulations are lower for HE dp6, are close for HE dp12−32, and are significantly higher for HE longer than dp36. These differences for longer GAGs could be attributed to the higher rigidity in the CG model, though significant bending in comparison to the linear HE model based on the NMR structure was observed. In terms of dynamic/diffusion properties, GAGs are anticipated to behave very similarly when comparing CG and AA models. Speeding up of dynamics like, for example, the one observed in the MARTINI CG force field[45] is not expected because we use the Boltzmann conversion method, which should not change the heights of energy barriers significantly.

In terms of local parameters, for all GAGs, there are very high correlations for the mean values of bonds, angles, and dihedral angles (larger than 0.99, 0.99, and 0.98, respectively) (Tables 8 and 9). The correlations for the variances are still high but significantly lower than for the means due to the fact that there are several high artificial force constants introduced to the bonds and angles parameters (see above) yielding lower variances in case of CG simulations (Tables S10 and S11, Supporting Information). The mean of the differences for local parameters is lower than 0.01 Å, 1°, and 7° for bonds, angles, and dihedral angles, respectively. The main reason for the higher differences for dihedral angle means originates from the procedure of the development of dihedral angle parameters in this CG model. Whereas in the AA simulation the dihedral angles probability density distribution can have several maxima and, therefore, several energy minima, only the global minimum is reproduced in the CG model (see examples for IY-OY-IA-OA, IV−OV−IA−OA, IY−OA−IA−OY, IV−OA−OA−OV, Figure 1). That is why the differences between the values of dihedral angles obtained by AA and CG simulations should not be simply compared in terms of mean and variance (Table S12, Supporting Information) as it is done for bonds and angles. Instead, this should be done in terms of visual representation of probability density distributions or free energy landscapes.

Due to the reasons mentioned above and because the density distributions of dihedral angle values for HE^{chair} and HE^{boat} are significantly different, we do not show the comparison of dihedral angles values obtained for HE in AA and CG simulations in terms of mean and variances in the table. Instead, we show the entire distributions graphically for all 12 analyzed dihedral angles (Figure S1, Supporting Information). The analysis of these distributions reveals good reproduction of the AA global maxima by the CG approach. However, as it is *a priori* implicated in the CG model, local maxima are not reproduced, which at the end contributes to higher rigidity of the GAG molecules' behavior in the CG model.

In order to check if there are possible long time scale artifacts in CG simulations, we carried out 1 μs MD simulations for HA46 and compared the obtained global and local parameters with the ones from 0.1 μs (Tables S13−S16, Supporting Information). The results show that neither global nor local parameters from these two simulations of different length significantly differ. This suggests stability of the derived CG approach at longer time scales.

dx.doi.org/10.1021/ci500669w | J. Chem. Inf. Model. 2015, 55, 114−124

Table 7. Dihedral Angle Parameters for CG Pseudoatoms in AMBER Consistent Format

[a]CG1	[b]CG2	[c]CG3	[d]CG4	[e]IDIVF	[f]PK, (kcal/mol)	[g]Phase (deg)	[h]PN	[i]AA MD
[IY,TY]	OY	IA	OA	1	1.21	−145.8	1	HA
[IV,TV]	OV	IA	OA	1	1.09	−143.2	1	deCS
[IY,TY]	OA	[IA,TA]	OY	1	0.91	−17.8	1	HA
[IV,TV]	OA	[IA,TA]	OV	1	0.93	−18.3	1	deCS
[IA,TA]	OA	IY	OY	1	2.79	−18.1	1	HA
[IA,TA]	OA	IV	OV	1	2.32	−18.9	1	deCS
[IA,TA]	OY	[IY,TY]	OA	1	2.29	55.0	1	HA
[IA,TA]	OV	[IV,TV]	OA	1	2.26	33.4	1	deCS
[IV,TV]	OV	[IA,TA]	KW	1	2.17	−161.2	1	DS
[IV,TV]	OA	[IA,TA]	KW	1	1.98	70.2	1	DS
[IY,TY]	OY	[IA,TA]	KY	1	1.85	139.8	1	HA
[IV,TV]	OV	[IA,TA]	KY	1	1.77	140.1	1	deCS
[IY,TY]	OA	[IA,TA]	KY	1	2.28	70.1	1	HA
[IV,TV]	OA	[IA,TA]	KY	1	2.54	74.3	1	deCS
[IY,TY]	OY	[IA,TA]	SD	1	2.66	−33.6	1	HA2′
[IV,TV]	OV	[IA,TA]	SD	1	2.65	−25.5	1	CS2′
[IY,TY]	OA	[IA,TA]	SD	1	1.89	−90.4	1	HA2′
[IV,TV]	OA	[IA,TA]	SD	1	1.92	−96.1	1	CS2′
[IY,TY]	OY	[IA,TA]	ST	1	2.59	−21.3	1	HA3′
[IV,TV]	OV	[IA,TA]	ST	1	2.64	−21.3	1	CS3′
[IY,TY]	OA	[IA,TA]	ST	1	1.78	−125.1	1	HA3′
[IV,TV]	OA	[IA,TA]	ST	1	1.68	−129.6	1	CS3′
[IA,TA]	OY	[IY,TY]	[SS,SF]	1	1.75	−88.5	1	HA6
[IA,TA]	OV	[IV,TV]	[SS,SF]	1	2.21	−125.5	1	CS6
[IA,TA]	OA	[IY,TY]	[SS,SF]	1	2.61	123.6	1	HA6
[IA,TA]	OA	[IV,TV]	[SS,SF]	1	2.21	158.5	1	CS6
[IA,TA]	OY	[IY,TY]	FC	1	2.64	−95.3	1	HA
[IA,TA]	OV	[IV,TV]	FC	1	2.29	−114.9	1	deCS
[IA,TA]	OA	[IY,TY]	FC	1	2.47	136.0	1	HA
[IA,TA]	OA	[IV,TV]	FC	1	2.27	151.9	1	deCS
[IA,TA]	OY	[IY,TY]	[SK,SL]	1	1.29	−85.4	1	HA4
[IA,TA]	OA	[IY,TY]	[SK,SL]	1	1.77	72.5	1	HA4
[IA,TA]	OV	[IV,TV]	[SQ,SP]	1	1.86	−38.6	1	CS4
[IA,TA]	OA	[IV,TV]	[SQ,SP]	1	2.41	62.3	1	CS4
[IA,TA]	OY	[IY,TY]	AC	1	2.78	66.1	1	HA
[IA,TA]	OV	[IV,TV]	AC	1	2.33	46.1	1	deCS
[IA,TA]	OA	[IY,TY]	AC	1	2.43	−27.9	1	HA
[IA,TA]	OA	[IV,TV]	AC	1	2.28	−29.0	1	deCS
[IH,TH]	OX	IA	OA	1	0.82	−176.1	1	HE^{chair,boat}
[IH,TH]	OA	IA	OX	1	0.19	53.8	1	HE^{chair,boat}
[IA,TA]	OX	[IH,TH]	OA	1	2.88	−26.6	1	HE^{chair,boat}
[IA,TA]	OA	[IH,TH]	OX	1	1.28	−76.6	1	HE^{chair,boat}
[IH,TH]	OX	[IA,TA]	KW	1	1.13	147.5	1	HE^{chair,boat}
[IH,TH]	OA	[IA,TA]	KW	1	0.40	79.9	1	HE^{chair,boat}
[IH,TH]	OX	[IA,TA]	SD	1	1.03	−37.7	1	HE^{chair,boat}
[IH,TH]	OA	[IA,TA]	SD	1	0.37	−100.7	1	HE^{chair,boat}
[IA,TA]	OX	[IH,TH]	SN	1	2.47	−117.2	1	HE^{chair,boat}
[IA,TA]	OA	[IH,TH]	SN	1	1.25	39.2	1	HE^{chair,boat}
[IA,TA]	OX	[IH,TH]	SS	1	2.63	98.4	1	HE^{chair,boat}
[IA,TA]	OA	[IH,TH]	SS	1	1.13	−173.0	1	HE^{chair,boat}
[IZ,TZ]	OZ	IA	OA	1	1.35	−148.4	1	deHS
[IZ,TZ]	OA	[IA,TA]	OZ	1	1.22	45.2	1	deHS
[IA,TA]	OA	IZ	OZ	1	1.90	−79.8	1	deHS
[IA,TA]	OZ	[IZ,TZ]	OA	1	3.37	−25.3	1	deHS
[IA,TA]	OZ	[IZ,TZ]	AC	1	2.64	−116.8	1	deHS
[IA,TA]	OA	[IZ,TZ]	AC	1	1.91	44.4	1	deHS
[IA,TA]	OZ	[IZ,TZ]	FC	1	3.02	82.9	1	deHS
[IA,TA]	OA	[IZ,TZ]	FC	1	1.80	−155.7	1	deHS
[IZ,TZ]	OZ	[IA,TA]	KY	1	1.39	140.4	1	deHS
[IZ,TZ]	OA	IA	KY	1	2.47	117.0	1	deHS

[a]CG atom 1. [b]CG atom 2. [c]CG atom 3. [d]CG atom 4 in the dihedral angle. [e]Factor by which the torsional barrier is divided. [f]Barrier height divided by a factor of 2. [g]Phase shift angle in the torsional function. [h]Periodicity of the torsional barrier. [i]AA MD simulation(s), from which the parameter was obtained.

dx.doi.org/10.1021/ci500669w | J. Chem. Inf. Model. 2015, 55, 114−124

Table 8. Comparison of Global and Local Properties for HA/CS from AA and CG Simulations

	AA vs $CG_{explicit}$				AA vs $CG_{implicit}$			
global parameters	mean difference	r_{mean}	SD difference	r_{SD}	mean difference	r_{mean}	SD difference	r_{SD}
RMSD (Å)	–	–	0.06 ± 0.18	0.099	–	–	0.11 ± 0.13	0.472
$R_{gyration}$ (Å)	0.01 ± 0.34	0.215	0.26 ± 0.15	0.303	0.02 ± 0.37	0.220	0.29 ± 0.17	−0.109
length (Å)	0.79 ± 3.36	0.236	2.02 ± 1.13	−0.408	−5.23 ± 2.10	−0.376	1.71 ± 1.23	−0.632
local parameters	mean difference	r_{mean}	SD difference	r_{SD}	mean difference	r_{mean}	SD difference	r_{SD}
bonds (Å)	−0.007 ± 0.053	0.998	0.044 ± 0.085	0.619	−0.002 ± 0.034	0.999	0.045 ± 0.088	0.593
angles (deg)	0.33 ± 1.66	0.999	1.77 ± 1.94	0.807	0.15 ± 1.76	0.999	1.88 ± 2.03	0.785
dihedrals (deg)	3.4 ± 21.6	0.980	12.4 ± 15.4	0.579	6.5 ± 20.7	0.983	12.6 ± 16.3	0.423

Table 9. Comparison of Local Properties for HE from AA and CG Simulations

	AA vs $CG_{explicit}$				AA vs $CG_{implicit}$			
local parameters	mean difference	r_{mean}	SD difference	r_{SD}	mean difference	r_{mean}	SD difference	r_{SD}
bonds (Å)	−0.018 ± 0.021	1.000	0.11 ± 0.18	−0.462	−0.013 ± 0.024	1.000	0.11 ± 0.18	−0.462
angles (deg)	0.27 ± 3.12	0.996	3.62 ± 4.73	0.322	0.33 ± 3.04	0.997	3.39 ± 4.77	0.338

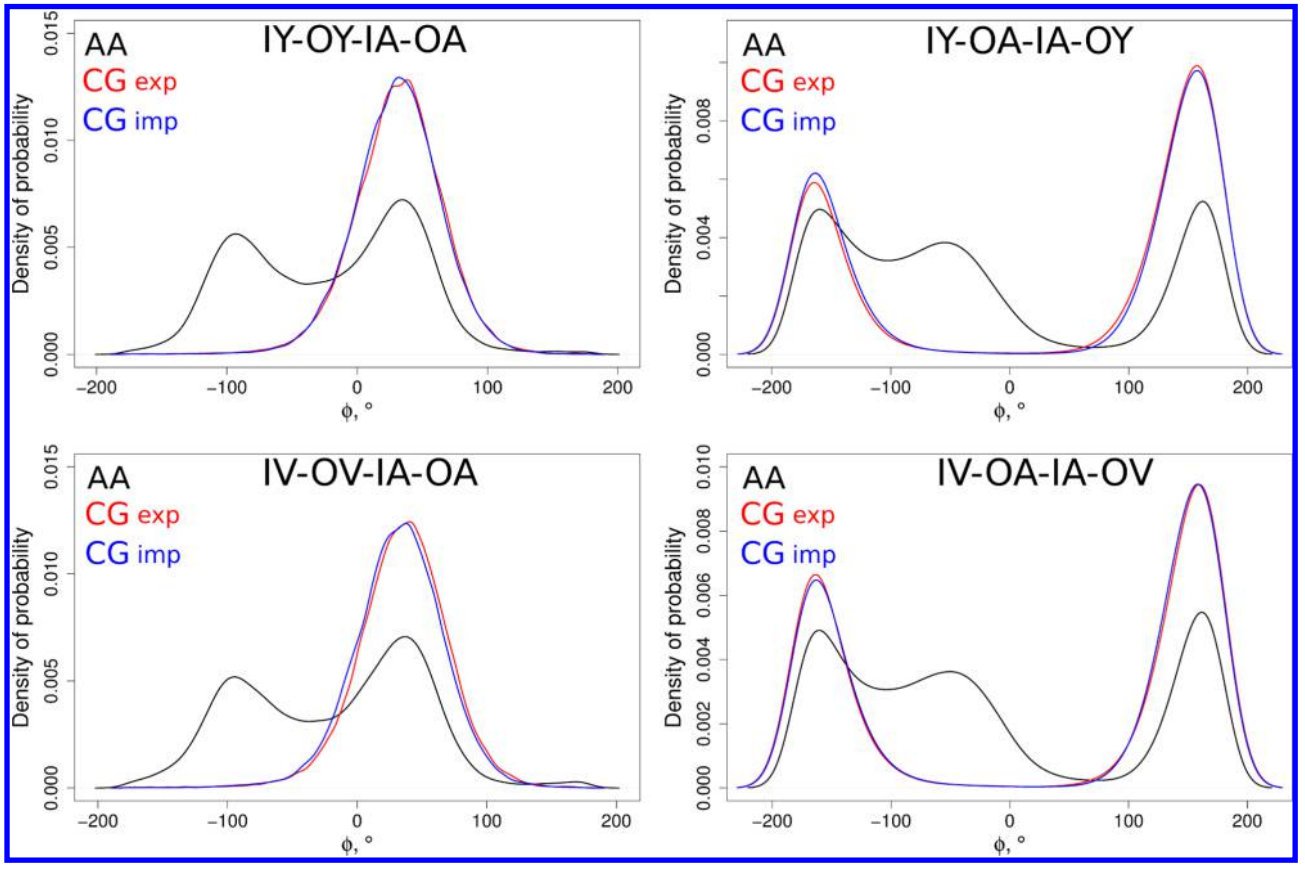


Figure 1. Probability density distribution for dihedral angles in AA and CG MD simulations: AA in black, CG in explicit solvent in red, and CG in implicit solvent in blue.

Comparison of Computational Expenses for AA and CG Approaches. Comparison of computational needs for simulating the same systems using AA and CG approaches shows that the speed up of the MD simulation in explicit solvent is insignificant (∼20%) because water molecules contribute the most to the computational expenses, whereas in implicit solvent the required calculation time decreases about 7-fold for GAGs of 10 dp and about 40-fold for GAGs of 100 dp. When using the CG method, this time decrease could allow long simulations for long GAG molecules within realistic time scales inaccessible by the AA approach.

To summarize, the CG model developed in this study yields a good general agreement for both global and local characteristics of the simulated GAGs. The limitations of the model include GAGs length overestimation in implicit solvent due to their higher rigidity in the CG model and principal differences in the distributions of several dihedral angles due to the procedure of dihedral angle parameters development. The approach used to treat electrostatic properties of pseudoatoms implies that our model is not constructed to distinguish specific hydration properties such as a degree of ionization for GAGs with the same total charge per disaccharide unit (like HA and deCS) shown to

dx.doi.org/10.1021/ci500669w | J. Chem. Inf. Model. 2015, 55, 114−124

be distinguashable in other studies.[60,61] In perspective, the developed CG parameters could be useful for docking and MD analysis of GAG−protein systems. In particular, future work will include the development of docking scoring functions and free energy calculations of GAG binding to proteins within the CG model based on AA-derived and experimental data.

CONCLUSIONS

In this study, we develop AMBER MD suite compatible CG parameters for 17 different GAGs using AA GLYCAM06 MD simulations and the Boltzmann conversion approach. $^{1}C_{4}$ and $^{2}S_{O}$ conformations of IdoU2S rings are treated implicitly within this CG model by allowing the parameters to vary so that their value ranges corresponding to both ring conformations can be sampled. Several bond and angle equilibrium value parameters had to be restricted artificially by applying strong force constants in order to maintain their mean values close to the corresponding mean values in AA simulations. Global and local GAG properties observed in AA simulations are well reproduced by CG simulations; although due to its approximations, the presented CG model yields slightly higher rigidity of GAG molecules in comparison to AA simulations. The potential use of the obtained CG parameters is promising, especially for the implicit solvent model, due to the substantial gain in speed provided compared to the AA model. The developed CG model allows efficient treatment of large systems containing GAGs at time scales currently inaccessible by the classical AA MD, and the obtained results contribute to the knowledge of conformational properties of GAGs.

ASSOCIATED CONTENT

Supporting Information

Tables S1−S16 and Figure S1. All the parameter and configuration files in AMBER-compatible format are available upon request. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: sergeys@biotec.tu-dresden.de.
*E-mail: mayte@biotec.tu-dresden.de.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Jan-Philip Gehrcke, Malte Lichtner, and Dr. Maciej Paszkowski-Rogacz for fruitful discussions and Malte Lichtner for the GAG/protein dataset. This work was supported by the German Research Council (SFB-TRR67, A7).

REFERENCES

(1) Esko, J. D.; Kimata, K.; Lindahl, U. *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2009.
(2) Hansen, S.; Miller, G.; Cole, C.; Rushton, G.; Avizienyte, E.; Jayson, G.; Gardiner, J. Tetrasaccharide Iteration Synthesis of a Heparin-Like Dodecasaccharide and Radiolabelling for in Vivo Tissue Distribution Studies. *Nat. Commun.* **2013**, *4*, 2016.
(3) Khan, S.; Fung, K. W.; Rodriguez, E.; Patel, R.; Gor, J.; Mulloy, B.; Perkins, S. The Solution Structure of Heparan Sulfate Differs from That of Heparin: Implications for Function. *J. Biol. Chem.* **2013**, *288*, 27737−27751.
(4) Raman, K.; Mencio, C.; Desai, U.; Kuberan, B. Sulfation Patterns Determine Cellular Internalization of Heparin-like Polysaccharides. *Mol. Pharmaceutics* **2013**, *10*, 1442−1449.
(5) Gandhi, N.; Mancera, R. The Structure of Glycosaminoglycans and Their Interactions with Proteins. *Chem. Biol. Drug Des.* **2008**, *72*, 455−482.
(6) Proudfoot, A. The Biological Relevance of Chemokine-Proteoglycan Interactions. *Biochem. Soc. Trans.* **2006**, *34*, 422−426.
(7) Hamel, D.; Sielaff, I.; Proudfoot, A.; Handel, T. *Methods Enzymol.* **2009**, *461*, 71−102.
(8) Shute, J. Glycosaminoglycan and Chemokine/Growth Factor Interactions. *Handb. Exp. Pharmacol.* **2012**, 307−324.
(9) Shipp, E.; Hsieh-Wilson, L. Profiling the Sulfation Specificities of Glycosaminoglycan Interactions with Growth Factors and Chemotactic Proteins Using Microarrays. *Chem. Biol.* **2007**, *14*, 195−208.
(10) Kim, S.-H.; Turnbull, J.; Guimond, S. Extracellular Matrix and Cell Signalling: The Dynamic Cooperation of Integrin, Proteoglycan and Growth Factor Receptor. *J. Endocrinol.* **2011**, *209*, 139−151.
(11) Linhardt, R.; Toida, T. Role of Glycosaminoglycans in Cellular Communication. *Acc. Chem. Res.* **2004**, *37*, 431−438.
(12) Lord, M.; Whitelock, J. Bioengineered Heparin: Is There a Future for This Form of the Successful Therapeutic? *Bioengineered* **2014**, *5*, 222−226.
(13) Schnabelrauch, M.; Scharnweber, D.; Schiller, J. Sulfated Glycosaminoglycans as Promising Artificial Extracellular Matrix Components To Improve the Regeneration of Tissues. *Curr. Med. Chem.* **2013**, *20*, 2501−2523.
(14) Scott, R.; Panitch, A. Glycosaminoglycans in Biomedicine. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **2013**, *5*, 388−398.
(15) DeMarco, M.; Woods, R. Structural Glycobiology: A Game of Snakes and Ladders. *Glycobiology* **2008**, *18*, 426−440.
(16) Samsonov, S.; Teyra, J.; Pisabarro, M. T. Docking Glycosaminoglycans to Proteins: Analysis of Solvent Inclusion. *J. Comput.-Aided Mol. Des.* **2011**, *25*, 477−489.
(17) Samsonov, S.; Teyra, J.; Anders, G.; Pisabarro, M. T. Analysis of the Impact of Solvent on Contacts Prediction in Proteins. *BMC Struct. Biol.* **2009**, *9*, 22.
(18) Faham, S.; Hileman, R.; Fromm, J.; Linhardt, R.; Rees, D. Heparin Structure and Interactions with Basic Fibroblast Growth Factor. *Science* **1996**, *271*, 1116−1120.
(19) Li, Z.; Kienetz, M.; Cherney, M.; James, M.; Brömme, D. The Crystal and Molecular Structures of a Cathepsin K:Chondroitin Sulfate Complex. *J. Mol. Biol.* **2008**, *383*, 78−91.
(20) Mulloy, B.; Forster, M.; Jones, C.; Davies, D. N.m.r. and Molecular-Modelling Studies of the Solution Conformation of Heparin. *Biochem. J.* **1993**, *293* (Pt3), 849−858.
(21) Hricovíni, M. B3LYP/6-311++G** Study of Structure and Spin-Spin Coupling Constant in Methyl 2-O-Sulfo-Alpha-L-Iduronate. *Carbohydr. Res.* **2006**, *341*, 2575−2580.
(22) Krylov, D.; Mikhailenko, I.; Vinson, C. A Thermodynamic Scale for Leucine Zipper Stability and Dimerization Specificity: E and G Interhelical Interactions. *EMBO J.* **1994**, *13*, 2849−2861.
(23) Mikhailov, D.; Mayo, K.; Vlahov, I.; Toida, T.; Pervin, A.; Linhardt, R. NMR Solution Conformation of Heparin-Derived Tetrasaccharide. *Biochem. J.* **1996**, *318* (Pt 1), 93−102.
(24) Gandhi, N.; Mancera, R. Can Current Force Fields Reproduce Ring Puckering in 2-O-Sulfo-α-L-Iduronic Acid? A Molecular Dynamics Simulation Study. *Carbohydr. Res.* **2010**, *345*, 689−695.
(25) Sattelle, B.; Hansen, S.; Gardiner, J.; Almond, A. Free Energy Landscapes of Iduronic Acid and Related Monosaccharides. *J. Am. Chem. Soc.* **2010**, *132*, 13132−13134.
(26) Nieto, L.; Canales, Á.; Giménez-Gallego, G.; Nieto, P.; Jiménez-Barbero, J. Conformational Selection of the AGA*IA(M) Heparin Pentasaccharide When Bound to the Fibroblast Growth Factor Receptor. *Chem. Weinh. Bergstr. Ger.* **2011**, *17*, 11204−11209.
(27) Samsonov, S. A.; Pisabarro, M. T. Importance of IdoA and IdoA(2S) Ring Conformations in Computational Studies of Glycosaminoglycan−Protein Interactions. *Carbohydr. Res.* **2013**, *381*, 133−137.

(28) Mulloy, B. The Specificity of Interactions Between Proteins and Sulfated Polysaccharides. *An. Acad. Bras. Ciênc.* **2005**, *77*, 651−664.
(29) Pichert, A.; Samsonov, S.; Theisgen, S.; Thomas, L.; Baumann, L.; Schiller, J.; Beck-Sickinger, A.; Huster, D.; Pisabarro, M. T. Characterization of the Interaction of Interleukin-8 with Hyaluronan, Chondroitin Sulfate, Dermatan Sulfate and their Sulfated Derivatives by Spectroscopy and Molecular Modeling. *Glycobiology* **2012**, *22*, 134−145.
(30) Van der Smissen, A.; Hintze, V.; Scharnweber, D.; Moeller, S.; Schnabelrauch, M.; Majok, A.; Simon, J.; Anderegg, U. Growth Promoting Substrates for Human Dermal Fibroblasts Provided by Artificial Extracellular Matrices Composed of Collagen I and Sulfated Glycosaminoglycans. *Biomaterials* **2011**, *32*, 8938−8946.
(31) Hintze, V.; Samsonov, S.; Anselmi, M.; Moeller, S.; Becher, J.; Schnabelrauch, M.; Scharnweber, D.; Pisabarro, M. T. Sulfated Glycosaminoglycans Exploit the Conformational Plasticity of Bone Morphogenetic Protein-2 (BMP-2) and Alter the Interaction Profile with Its Receptor. *Biomacromolecules* **2014**, *15*, 3083−3092.
(32) Gandhi, N.; Coombe, D.; Mancera, R. Platelet Endothelial Cell Adhesion Molecule 1 (PECAM-1) and Its Interactions with Glycosaminoglycans: 1. Molecular Modeling Studies. *Biochemistry* **2008**, *47*, 4851−4862.
(33) Gandhi, N.; Mancera, R. Free Energy Calculations of Glycosaminoglycan−Protein Interactions. *Glycobiology* **2009**, *19*, 1103−1115.
(34) Gandhi, N.; Mancera, R. Molecular Dynamics Simulations of CXCL-8 and Its Interactions with a Receptor Peptide, Heparin Fragments, and Sulfated Linked Cyclitols. *J. Chem. Inf. Model.* **2011**, *51*, 335−358.
(35) Agostino, M.; Gandhi, N.; Mancera, R. Development and Application of Site Mapping Methods for the Design of Glycosaminoglycans. *Glycobiology* **2014**, *24*, 840−851.
(36) Munakata, H.; Takagaki, K.; Majima, M.; Endo, M. Interaction Between Collagens and Glycosaminoglycans Investigated Using a Surface Plasmon Resonance Biosensor. *Glycobiology* **1999**, *9*, 1023−1027.
(37) Ho, L.; Harris, A.; Tanioka, H.; Yagi, N.; Kinoshita, S.; Caterson, B.; Quantock, A.; Young, R.; Meek, K. A Comparison of Glycosaminoglycan Distributions, Keratan Sulphate Sulphation Patterns and Collagen Fibril Architecture from Central to Peripheral Regions of the Bovine Cornea. *Matrix Biol.* **2014**, *38*, 59−68.
(38) Raspanti, M.; Viola, M.; Forlino, A.; Tenni, R.; Gruppi, C.; Tira, M. E. Glycosaminoglycans Show a Specific Periodic Interaction with Type I Collagen Fibrils. *J. Struct. Biol.* **2008**, *164*, 134−139.
(39) Marrink, S.; Risselada, J.; Yefimov, S.; Tieleman, P.; de Vries, A. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(40) Korkut, A.; Hendrickson, W. A Force Field for Virtual Atom Molecular Mechanics of Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15667−15672.
(41) Baaden, M.; Marrink, S. Coarse-Grain Modelling of Protein−Protein Interactions. *Curr. Opin. Struct. Biol.* **2013**, *23*, 878−886.
(42) Monticelli, L.; Kandasamy, S.; Periole, X.; Larson, R.; Tieleman, P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(43) Gu, J.; Bai, F.; Li, H.; Wang, X. A Generic Force Field for Protein Coarse-Grained Molecular Dynamics Simulation. *Int. J. Mol. Sci.* **2012**, *13*, 14451−14469.
(44) Liwo, A.; Baranowski, M.; Czaplewski, C.; Gołaś, E.; He, Y.; Jagieła, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M.; Niadzvedtski, A.; Ołdziej, S.; Scheraga, H.; Sieradzan, A.; Ślusarz, R.; Wirecki, T.; Yin, Y.; Zaborowski, B. A Unified Coarse-Grained Model of Biological Macromolecules Based on Mean-Field Multipole−Multipole Interactions. *J. Mol. Model.* **2014**, *20*, 1−15.
(45) Marrink, S.; de Vries, A.; Mark, A. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2003**, *108*, 750−760.
(46) López, C.; Rzepiela, A.; de Vries, A.; Dijkhuizen, L.; Hünenberger, P.; Marrink, S. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195−3210.
(47) Vicatos, S.; Rychkova, A.; Mukherjee, S.; Warshel, A. An Effective Coarse-Grained Model for Biological Simulations: Recent Refinements and Validations. *Proteins* **2014**, *82*, 1168−1185.
(48) Spiga, E.; Alemani, D.; Degiacomi, M.; Cascella, M.; Peraro, M. Electrostatic-Consistent Coarse-Grained Potentials for Molecular Simulations of Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 3515−3526.
(49) Bathe, M.; Rutledge, G.; Grodzinsky, A.; Tidor, B. A Coarse-Grained Molecular Model for Glycosaminoglycans: Application to Chondroitin, Chondroitin Sulfate, and Hyaluronic Acid. *Biophys. J.* **2005**, *88*, 3870−3887.
(50) Chemical Computing Group, Inc., Montreal, Canada, 2006.
(51) Case, D.; Darden, T.; Cheatham, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A.; Kolossváry, I.; Wong, K.; Paesani, F.; Vanicek, J.; Wolf, R.; Liu, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.; Cui, G.; Roe, D.; Mathews, D.; Seetin, M.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. 2012. AMBER 12, University of California, San Francisco.
(52) Huige, C.; Altona, C. Force Field Parameters for Sulfates and Sulfamates Based on Ab Initio Calculations: Extensions of AMBER and CHARMm Fields. *J. Comput. Chem.* **1995**, *16*, 56−79.
(53) Onufriev, A.; Case, D.; Bashford, D. Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect. *J. Comput. Chem.* **2002**, *23* (1304), 1297.
(54) Tschöp, W.; Kremer, K.; Batoulis, J.; Bürger, T.; Hahn, O. *Acta Polym.* **1998**, *49*, 61−74.
(55) *R-package. R: A Language and Environment for Statistical Computing*; R. Development Core Team: Vienna, Austria, 2006.
(56) Kerzmann, A.; Neumann, D.; Kohlbacher, O. SLICK−Scoring and Energy Functions for Protein−Carbohydrate Interactions. *J. Chem. Inf. Model.* **2006**, *46*, 1635−1642.
(57) Jin, L.; Hricovíni, M.; Deakin, J.; Lyon, M.; Uhrín, D. Residual Dipolar Coupling Investigation of a Heparin Tetrasaccharide Confirms the Limited Effect of Flexibility of the Iduronic Acid on the Molecular Shape of Heparin. *Glycobiology* **2009**, *19*, 1185−1196.
(58) Khan, S.; Gor, J.; Mulloy, B.; Perkins, S. J. Semi-Rigid Solution Structures of Heparin by Constrained X-ray Scattering Modelling: New Insight into Heparin-Protein Complexes. *J. Mol. Biol.* **2010**, *395*, 504−521.
(59) Pavlov, G.; Finet, S.; Tatarenko, K.; Korneeva, E.; Ebel, C. Conformation of Heparin Studied with Macromolecular Hydrodynamic Methods and X-ray Scattering. *Eur. Biophys. J.* **2003**, *32*, 437−449.
(60) Horkay, F.; Basser, P. J.; Londono, D. J.; Hecht, A. M.; Geissler, E. Ions in Hyaluronic Acid Solutions. *J. Chem. Phys.* **2009**, *131*, 184902−184908.
(61) Kolesnikov, A. L.; Budkov, Y. A.; Nogovitsyn, E. A. Coarse-Grained Model of Glycosaminoglycans in Aqueous Salt Solutions. A Field-Theoretical Approach. *J. Phys. Chem. B* **2014**, *118*, 13037−13049.