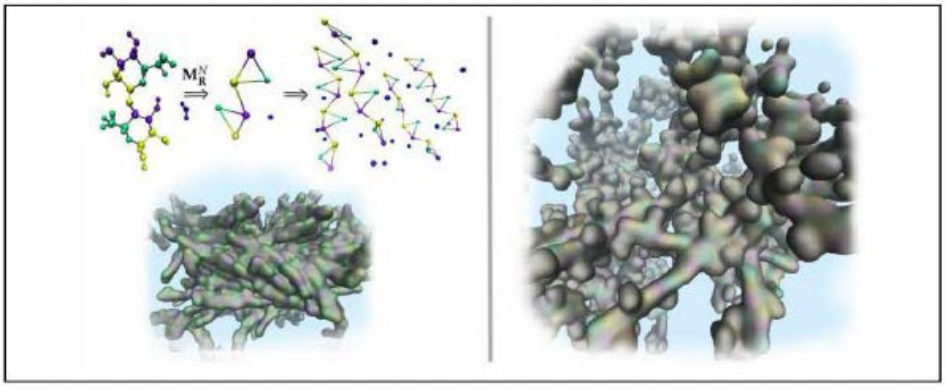
Published in final edited form as:
Sauter, J., & Grafmüller, A. (2017). Procedure for Transferable Coarse-Grained Models of Aqueous Polysaccharides. Journal of Chemical Theory and Computation. 13(1), 223-236.
doi: 10.1021/acs.jctc.6b00613

Procedure for Transferable Coarse-Grained Models of Aqueous Polysaccharides

Abstract

We present a procedure to obtain Coarse-Grained (CG) models for aqueous polysaccharide solutions that are transferable over different degrees of polymerization and different polysaccharide concentrations based on atomistic Molecular Dynamics (MD) simulations. This is achieved by a hybrid procedure combining Boltzmann Inversion (BI) and the Multiscale Coarse-Graining (MS-CG) method. In order to overcome problems that have been previously reported with this approach, namely differences in the aggregation behavior and the end to end distance between the atomistic reference simulation and the coarse-grained simulation, we employ a separation-ansatz and explicit 1–3 and 1–4 nonbonded intramolecular interactions. This allows the use of the model for long polysaccharides. We demonstrate the transferability over both concentration and degrees of polymerization, evaluate the scope for which the coarse-grained model can be applied, and then present a scheme to extend the concentration transferability. In addition, we show that the procedure can be applied to generate a transferable implicit solvent model and demonstrate that it can be used for different atomistic force fields (FFs) as well. The procedure is then applied to derive a coarse- grained model of different hemicellulose polysaccharides. The resulting model is used to demonstrate that branching with monomer side-chains significantly increases the water uptake capacity of the molecules in comparison to linear polysaccharides which is consistent with experimental results.



Procedure for Transferable Coarse-Grained Models of Aqueous Polysaccharides

Jörg Sauter and Andrea Grafmüller*

Theory and Bio-Systems, Max Planck Institute of Colloids and Interfaces, Potsdam, Germany

E-mail: andrea.grafmueller@mpikg.mpg.de

Abstract

We present a procedure to obtain coarse-grained models for aqueous polysaccharide solutions that are transferable over different degrees of polymerization and different polysaccharide concentrations based on atomistic Molecular Dynamics (MD) simulations. This is achieved by a hybrid procedure combining Boltzmann Inversion (BI) and the Multiscale Coarse-Graining (MS-CG) method. In order to overcome problems that have been previously reported with this approach, namely differences in the aggregation behavior and the end to end distance between the atomistic reference simulation and the coarse-grained simulation, we employ a separation-ansatz and explicit 1-3 and 1-4 non-bonded intra-molecular interactions. This allows the use of the model for long polysaccharides. We demonstrate the transferability over both concentration and degrees of polymerization and evaluate the scope for which the coarse-grained model can be applied, then present a scheme to extend the concentration transferability. In addition, we show that the procedure can be applied to generate a transferable implicit solvent model and demonstrate that it can be used for different atomistic force fields (FFs) as well. The procedure is then applied to derive a coarse-grained model of different hemicellulose polysaccharides. The resulting model is used to demonstrate that

branching with monomer side-chains significantly increases the water uptake capacity of the molecules in comparison to linear polysaccharides which is consistent with experimental results.

1 Introduction

Polysaccharides are among the most important molecules in biological systems. They play a fundamental role in immune recognition,[1] energy transfer and storage,[2] and as constituents of structuring materials.[3–5] There has been a continuously growing interest in the study of these systems by Molecular Dynamics simulations and improved simulation methods as well as better FFs are an active research field. While properties that involve only small saccharide systems and short time-scales can be studied using MD simulations with atomistic resolution, many important phenomena exceed the limitations imposed by atomistic MD simulations.

A particularly interesting simulation task is to explore the possibilities for carbohydrates in bio-inspired materials such as drug delivery systems. However, the relevant natural materials typically involve high degrees of polymerization making these systems challenging to study using atomistic simulations as they involve both large system sizes and slow dynamics. An interesting example of a natural functional polysaccharide material, is the direction dependent swelling of the primary plant cell wall that plays an important role in a variety of plant systems such as, for instance, in the opening of ice plant seed capsules[6] or pine cones.[7] The primary plant cell wall consists predominantly of polysaccharides in the form of crystalline cellulose microfibrils embedded in an amorphous matrix of hemicellulose and pectin.[3] These systems involve high polysaccharide concentrations and degrees of polymerization. Hemicellulose molecules reach degrees of polymerization (DP) up to 3500[8] and are in general shorter than cellulose which can possess a DP of up to 15000.[9] The primary plant cell wall is highly stress resistant while it is at the same time sensitive to hydration and can significantly swell as a function of water content,[4,5] making the polysaccharide-water interactions a key factor. For hemicellulose polysaccharides it was suggested previoulsy,

that monosaccharide side-chains increase the water affinity of branched polysaccharides in comparison to linear polysaccharides.[10] However, this increased water affinity for monomer side-chains is not large enough to prevent self-association, since gel-formation[11] occurs.

A similar example are proteoglycans, which reside in the extracellular matrix in animal tissues. These molecules reach masses of up to 400 kD, most of which is given by the glycosaminoglycan carbohydrates. These molecules fill the space between cells or form a stable matrix with Hyaluronan in cartilage, capable of absorbing high compressive loads by water desorption and resorption[12]

To make simulations of such complex systems feasible, a possibility is to extend the accessible simulation scales by the use of Coarse-Grained (CG) models that preserve as much of the predictive capabilities of the atomistic model as possible, while reducing the computational cost significantly.

In the past, different strategies to obtain CG-FFs of polysaccharides have been pursued. Besides CG-FFs that have been empirically parameterized for a limited amount of molecules, such as the popular MARTINI model,[13,14] there has been an increasing interest in the development of bottom up CG-FFs based on atomistic FFs. This approach has the advantage that, in principle, the full flexibility of atomistic FFs can be used, i.e. a suitable FF can be chosen at the atomistic scale to accurately describe the molecules and properties of interest. Using an appropriate method, a CG-FF is derived to extend the specific predictive power of the FF to a larger scale.

The most common methods to obtain the interactions between the CG sites are (iterative-) Boltzmann Inversion,[15] Inverse Monte Carlo,[16] and Multiscale Coarse-Graining[17] also known as Force Matching. In addition, a method that aims at the minimization of the relative entropy between the atomistic and the CG system has been introduced[18] recently. While BI and IMC seek to reproduce structural properties such as the Radial Distribution Function (RDF), the MS-CG method aims effectively at reproducing the forces of the atomistic system in the CG system.

Besides the close reproduction of atomistic properties, a key challenge in coarse-graining is to obtain a model that is transferable over different thermodynamic conditions or over different system compositions, i.e. a model that can be used for different solute concentrations and for different degrees of polymerization of the solute molecules. Because the CG-FFs represent potentials of mean force they depend on the system properties. However, for the application to polysaccharides, transferability over different degrees of polymerization is essential to construct the large systems of interest from atomistically accessible systems.

Transferabiliy with respect to system composition can not be expected for the structural methods that reproduce properties which significantly change with system composition. In particular, this is the case for the RDFs of non-bonded interactions at different polysaccharide concentrations.[19] On the other hand, the MS-CG method attempts to reproduce the mean force of the atomistic system in the CG system and can therefore be expected to be less system dependent, because the mean force is a more fundamental property of the system. In fact, it has been shown by Wagner *et al.*[20] that the CG-FF obtained by the MS-CG method can be directly linked to the underlying atomistic FF parameters within certain limitations. More precisely, it was found that it is possible to approximate the CG-FF that is obtained when atomistic parameters are slightly altered, provided that many-body components are negligible. The transferability of potentials over different thermodynamic condition obtained by MS-CG has been explored[21] in detail, and it has been shown that a modified MS-CG method can yield thermodynamic and composition transferable potentials for ionic liquids.[22–24]

Building a reliable CG-FF from atomistic simulations for aqueous polysaccharides has proven to be difficult in the past. Molinero *et al.*[25] have introduced the M3B model which uses three sites for a glucose molecule. A hybrid approach combining MS-CG (non-bonded interactions) and BI (bonded interactions) was introduced by Liu *et al.*[26] for glucose monomers in solution and for polysaccharides up to a Degree of Polymerization (DP) of fourteen, however, using only one polysaccharide in water. It was shown that, for α-D-glucose, the CG

model essentially preserves the average density of molecules at a certain distance from a given molecule and the relative orientation between two molecules. Their investigations indicated that, for monomers, a modified hybdrid approach using a correction term for the pressure can reproduce most thermodynamic quantities reasonable well and can produce CG-FFs that are transferable over different thermodynamic conditions.

Furthermore, different mapping schemes using up to four sites per monomer have been studied[27] for β-D-glucose. Hynninen *et al.*[28] used the hybrid method for multiple polysaccharides in solution up to DP = 4 and demonstrated the necessity to use a hybrid procedure, since using the MS-CG approach for bonded interactions lead to un-physical conformations. In addition, starting with DP = 2, aggregation problems became apparent and problems with the end to end distance at DP = 4 have been reported.

In this article, we show that applying the MS-CG method to the solvent interactions and solute-solute interactions separately results in a CG-FF that leads to excellent agreement with the atomistic reference simulation. Using explicit 1-3 and 1-4 interactions obtained by BI, we demonstrate that for long polysaccharides with DP = 16 the CG-FF is able to reproduce the end to end distance and the radius of gyration with high accuracy. We show that, within a certain range, the CG-FF can be used at different polysaccharide concentrations and we present a system dependent scheme to extend the concentration range within which the model can be applied. Furthermore, we demonstrate that the CG-FF can be transferred over different degrees of polymerization. Therefore, the method presented should allow the construction of polymers of indefinite length using the potentials obtained at DP = 16. In addition, we demonstrate that the procedure can also be used to obtain an implicit solvent model which is transferable to different solute concentrations, similarly to the explicit solvent CG-FF. The CHARMM36 FF[29,30] with TIP3P[31] water was used as the standard FF in this paper to ensure continuity with previous studies in this field. On the other hand, we have shown before[19] that the optimal FF to study aqueous polysaccharides, with respect to the solution properties, is the GLYCAM06[32] FF with TIP5P[33] water. Therefore, we further test

the performance of the coarse-graining procedure with two variants of that FF.

Finally, we use the developed procedure to derive a CG model of hemicellulose polysaccharides to investigate the water uptake of aggregates of branched polysaccharides with monomer side-chains in water in comparison to aggregates of linear polysaccharides.

2 Procedure: Methods

2.1 Atomistic MD-Simulation

The following four types of atomistic systems were simulated for the coarse-gaining and for reference purposes.

1. A system with 64 1-4 linked β-D-glucose polysaccharides with DP = 16 using (a) 36, (b) 18, (c) 54, and (d) 222 water molecules per polysaccharide.

2. A system with a single 1-4 linked β-D-glucose polysaccharide with DP = 16 using 19673 water molecules.

3. A system with 64 1-4 linked β-D-glucose polysaccharides with DP = 8 using 18 water molecules per polysaccharide.

4. A system with 64 β-D-glucose monomers using 14 water molecules per monomer. For this system we used (a) the GLYCAM06 FF with TIP5P water, (b) the GLYCAM06$^{\text{TIP5P}}_{\text{OSMOr14}}$ FF,[34] and (c) the CHARMM36 TIP3P FF.

Systems 1 (a) to (d) are used to investigate concentration transferability. The purpose of system 2 is to compare the atomistic and coarse-grained end to end distance as well as the radius of gyration. System 3 is used to demonstrate the DP transferability. Systems 4 (a) to (c) are used to evaluate FF dependencies. For systems 1-3 the CHARMM36 FF with TIP3P water was used and the system was set up directly in GROMACS after the CHARMM36 parameters were converted to GROMACS.

The GLYCAM06 topology and coordinate files were generated in Amber 12[35] using pdb files from the GLYCAM builder.[36] Afterwards, they were converted to the GROMACS format using the glycam2gmx script,[37,38] followed by hydration in GROMACS. All computations were performed in GROMACS 4.6.7.[39] Covalent bonds involving hydrogen atoms were constrained with LINCS[40] while the water models use Settle[41] constraints. Note that in this setup no constraints exist between atoms that were mapped onto different CG-sites. Electrostatic interactions were treated with the Particle Mesh Ewald method[42] using a 1.4 nm cut-off. For all reruns (see below) the Reaction-Field[43] method was used with the same cut-off. A 1.4 nm cut-off was also used for the Lennard-Jones interactions. The GROMACS dispersion correction was used for the energy. We used the Leap-Frog integrator[44] with a 2 fs time-step size and the temperature was controlled at T=298.15K using the Nosé-Hoover thermostat.[45,46] An output frame was written every 2ps. Energy minimization and equilibration were performed using standard protocols. As usual with the MS-CG method, the simulations were performed in the NVT ensemble. The volume was obtained as the average of 50ns NPT simulations with pressure set to 1 bar with the Parinello-Rahmann barostat[47,48] and additional dispersion correction for the pressure. All atomistic NVT simulations used a 100ns trajectory with 50000 frames after a total of 400 ns equilibration. The long equilibration time is required due to the slow dynamic of the systems containing polysaccharides high DP. In an extended atomistic simulation using 200 ns and 1000000 frames, no significant differences in the potentials derived from the trajectory were detectable, thus we assume that 100ns offer sufficient sampling. For visualization Visual Molecular Dynamics (VMD)[49] was used.

2.2 Coarse-Graining Methods

The non-bonded CG interactions were obtained by the MS-CG method i.e. the forces $\mathbf{F}_I$ on the CG site I in a given CG configuration $\mathbf{R}^N$ are assumed to be given by

$$\mathbf{F}_I(\mathbf{R}^N) = \langle \boldsymbol{\mathcal{F}}_I(\mathbf{r}^n) \rangle_{\mathbf{R}^N} \tag{1}$$

where $\boldsymbol{\mathcal{F}}_I$ is a weighted linear combination of the forces acting on the atoms that are mapped to the CG-site I, depending on the respective atomistic configuration $\mathbf{r}^n$. Here, n is the number of atoms and N the number of CG-sites. The angular brackets denote a filtered average over the atomistic reference trajectory of all configurations $\mathbf{r^n}$ that map to a given CG configuration $\mathbf{R^N}$. The complete CG-FF is obtained by a least squares approximation or using L_2-regularization where pairwise additive, radially symmetric, forces between the CG-sites are assumed. A more detailed theoretical description is given by Noid *et. al.*[50]

Bonds, angles, dihedrals, and explicit short range intra-molecular (1-3 and 1-4) non-bonded interactions are obtained by BI i.e. a potential $U(\mathbf{R}^N)$ for an interaction is given by

$$U(\mathbf{R}^N) = -k_B T \ln\left(p(\mathbf{R}^N)\right) \tag{2}$$

were $p(\mathbf{R}^N)$ is the probability of a given CG configuration $\mathbf{R}^N$ obtained from an atomistic reference trajectory and k_B denotes Boltzmann's constant.

For BI we used VOTCA[51,52] 1.2.4 and the MS-CG was performed with the MSCGFM 1.2 code obtained from the Voth group[53] and used L_2-regularization with $\lambda = 0.25^2$. The code was modified by us to exclude all non-bonded intra-molecular interactions. Standard post-processing of all interactions was performed to remove anomalies from poorly sampled regions and all forces were linearily extended to unsampled regions.

We performed the coarse-graining of the atomistic systems 1 to 4 and simulated the respective CG systems with the derived CG-FF in the NVT-ensemble using the Leap-Frog integrator. The temperature was controlled using the Nosé-Hoover thermostat at T=298.15K.

In the MS-CG method the interaction range was 1.0 nm. The smaller interaction range became necessary because at long ranges the aggregation behavior of the system becomes more sensitive to perturbations. This sensitivity is discussed in detail in Section 3.3.2. No constraints were used and no explicit electrostatic interactions were present. A time step-size of 1 fs was used and each system was equilibrated for 10ns. In principle however, a significantly larger time step-size is possible in the CG system. It should be noted here, that the time-scale in the CG-system does not correspond to the atomistic time.[54]

3 Procedure: Results and Discussion

3.1 Method development

Several important considerations have to be made when devising the hybrid method which are discussed in this section. A structural criterion for the comparison of the CG results to the atomistic reference trajectory is given by the RDF

$$g(r) = \frac{V}{N^2} \left\langle \sum_i \sum_{i \neq j} \delta(r - |\mathbf{r_i} - \mathbf{r_j}|) \right\rangle \tag{3}$$

where $\mathbf{r_i}$ and $\mathbf{r_j}$ are the position of the site i and j respectively, N denotes the total number of sites, V the volume, and $\langle \cdot \rangle$ the ensemble average. A comparison of the RDFs obtained from the initial and final frames of the trajectory revealed only minor deviations for all concentrations indicating that the equilibration, in terms of the RDFs, is sufficient for the atomistic and the CG systems. The RDFs were calculated using 1000 frames,

For the water molecules a one site center of geometry (COG) mapping was used because a COG mapping was found to work best for water.[17] The coordinates for the saccharide CG-sites were obtained using a center of mass (COM) mapping. We use the three site mapping $\mathbf{M_R^N}$ shown in Figure 1 for DP = 4 to map the atomistic coordinates $\mathbf{r}^n$ of the polysaccharide onto CG coordinates $\mathbf{R}^N$. We denote by A_i site A of the i-th monomer in the polysaccharide.

B_i and C_i are defined analogously. The mapping is defined in the same way for other DPs. In contrast to Hynninen *et al.*,[28] we do not treat the terminal sites differently and we do not

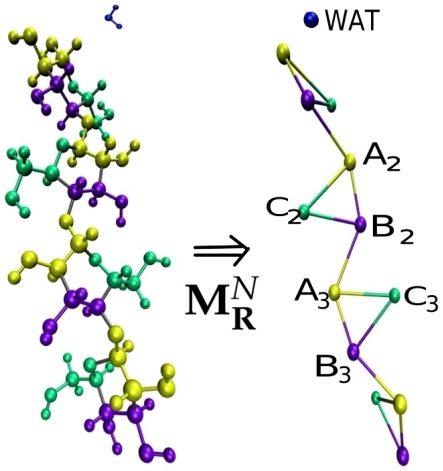


Figure 1: Example of the CG Mapping $\mathbf{M}_\mathbf{R}^\mathbf{N}$ of the atomistic coordinates $\mathbf{r}^n$ onto CG coordinates $\mathbf{R}^N$ for 1-4 linked β-D-glucose with DP = 4.

distinguish between different C sites since the differences obtained from such treatment were negligible. For the definition of the molecule interactions within one monomer three bonds were used. The connection between two monomers involved one (BI-)bond, two (BI-)angles, and three (BI-)dihedrals. The only choice for the bond is B_i - $A_{\mathrm{i}+1}$. The connecting angles C_i - B_i - $A_{\mathrm{i}+1}$ and B_i - $A_{\mathrm{i}+1}$ - $C_{\mathrm{i}+1}$ were chosen and we used the dihedrals $A_{\mathrm{i}+1}$ - B_i - C_i - A_i and B_i - $A_{\mathrm{i}+1}$ - $C_{\mathrm{i}+1}$ - $B_{\mathrm{i}+1}$ as well as $C_{\mathrm{i}+1}$ - $A_{\mathrm{i}+1}$ - B_i - C_i. In addition, two explicit (BI-)1-3 non-bonded intra-molecular interactions connecting the sites A_i - $A_{\mathrm{i}+1}$ as well as B_i - $B_{\mathrm{i}+1}$ were used. Similarly, the (BI-)1-4 interactions C_i - $C_{\mathrm{i}+1}$ as well as A_i - $B_{\mathrm{i}+1}$ were explicitly included.

To ensure a consistent hybrid method, we excluded all bonded interactions directly from the atomistic trajectory using the GROMACS rerun method before the MS-CG method was applied and consequently, these interactions are excluded in the MS-CG procedure as well. Note that this approach is superior by design to subtracting tabulated BI forces because the BI interactions are only an approximation, i.e. they contain no cross-correlations and are interpolated. Therefore, the errors stemming from BI are propagated to the MS-CG method,

when tabulated BI forces are subtracted before the MS-CG method is applied.

As done by Hynninen *et al.*, all intra-molecular non-bonded interaction are excluded from the MS-CG method since otherwise intra-molecular interactions are projected onto the inter-molecular interactions in the CG-FF, which we found to lead to incorrect aggregation behavior. All non-bonded intra-molecular interactions beyond 1-4 are given by the non-bonded inter-molecular solute-solute interactions obtained by MS-CG for the respective system, except for system 2. For system 2, a single polymer, the non-bonded intra-molecular interactions of system 1 were used as the system does not contain inter-molecular solute-solute interactions and sufficiently sampled intra-molecular solute-solute interactions can not be obtained with reasonable computational effort.

In the RDFs shown by Hynninen *et al.*,[28] starting with DP = 2, aggregation problems become apparent. We find that these become more pronounced for longer polysaccharides. For instance discrepancies between the aggregation of the CG and the atomistic system are clearly visible in the RDFs for system 1 (a) shown in Figure 2. The solute-solute RDFs, represented by the A-A RDF, show over-aggregation and consistently solute-solvent RDFs, represented by the A-WAT RDF, show under-aggregation.

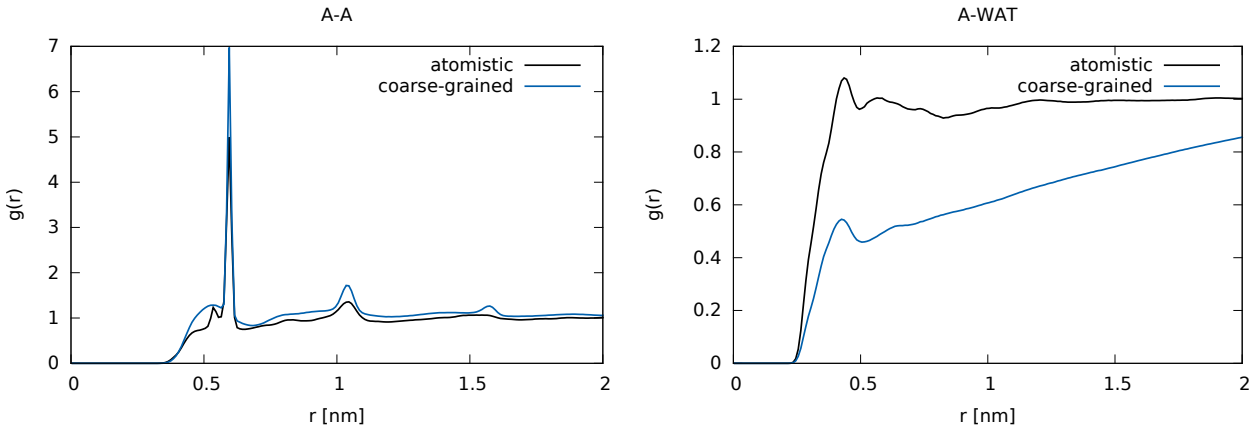


Figure 2: RDFs of the atomistic and the CG simulation with the CG-FF obtained for system 1 (a) without the separation-ansatz.

To solve the problem with the aggregation behavior, we introduce a separation-ansatz, i.e., we separate solvent interactions (that includes solute-solvent as well as solvent-solvent

interactions) and solute-solute interactions in the CG procedure. This was accomplished with two separate versions of atomistic trajectories containing only one class of interaction and two separate MS-CG runs. The atomistic trajectories were obtained using reruns as described above for the exclusion of intra-molecular non-bonded interaction. This separation ensures that the MS-CG-approximation (least squares or L_2-regularization) is not performed over these different interaction classes i.e., force contributions from solute-solvent interactions and solute-solute interactions cannot be mixed in the MS-CG procedure, so that, solute-solvent interactions can not perturb the sensitive solute-solute interactions. We found that the L_2-regularization used in the MS-CG method has only a negligible impact on the aggregation behavior compared to the standard least squares approximation both, with and without the separation-ansatz.

3.2 Evaluation of the CG-FF

To assess the performance of the procedure we first analyze system 1 (a) (DP = 16 and 36 water molecules per polysaccharide). The RDFs of the atomistic and the CG trajectory are shown in Figure 3 for all interactions involving sites of type A. All pairs show good overall agreement between the RDFs obtained from the atomistic and the CG-FF. In particular, it can be seen in the A-WAT RDF that now the aggregation behavior is very well captured by the CG-FF. The RDFs show that the CG interactions effectively capture atomistic behavior that is not directly resolved in the CG model. This is best recognizable in the A-B RDF in Figure 3. The first and, less pronounced, the second bonded neighbor interactions contain two peaks corresponding to different hydroxyl group orientations. However, looking at the details of the solute-solute RDFs of the CG-FF, the peaks corresponding to the bonded first, second, and third neighbor correlations, marked in Figure 3, are slightly more pronounced than in the atomistic simulation. This can be attributed to the BI method which does not take into account that the interactions of the intra-molecular degrees of freedom are influenced by multiple interaction types, i.e. bonds, angles, dihedrals and 1-3 or 1-4 non-

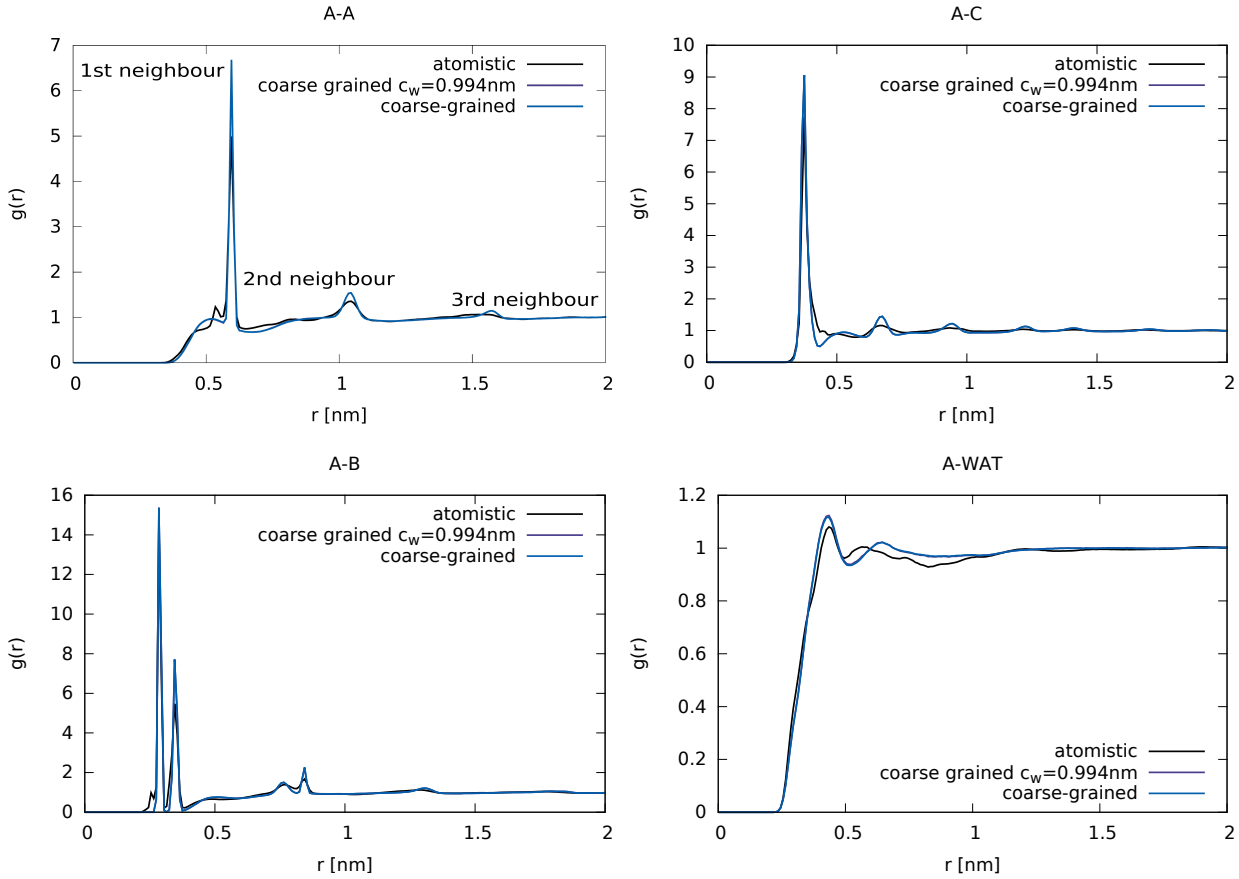


Figure 3: RDFs for all interactions involving sites of type A of system 1 (a). The RDFs for the remaing sites are shown in the Supporting Information. In addition to the results of the atomistic and CG system, the results of the CG system with a cut-off c_w applied to all solvent interactions are shown (see Section 3.3.3). The RDFs contain contributions from both intra- and inter-molecular monomers.

bonded interactions so that the individual interactions have a tendency to be over-restricted.

The effects of the over-represented nearest neighbor correlations also show up in the distribution of the end to end distance and the radius of gyration shown in Figure 4 of a single polymer with DP = 16 in solution, (system 2) Overall, the atomistic probability density functions are very well reproduced by the CG-FF. However, a slight underestimation of the occurrence frequency for short end to end distances and radii of gyration is found in both distributions since the over-restricted bonds translate to a reduced flexibility.Comparable results for the corresponding RDFs, end to end distances, and radii of gyration were also found for the DPs 1,2,4, and 8. For DP = 8 the RDFs as well as the end to end distance and

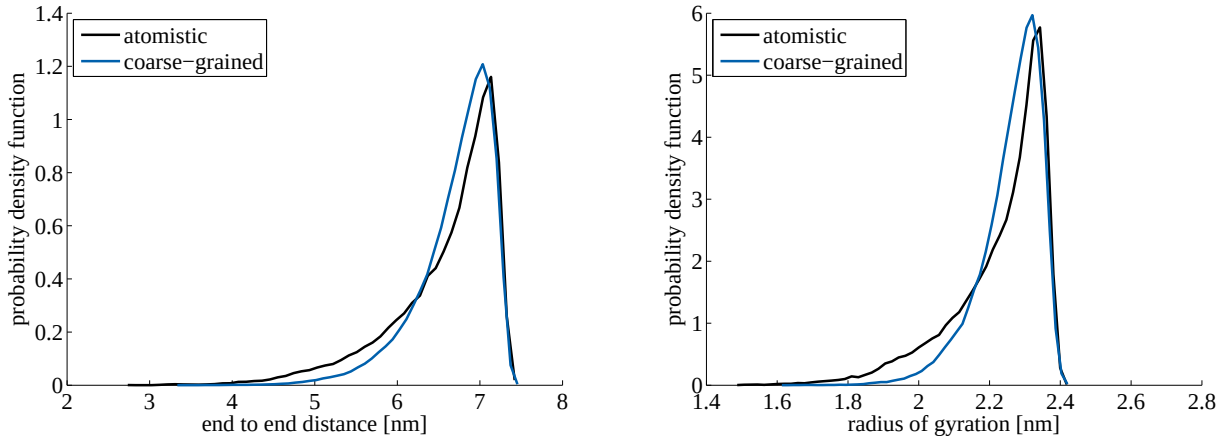


Figure 4: End to end distance and radius of gyration of CG system 2 in comparison to the atomistic reference system.

radius of gyration probability density function are shown in the Supporting Information.

3.3 Concentration Transferability

The aim of the procedure was to derive a CG-FF that is transferable over systems of different composition. To discuss the transferability of the obtained CG-FF over different concentrations, we use the CG interactions (bonded and non-bonded) obtained from system 1 (a) (36 water molecules) for the simulation of the CG systems at higher (system 1 (b), 18 water molecules) and lower solute concentration (system 1 (c), 54 water molecules). The ability of the transferred model to reproduce atomistic data is compared to that of the CG model obtained at the respective native polysaccharide concentration. The results are shown in Figure 5. For the higher concentration, i.e. 18 water molecules per polysaccharide (system 1 (b)), the RDFs shown on the left side of Figure 5 indicate excellent agreement between the atomistic and the CG simulations for both the native CG-FF as well as for the transferred CG-FF. For the lower concentration, the RDFs (Figure 5, right side) show that especially the native model performs poorly and fails to capture the correct aggregation behavior of the system. In the A-WAT RDF a strong under-aggregation becomes apparent at short distances in the region around 0.5 nm. Consistently, over-aggregation is observed in the

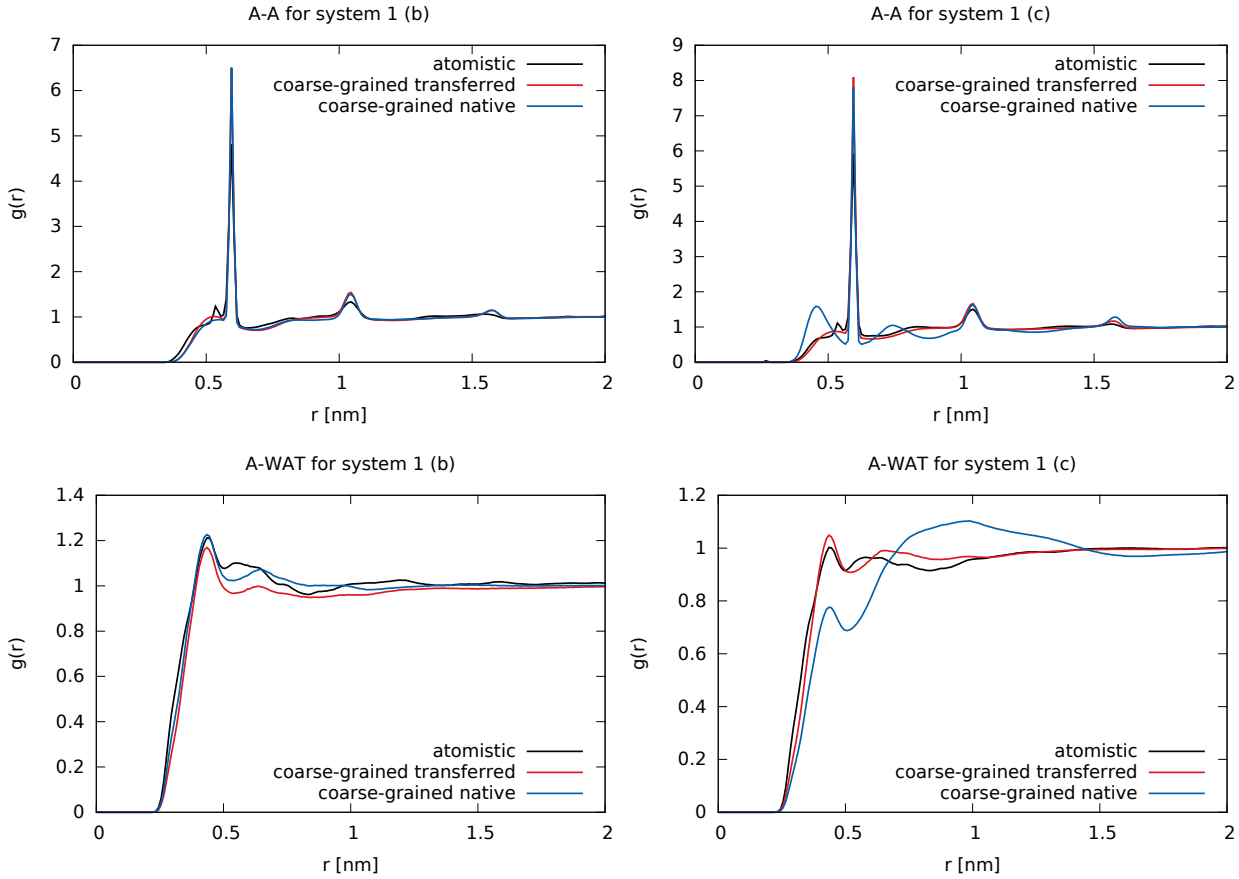


Figure 5: Left side: Example RDFs for system 1 (b). Right side: Example RDFs for the system 1 (c). In addition to the results of the atomistic and CG system, there are also the results of the CG system transferred from system 1 (a).

A-A RDF in the same region. An over-aggregation in the A-WAT RDF at about 1.0 nm partly corrects the overall aggregation behavior so that no phase separation is observed at this concentration.

Several modifications to the method were tested, to see whether the discrepancy in the aggregation behavior of the native CG-model could be corrected: We varied the number of sites and investigated different mapping schemes. Changes in the aggregation behavior were observed for each scheme, however, none was able to capture the solute aggregation correctly. Additional separation of the interactions between solvent-solvent and solute-solvent interactions, similar to the separation of solute-solute interactions and solvent interactions, brought no improvement. The same was true for the additional separation of individual site

types. The reasons behind the failure of the native model will be discussed in the context of the limits of the transferability, investigated in the next Sections.

Surprisingly, the transferred model performs much better at capturing the correct aggregation trends and again leads to very good overall agreement with the atomistic reference simulation. Only a minor trend to over-aggregation is found in the A-WAT distribution. The good performance of the transferred potentials for both concentrations suggests that it is in principle possible to transfer the CG-FF to different system conditions, provided that the native concentration is high enough to correctly capture the aggregation behavior of the solution. The slightly too strong solute-solute aggregation of the transferred CG-FF at lower concentration suggests that transferability may have a lower limit for solute concentrations to which the model can be applied. To investigate further, we now compare both, the native and the transferred CG-FF at a very low solute concentration where the described effects can be expected to be magnified.

3.3.1 Very Low Concentrations

We now apply the coarse-graining to system 1(d) in which the solute concentration is further reduced to 222 water molecules per solute molecule, and test the performance of the transferred potentials for the same conditions. The results shown in Figure 6 (left side) reveal that the overall aggregation behavior is not well captured, by either the native or the transferred CG-FF. The native model leads to over-aggregation of the solute molecules at all distances. In contrast to the previous case at low concentration (system 1 (c)), this effect is not compensated at longer ranges and leads to phase separation in the CG-system. Since the aggregation problems appear more and more pronounced the lower the concentration, it is a fair assumption that solvent interactions are involved in the creation of these artefacts. In fact, it is known that the water-water interaction obtained for a one-site water model reproduces the water-water RDF relatively poorly. This can be attributed to many-body effects and has been corrected by the use of (short range) three-body interactions[55] or itera-

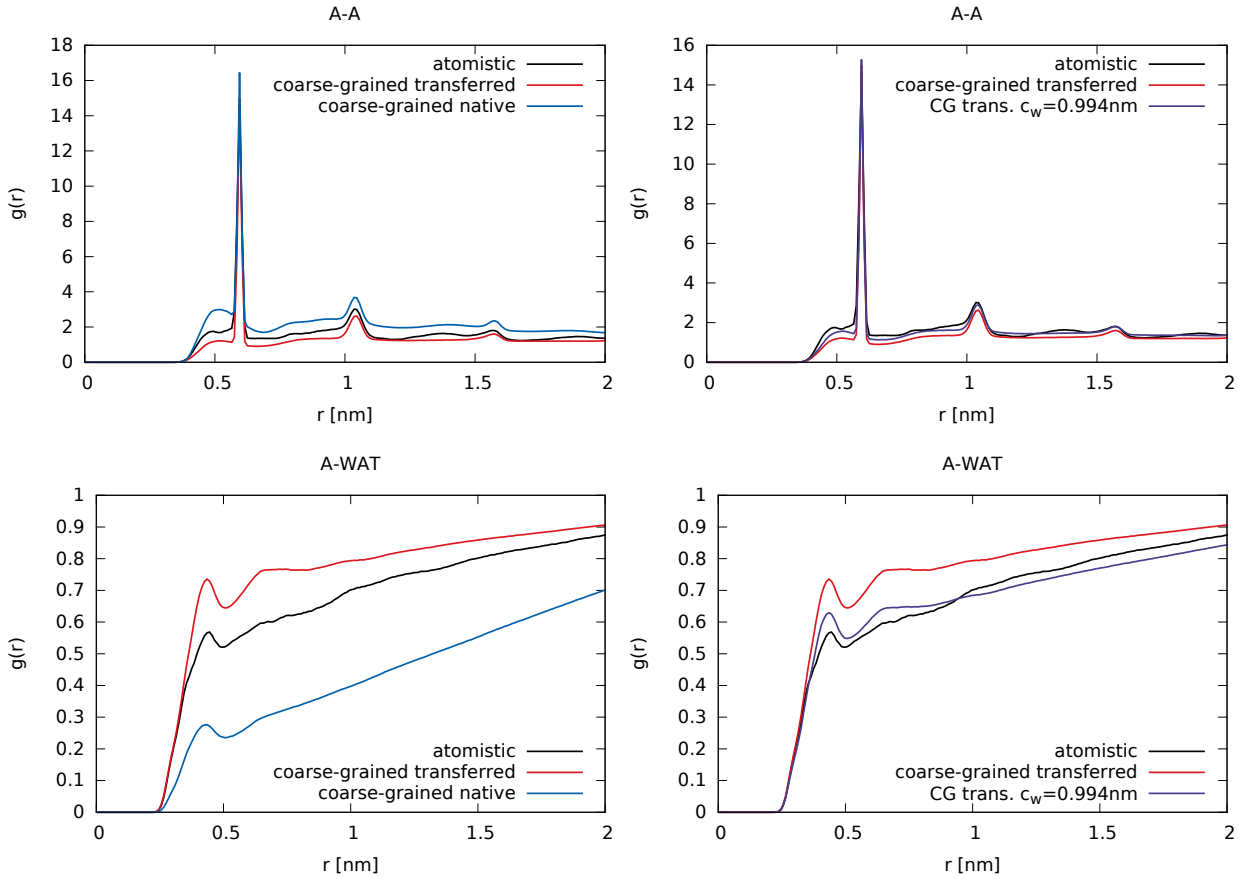


Figure 6: Left side: Transferred and native CG RDFs in comparison to the atomistic RDFs. Right side: Transferred CG RDFs and CG RDFs obtained with a cut-off c_w applied to all solvent interactions compared to the atomistic RDFs.

tive methods,[56,57] however, reducing the computational efficiency. Similar many-body effects can introduce severe problems for solute-solvent systems as they affect the sensitive balance between solute-solvent and solute-solute interactions. In addition, many-body effects of the solute interactions may also become relevant and affect the solute-solvent interactions. Such effects become especially relevant for strong solute interactions, as discuss in Section 3.6, in the context of the FF dependence of the method.

In contrast, the transferred model, which still performed reasonably well for the low solute concentration, now shows significant over-aggregation of the A-WAT RDF shown in Figure 6 (left side), indicating under-aggregation of the solute molecules. To understand the decreased performance of the transferred model, consider that at such low solute concentra-

tions small long range errors arising from many-body effects, or regularization become more relevant, whereas at higher concentrations the systems are dominated by the significantly larger short range interactions. To illustrate the larger relevance of short range interaction in dense systems, the average CG forces of the solute-solute, solute-solvent, and solvent-solvent interactions from system 1 (a) are shown in Figure 7. It can be seen that for the solute-solute interactions there is a pronounced minimum at about 0.6 nm. The minimum for the solute-solvent interactions is wider and shallower at about 0.5 nm and the solvent-solvent interactions have an even shallower minimum at this distance, but possess another deep short range minimum corresponding to the first hydration shell. Figure 7 also reveals that for all interactions, the net long range forces are attractive and that the solute-solute interactions are significantly stronger than the other interaction classes.

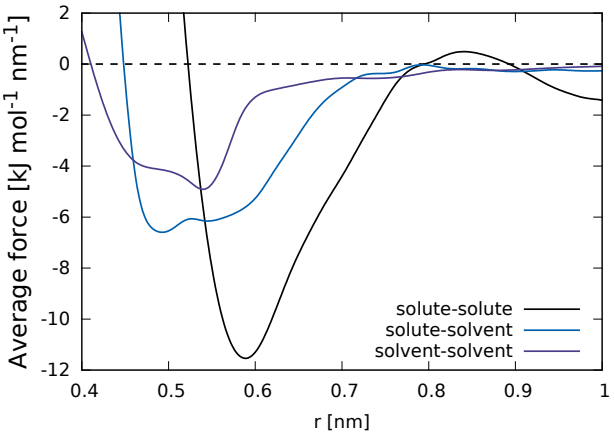


Figure 7: Average CG forces of all solute-solute and all solute-solvent interactions obtained from system 1 (a). In addition, the solvent-solvent CG forces are shown, however here, the first minimum is not shown.

3.3.2 Sensitivity Analysis

In order to better understand the dependence of the aggregation behavior on the long range contributions of the forces we performed a sensitivity analysis of the CG-FFs obtained from system 1 (a) and (d) by separately adding a well defined perturbation to the long range interactions of each interaction class. This perturbation consists of a small constant force

of ± 0.5 kJ mol^{-1}nm^{-1}added to all interactions within an interaction class in the range of 1.0 nm to 1.05 nm. The interactions of the other classes remain unchanged. Analysis of the RDFs reveals that at both concentrations perturbation of the solvent-solvent interactions have virtually no effect. For high concentration (system 1 (a)) the results are also very stable with respect to perturbations of solute-solvent interactions and only a minor variations are observed for perturbations of the solute-solute interactions. At very low concentration (system 1 (d)) on the other hand, there is virtually no solute-solute sensitivity but a profound dependence on the solute-solvent interactions.

In light of this analysis it can be understood why the native model fails at higher concentrations than the transferred model. Due to the increased sensitivity at low concentrations the long range solute-solvent interactions start perturbing the systems. However the concentration is still sufficiently high, and therefore the system is still dominated by short range interactions. This explains the partial, however insufficient, compensation of long range errors by short and medium range interactions. For even lower concentrations the solute-solvent long range sensitivity becomes too pronounced and begins to dominate the overall aggregation behavior. Since the highly concentrated system has a higher density the short range forces are slightly larger than for the very low concentration system. That means the transferred model, from the high concentration, is in general more dominated by short range interactions and is therefore generally less sensitive to long range perturbations than the low concentration CG-FF. This can be seen by the forces obtained with the MS-CG procedure at the high and the very low solute concentration shown in Figure 8.

3.3.3 Concentration Range Extension Procedure

To overcome the slightly too attractive long range interaction for solute molecules, the solvent interactions can be tuned to reproduce the aggregation behavior of the system, by applying a correction cut-off c_w to the potentials. The correction cut-off c_w is not to be confused with the interaction range in the MS-CG method; it is a post-MS-CG potential modification.

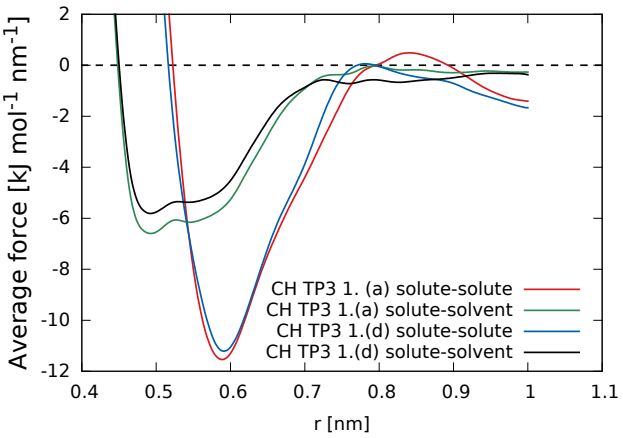


Figure 8: Average forces of all solute-solute and all solute-solvent CG interactions for the forces obtained from system 1 (a) and (d).

The same correction cut-off was applied to all solvent interactions and was chosen in an iterative trial and error process informed by the attractive nature of the long range solute-solvent and solvent-solvent forces. We found that a very small cut-off using 99.4% of the original interaction range is sufficient, once again underlining the significant sensitivity of the aggregation behavior to small long range modifications of the solvent interactions. It should be noted that the small concentration/density dependence, seen in Figure 8, has virtually no effect on key characteristic of the resulting RDFs from the transferred model at the very low concentration.

The corresponding RDFs of a CG system with a very low solute concentration and adjusted potentials are also shown in Figure 6 (right side). They demonstrate that this adjustment can successfully correct the aggregation behavior of the CG system. The cut-off means that a small fraction of the attractive solute-solvent interactions are omitted which leads to the good agreement now observed with the atomistic reference RDFs. Next, we applied the potentials with the correction cut-off c_w to the CG system 1 (a) to ensure that the long range changes do not significantly affect the performance of the CG-FF for the native concentration. The resulting RDFs, shown in Figure 3 alongside the original CG-FF, are virtually indistinguishable from the ones obtained with the unmodified potentials. Similar results have been obtained for the intermediate system 1 (c).

Here, only the solvent interactions have been adjusted in congruence with the sensitivity analysis, which suggests, that such a modification will have little effect at higher concentrations. In principle, the aggregation behavior of the model can also be adjusted by tuning the solute-solute interactions. However, the modification has to take into account the attractive nature of the long range solute-solute interactions, seen in Figure 7. Therefore, shortening the radius would result in even more repulsive solute-solute-interactions which would increase the error.

In principle, an alternative strategy would be to increase the interaction range in the MS-CG method, to possibly include more attractive solute-solute interactions at longer ranges. However, for longer interaction ranges the sensitivity of the CG potentials to long range perturbations increases, making this strategy very prone to errors. Even for higher solute concentrations, for which the method works well with a short MS-CG interaction range, the aggregation behavior begins to deviate from that of the atomistic reference system, if the MS-CG interaction range is increased.

In conclusion, a general procedure to obtain a concentration transferable CG-FF for polysaccharide solutions with correct aggregation behavior for an extended concentration range is as follows:

1. Obtain the CG-FF at a high solute concentration.

2. Apply a cut-off to the potentials (preferably solvent interactions) to reproduce the correct aggregation behavior at the lowest concentration.

3. Use the modified potentials for all concentrations within the range limited by the lowest concentration.

3.4 Degree of Polymerization Transferability

The next step is to investigate the transferability with respect to the DP. Several molecular properties of short polysaccharides in solution normalized with respect to the DP show

convergence with increasing DP beginning at DP = 8 to DP = 16.[19] This is due to the diminishing influence of the end monomers. As the properties are strongly DP dependent for small DPs, transferability of the CG-FF can not be expected and we focus on longer polysaccharides. CG interaction potentials are obtained for DP = 8 and DP = 16. The RDFs comparing the transferred and the native CG-FFs are shown in Figure 9 and reveal that the CG-FF obtained for polymers with DP=8 can be transferred to DP = 16 with only a minor decrease in accuracy as compared to the CG-FF that has been obtained at the native DP. For all sites, except the ones that involve water, the RDFs are virtually

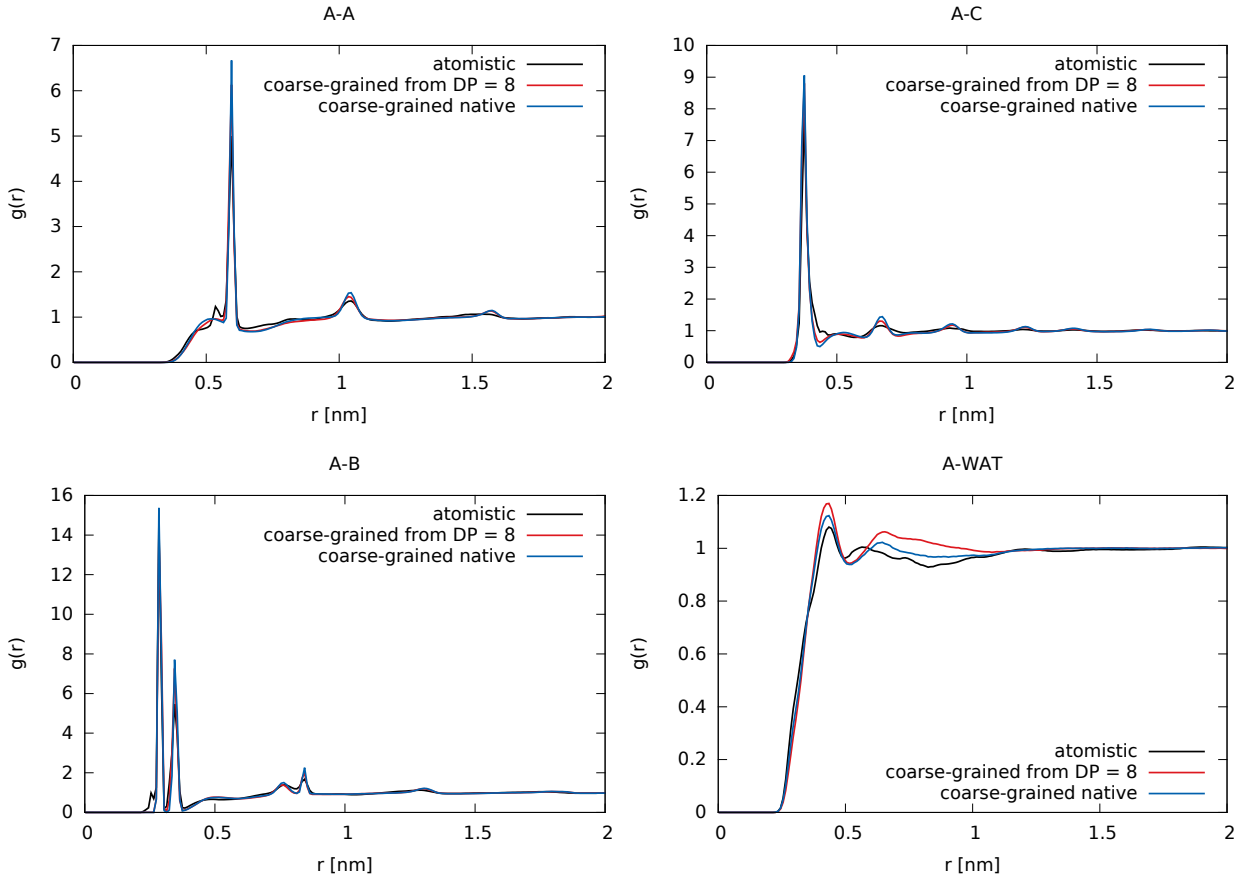


Figure 9: RDFs of the CG system 1 (a) with the potentials obtained from system 3 in comparison to the native and atomistic RDFs.

indistinguishable. For the sites involving water the transferred FF leads to slightly stronger short range correlations. Inversely, we transferred the potentials coarse-grained from DP = 16 (system 1 (a)) to DP = 8 (system 3). Again, the RDFs (shown in the Supporting

Information) show good overall agreement with each other and the atomistic reference data.

3.5 Implicit Solvent Models

Previously, it has been shown that the MS-CG method is capable of generating implicit solvent models.[58] Here, we note that the method presented also works to generate an implicit solvent model and in general all observations discussed above, including transferability, also apply in this case. As an example, Figure 10 shows the simulation result for the implicit solvent CG-FF obtained from system 1 (a) (i.e. 36 water molecules per polysaccharide) applied to system 1 (d) (i.e. 222 water molecules per polysaccharide). Note that, in contrast to the explicit solvent systems, no separation of solvent forces was used. The correction cutoff c_s had to be applied to the solute-solute interactions.

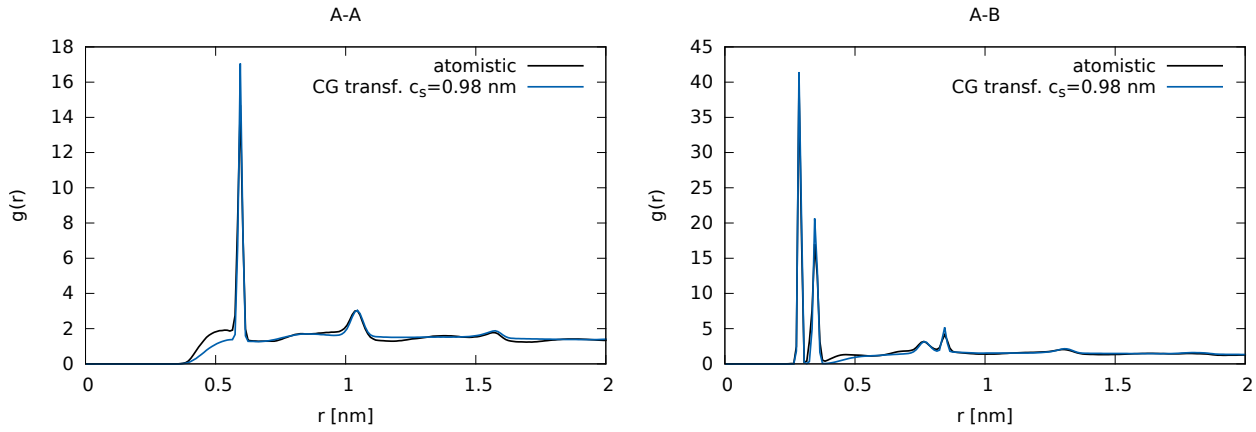


Figure 10: Resulting RDFs of the implicit solvent CG-FF obtained from system 1 (b) transferred to system 1 (d) with a cut-off c_s applied to all solute-solute interactions.

The correction cut-off c_s had to be applied to the solute-solute interactions, since explicit solute-solvent interactions are unavailable. The cut-off $c_s = 0.98$ nm was adjusted to capture the aggregation behavior of system 1 (d). A slightly greater cut-off than in the explicit solvent case was necessary here, which is consistent with the previously described lower sensitivity of the aggregation behavior to solute-solute than to solvent interactions. It can be seen that there is excellent agreement for the RDFs between the atomistic and the CG system. Again,

the CG-FF using an adjusted correction cut-off also works for the native system.

3.6 Force Field Effects

The FF used in the above procedure was CHARMM36 FF with TIP3P water. However, a recent FF comparison[19] showed that the optimal FF to study aqueous polysaccharide solutions, with respect to solution properties, is the GLYCAM06 TIP5P FF. Therefore, we now evaluate the performance of the procedure presented above with this FF. The results for system 4, a high concentration β-D-glucose system, for which no correction cut-off was necessary, are shown in Figure 11 (a). The RDFs demonstrate that the CG procedure performs

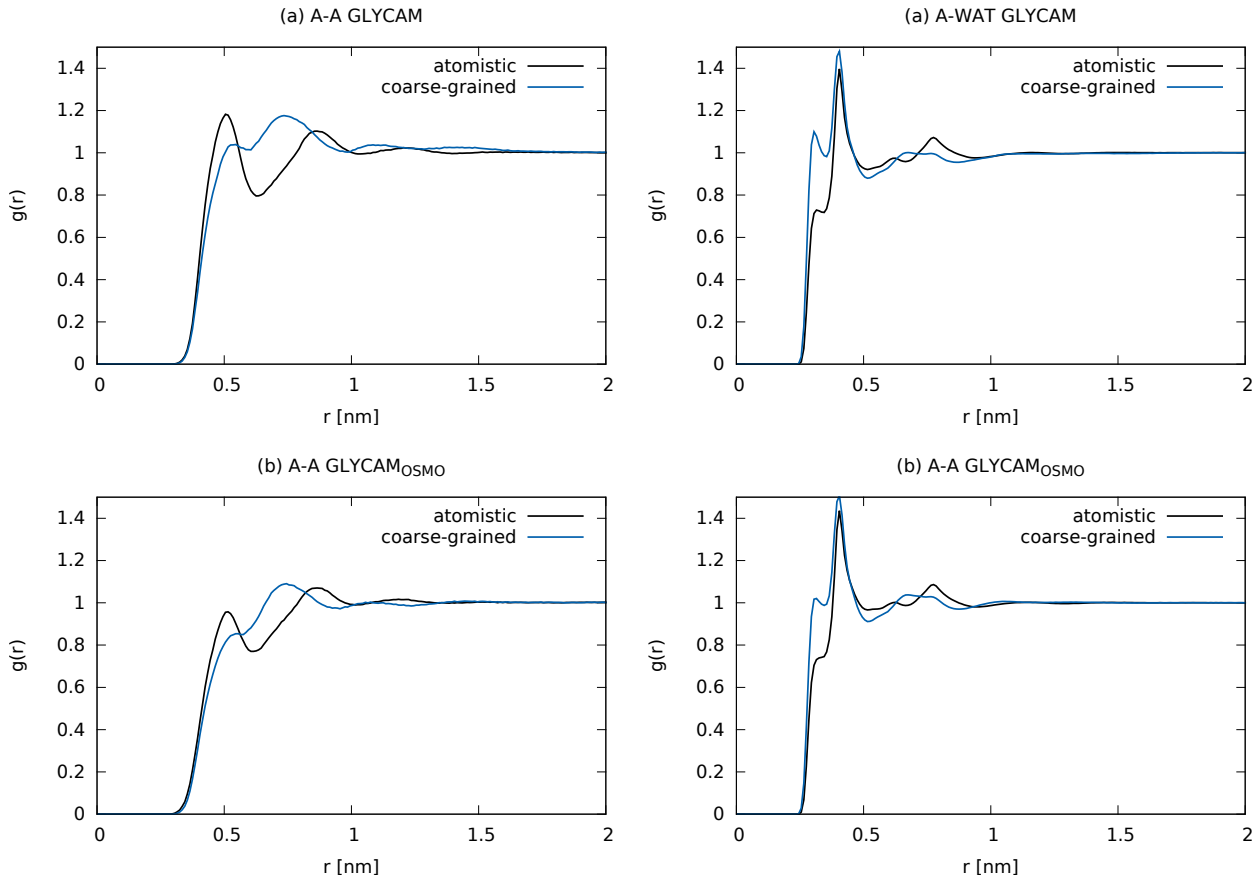


Figure 11: Example RDFs of system 4 using (a) the GLYCAM06 FF for β-D-glucose monomers and (b) the GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ FF

considerably worse using the GLYCAM06 TIP5P FF than with CHARMM36 TIP3P, espe-

cially for the solute-solute RDFs. We have seen previously,[19] that the first coordination shell in the RDF between glucose monomers (COM) in solution is significantly more populated with GLYCAM06 TIP5P than with the CHARMM36 TIP3P FF due to stronger Lennard-Jones interactions in the GLYCAM06 TIP5P FF. The stronger solute-solute coordination, makes the systems more sensitive to many-body effects similar to those observed in water. This can explain the reduced accuracy of the method.

In a separate study we introduced the GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ FF[34] parametrized to reproduce the osmotic pressure of polysaccharides in solution. In the new FF the overall solute-solute Lennard-Jones interactions are reduced. As seen in Figure 11 (b) the MS-CG method performs better with the GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ FF for the solute-solute interactions while the solute-solvent interactions are virtually unaffected, as expected, because only the solute-solute Lennard-Jones interactions were modified. Note that, in principle, this type of modification has been studied by Wagner *et al.*, however here, the changes in ϵ are larger and many-body effects are highly relevant in this system, so that a sufficiently precise prediction of the CG-FF by the formulas developed appears unlikely in this case. In addition, the situation is complicated as the ϵ parameters are scaled differently for different atomtypes in the GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ in comparison to the original GLYCAM06 FF.

Due to the stronger degree of coarse-graining of the solute molecules compared to water, 8 atoms per site as compared to 3 or 5 (virtual) atoms per site, the solute-solute interactions are more sensitive to many-body effects. With the CHARMM36 FF this stronger degree of coarse-graining is less relevant due to the smaller degree of the solute-solute coordination. However, for both GLYCAM06 FFs there is an increase in the solute-solute interaction strength, thus increasing the importance of many-body effects for solute-solute interactions. It is therefore not surprising that the observed short range errors in the RDFs are more pronounced with the GLYCAM06 FFs than with the CHARMM36 FF. This is consistent with the observation that GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ performs better than the original GLYCAM06 TIP5P FF where only solute-solute interaction have been reduced.

To better understand the balance of interactions in the different FFs we analyze the average CG forces for solute-solvent and solute-solute interactions for the CHARMM36 TIP3P FF and the GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ FF in Figure 12. The comparison of the average forces

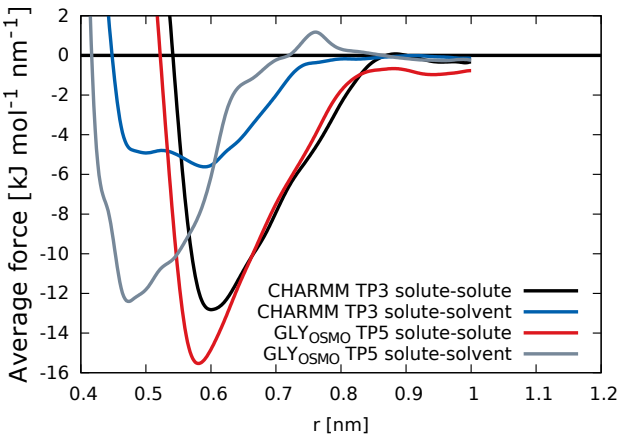


Figure 12: Average forces of solute-solute,solute-solvent, and solvent-solvent CG interactions for the forces obtained from system 4 (b) and (c).

confirms that, with the GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ FF, for all interaction classes, the forces are stronger for the GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ FF. Especially for the interactions involving solvent the minima are much deeper than in the CHARMM36 TIP3P FF based model.

To improve the performance of the coarse-graining procedure for GLYCAM06, it may be worthwhile to modify the MS-CG method to better capture many-body effects by including three-body interactions or by the use of more elaborate mapping schemes [59] to increase the short range accuracy.

4 Application: Water uptake of Branched and Linear Polysaccharides

We now investigate the water uptake of a $(\mathrm{XXXG})_4$ xyloglucan which consist of a β-D-glucose(1-4) backbone with side-chains. Here, a backbone monomer with no side-chain attached is denoted by G, a monomer with a α-D-xylose(1-6) side-chain is denoted by X,

see also Figure 13 for the chemical structure. The water uptake is compared to $(GGGG)_4$, the linear polysaccharide built of β-D-glucose with DP = 16 that was studied in the previous sections. The comparison is performed by inserting polysaccharide aggregates with low water content into a water bath. To that end, we derive a CG-FF employing the concentration range extension procedure for both molecules at the resulting very low concentration. The GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ FF is best suited to describe solution properties[19] as well as the solute-solute aggregation[34] with atomistic resolution and is therefore used as the basis for the CG-FF. In this Section, an interaction rage of 0.8 nm and a 2 fs time-step size is used in the CG-model.

4.1 Systems and CG procedure

We simulated the systems

(i) A system with 64 $(XXXG)_4$ molecules using 64 water molecules per polysaccharide.

(ii) A system with 64 $(GGGG)_4$ molecules using 432 water molecules per polysaccharide.

(iii) A system with 8 $(XXXG)_4$ molecules using 1400 water molecules per polysaccharide.

(iv) A system with 8 $(GGGG)_4$ molecules using 1400 water molecules per polysaccharide.

(v) A system with 64 $(XXXG)_4$ molecules using a total of 87988 water molecules.

(vi) A system with 64 $(GGGG)_4$ molecules using a total of 87988 water molecules.

Systems (i) and (ii) were used to obtain the CG-FF. Systems (iii) and (iv) were used to obtain the reference aggregation at an extremely low concentration where in the initial conformation the molecules are evenly distributed in the box. Due to the strong interactions in the GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ FF, discussed in Section 3.6, the dynamics of systems with high solute concentrations and high DPs is slow and, therefore, the equilibration is more challenging. However, the more balanced interactions also make the system less sensitive to perturbations,

and therefore, in system (ii) a lower solute concentration could be used to obtain the CG-FF. To obtain the initial conformations for systems (v) and (vi), we used conformations from very low water content systems *i.e.* 18 water molecules per polysaccharide for $(GGGG)_4$ and 64 for $(XXXG)_4$. The whole low water content systems were inserted into water baths so that both resulting systems contain a total of 87988 water molecules. Systems (i) to (iv) were simulated in both atomistic and coarse-grained resolution. Note that for system (vi) we now have 474,692 atoms in the system, corresponding to 93364 CG sites i.e. long atomistic reference simulations are no longer feasible for system (vi). The same hold for system (v).

Again, equilibration of the aggregation behavior was assessed by monitoring the RDFs over time and 200 ns simulation time are sufficient for the atomistic systems to obtain equlibrated RDFs. For the CG systems (iii) and (iv) 10 ns were found to be sufficient for the equilibration, whereas for systems (v) and (vi) 100 ns were used. For system (iii), the aggregation behavior fluctuates strongly over the simulation time due to the higher solubility and the small number of molecules. Therefore, the RDFs are measured over 300ns (using every 10th frame) for both the atomistic and the CG model. For $(GGGG)_4$, (system (iv)), the aggregates are more stable. However, for simplicity the same protocol as for system (iii) was used. The CG mapping for the $(XXXG)_4$ polysaccharide is shown in Figure 13. In

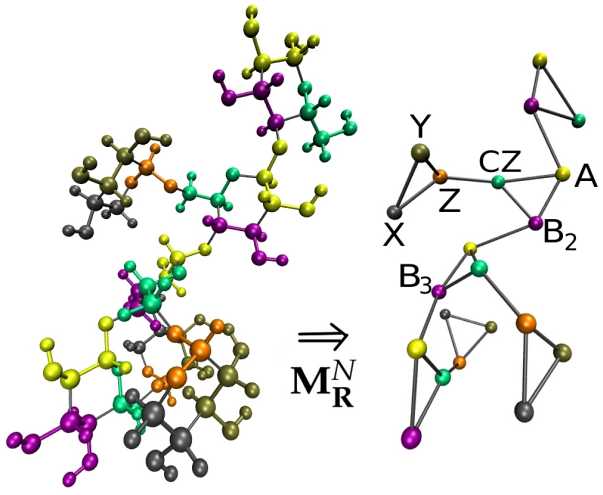


Figure 13: Example of the CG Mapping $\mathbf{M}_{\mathbf{R}}^{\mathbf{N}}$ for XXXG of the atomistic coordinates $\mathbf{r}^n$ onto CG cordinates $\mathbf{R}^N$.

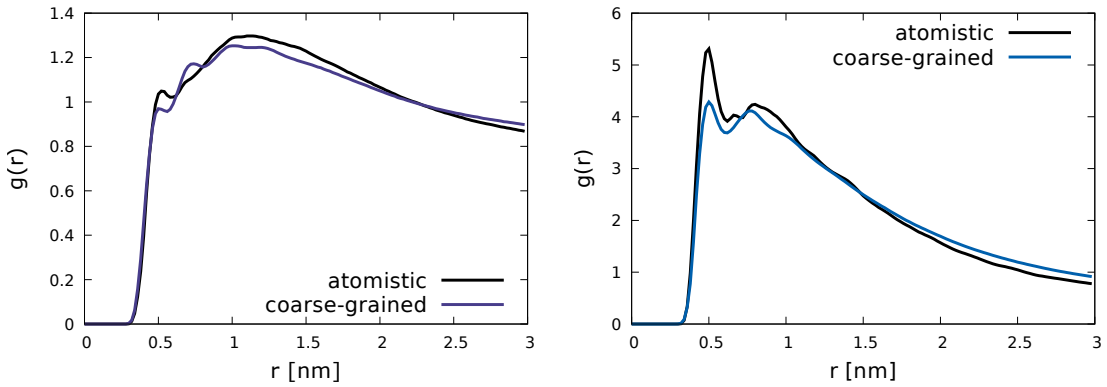


Figure 14: Solute-solute RDFs (COM) of (left) $(XXXG)_4$ and (right) $(GGGG)_4$, i.e. systems (iii) and (iv) for both the atomistic and the coarse-grained model

addition to the interactions described in Section 3.1 for $(GGGG)_4$, three (BI-) bonds were used within a side-chain monomer of $(XXXG)_4$. The side-chain monomer was connected to the backbone using one (BI-) bond Z - CZ and the (BI-) angle CZ - Z - X and the (BI-) dihedrals CZ - Z - X - Y and X - Z - CZ - B2. The additional, explicit (BI-)1-3 non-bonded intra-molecular interactions connecting the sites A - Z and B2 - Z were used. Similarly, the (BI-)1-4 interactions B3 - Y was explicitly included.

For $(GGGG)_4$ and $(XXXG)_4$ the CG-FF had to be modified to match the reference aggregation at the extremely low concentration, in the same way as described in Section 3.3.2, *i.e.* a small constant force of 1.0 kJ $mol^{-1}nm^{-1}$ was added to all solute-solvent interactions within the range of 0.8 nm to 0.806 and 0.8 nm to 0.864 nm for $(GGGG)_4$ or $(XXXG)_4$, respectively. The forces of the other interaction classes remained unchanged. The resulting overall solute-solute RDFs are shown in Figure 14 for both molecules at the extremely low concentration for the atomistic and the CG system. It can be seen that with this correction the aggregation behavior is in good agreement with the atomistic reference simulation for the extremely low concentration.

4.2 Results

The final solute-solute RDFs of system (v) and (vi) are shown in Figure 15, and representative snapshots of both systems are shown in Figure 16. The RDF in Figure 15 demonstrates that the $(GGGG)_4$ molecules have a strong tendency to aggregate, resulting is a single cluster with few connections over periodic boundary conditions. Note that the aggregation in the RDFs is more pronounced than in the RDF of system (iv) (Figure 14 b) because of the greater number of solute molecules. The $(XXXG)_4$ polysaccharides aggregate much less strongly. However, there is still a clear self-assembling tendency, recognizable in the RDF values which remain larger than 1. This leads to the formation of a network spanning the whole box.

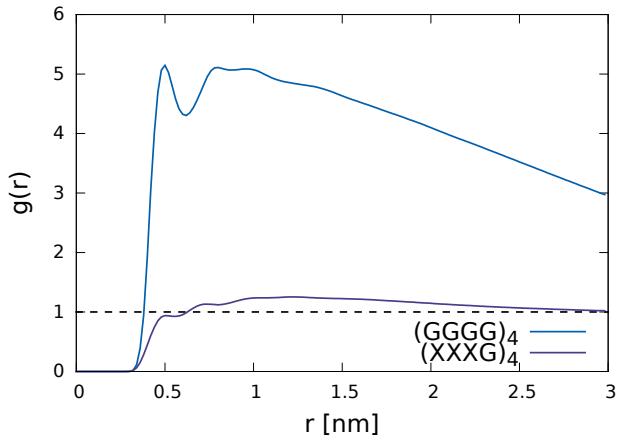


Figure 15: Solute-solute RDFs (COM) of $(XXXG)_4$ and $(GGGG)_4$ obtained with the derived CG-FF based on the GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ FF with an aggregated starting conformation using 87.988 water molecules i.e. systems (v) and (vi).

To check for initial state effects, and test whether these RDFs are a result of the aggregated starting conformation, system (v) and (vi) were also simulated with an evenly distributed initial conformation and the RDFs obtained were virtually identical to the ones shown in Figure 15. These results support the previous suggestion[10] that monosaccharide side-chains increase the water affinity in comparison to linear polysaccharides because of the reduced aggregation of the brached polysaccharides in comparison to the linear ones. In addition, the branched polysaccharides still have a tendency to aggregate which is constistent

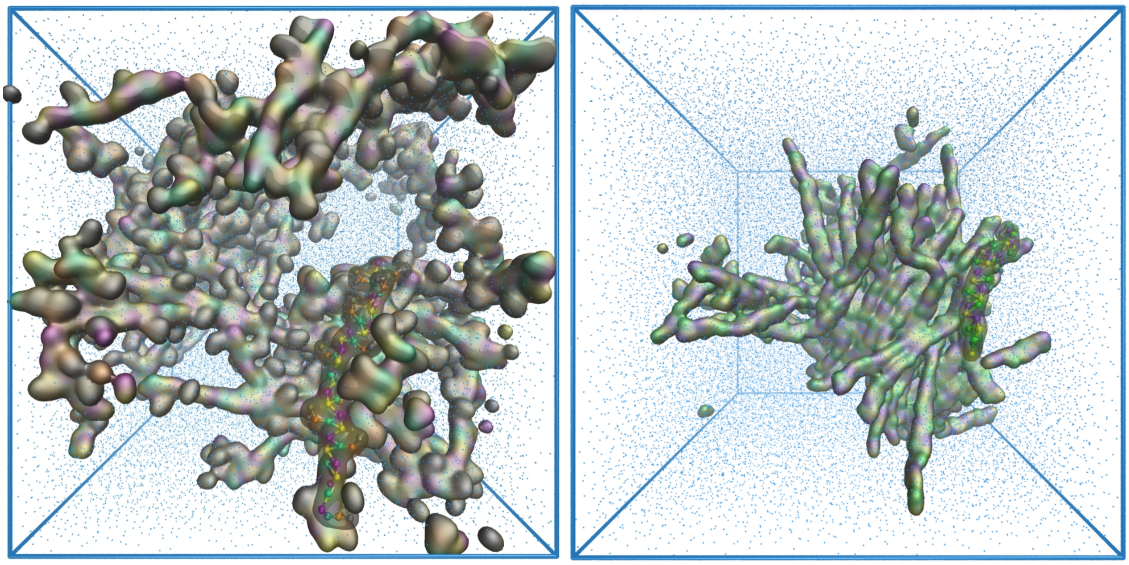


Figure 16: Snapshots of systems (v) (left) and (vi) (right) after equilibration.

with the formation of a gel[11] for xyloglucans with side-chain length one.

This sample application illustrates, that the CG model is consistent with both the atomistic reference simulation and experimental data. Thus it is well suited to be applied for instence in an in depth study of the structure-function relationship of hemicellulose, as well as wide range of other potential applications. A property worth exploring for instance is the swelling capacity quantified by the osmotic pressure as a function of the chemical structure. Atomistic calculation of the osmotic pressure requires microsecond sampling to acquire converged results, even for simple systems like XXXG. A CG-FF on the other hand offers a possibility to study longer polysaccharides and the effects of longer side-chains and side-chain variations within affordable computational cost.

5 Conclusions

We have demonstrated that a hybrid procedure combining the Multiscale Coarse-Graining method and Boltzmann Inversion, can be used to generate an accurate model for aqueous polysaccharides that is transferable over both, different polymer lengths and different concentrations. If the model is intended for use at very low concentrations an extended procedure

can be used to correct the aggregation at the lowest concentration, which leads to good agreement with the atomistic model over the entire concentration range. The same method can be used to generate implicit solvent models. When the coarse-graining procedure is applied to systems with more pronounced many-body effects it can still be used, however, the agreement between coarse-grained and atomistic data should be expected to be less accurate. Although here, the procedure was used for aqueous polysaccharide solutions it is in principle applicable to other solute-solvent systems, making use of the presented considerations. Although the method was applied here using a 3-site mapping scheme, in principle, finer or coarser mappings are also possible.

With this CG procedure we performed simulations to show that aggregates of branched polysaccharides with monomer side-chains dissolve better in comparison to linear polysaccharides when coupled to a water bath.

6 Acknowledgements

The project was funded by the Deutsche Forschungsgemeinschaft (GR 3661/2-1). The authors thank Reinhard Lipowsky, Gregory A. Voth, Jacob W. Wagner and James F. Dama for helpful discussions.

References

(1) Parham, P. *The Immune System, Fourth Edition*; Taylor & Francis Group: Abingdon, United Kingdom, 2014.

(2) Kleiber, M. *The Fire of Life. An Introduction to Animal Energetics.*; John Wiley & Sons, Inc.: Hoboken, New Jersey, 1961.

(3) Pauly, M.; Keegstra, K. *Plant J.* **2008**, *54*, 559–568.

(4) Fratzl, P.; Gupta, H.; Burgert, I. *Comp Biochem Physiol A Mol Integr Physiol* **2007**, *146*, S132–S132.

(5) Yamamoto, H.; Ruelle, J.; Arakawa, Y.; Yoshida, M.; Clair, B.; Gril, J. *Wood Sci. Technol.* **2009**, *44*, 149–163.

(6) Harrington, M. J.; Razghandi, K.; Ditsch, F.; Guiducci, L.; Rueggeberg, M.; Dunlop, J. W. C.; Fratzl, P.; Neinhuis, C.; Burgert, I. *Nat. Commun.* **2011**, *2*, 337.

(7) Dawson, C.; Vincent, J. F. V.; Rocca, A.-M. *Nature* **1997**, *390*, 668–668.

(8) Fry, S. C. *J. Exp. Bot.* **1989**, *40*, 1–11.

(9) Wyman, C.; Decker, S.; Himmel, M.; Brady, J.; Skopec, C.; Viikari, L. *Polysaccharides*; 2004; pp 994–1033.

(10) Hayashi, T.; Takeda, T.; Ogawa, K.; Mitsuishi, Y. *Plant and cell physiology* **1994**, *35*, 893–899.

(11) Reid, J. G.; Edwards, M.; Dea, I. C. *Gums and stabilizers for the food industry*; IRL Press: Oxford, United Kingdom, 1988; Vol. 4; pp 391–398.

(12) Esko, J. D.; Kimata, K.; Lindahl, U. *Essentials of Glycobiology*; 2009; p 784.

(13) Marrink, S. J.; De Vries, A.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750–760.

(14) López, C. A.; Rzepiela, A. J.; De Vries, A. H.; Dijkhuizen, L.; Hunenberger, P. H.; Marrink, S. J. *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210.

(15) Reith, D.; Pütz, M.; Müller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624–1636.

(16) Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1995**, *52*, 3730.

(17) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.

(18) Shell, M. S. *J. Chem. Phys* **2008**, *129*.

(19) Sauter, J.; Grafmüller, A. *J. Chem. Theory Comput.* **2015**, *11*, 1765–1774.

(20) Wagner, J. W.; Dama, J. F.; Voth, G. A. *J. Chem. Theory Comput.* **2015**, *11*, 3547–3560.

(21) Krishna, V.; Noid, W. G.; Voth, G. A. *J. Chem. Phys* **2009**, *131*.

(22) Wang, Y.; Noid, W.; Liu, P.; Voth, G. A. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2002–2015.

(23) Wang, Y.; Feng, S.; Voth, G. A. *J. Chem. Theory Comput.* **2009**, *5*, 1091–1098.

(24) Zhang, J.; Guo, H. *J. Phys. Chem. B* **2014**, *118*, 4647–4660, PMID: 24712306.

(25) Molinero, V.; Goddard, W. A. *J. Phys. Chem. B* **2004**, *108*, 1414–1427.

(26) Liu, P.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 11566–11575.

(27) Markutsya, S.; Devarajan, A.; Baluyut, J. Y.; Windus, T. L.; Gordon, M. S.; Lamm, M. H. *J. Chem. Phys* **2013**, *138*.

(28) Hynninen, A.-P.; Matthews, J. F.; Beckham, G. T.; Crowley, M. F.; Nimlos, M. R. *J. Chem. Theory Comput.* **2011**, *7*, 2137–2150.

(29) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. *J. Comput. Chem.* **2008**, *29*, 2543–2564.

(30) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell Jr, A. D. *J. Chem. Theory Comput.* **2009**, *5*, 2353–2370.

(31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys* **1983**, *79*, 926–935.

(32) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.

(33) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys* **2000**, *112*, 8910–8922.

(34) Sauter, J.; Grafmüller, A. *J. Chem. Theory Comput.* **2016**, *12*, 4375–4384.

(35) Case, D.; Darden, T.; Cheatham III, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K. AMBER 12. 2012.

(36) Woods Group GLYCAM Web. Complex Carbohydr. Res. Center, University of Georgia, Athens, GA (http://www.glycam.com), *Polysaccharides: Structural diversity and functional versatility* **(2005-2016)**,

(37) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472–2493.

(38) Wehle, M.; Vilotijevic, I.; Lipowsky, R.; Seeberger, P. H.; Varon Silva, D.; Santer, M. *J. Am. Chem. Soc.* **2012**, *134*, 18964–18972.

(39) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D. *Bioinformatics* **2013**, *29*, 845–854.

(40) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(41) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.

(42) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(43) Barker, J.; Watts, R. *Mol. Phys.* **1973**, *26*, 789–792.

(44) Birdsall, C. K.; Langdon, A. B. *Plasma Physics Via Computer Simulation*; McGraw-Hill, Inc.: New York City, New York, 1986.

(45) Nosé, S. *J. Chem. Phys* **1984**, *81*, 511–519.

(46) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

(47) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(48) Nose, S.; Klein, M. *Mol. Phys.* **1983**, *50*, 1055–1076.

(49) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38.

(50) Noid, W.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys* **2008**, *128*.

(51) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. *J. Chem. Theory Comput.* **2009**, *5*, 3211–3223.

(52) Mashayak, S.; Jochum, M. N.; Koschke, K.; Aluru, N.; Rühle, V.; Junghans, C. *PloS one* **2015**, *10*, e0131754.

(53) Lu, L.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. *J. Chem. Theory Comput.* **2010**, *6*, 954–965.

(54) Izvekov, S.; Voth, G. A. *J. Chem. Phys* **2006**, *125*.

(55) Larini, L.; Lu, L.; Voth, G. A. *J. Chem. Phys* **2010**, *132*.

(56) Lu, L.; Dama, J. F.; Voth, G. A. *J. Chem. Phys* **2013**, *139*.

(57) Maiolo, M.; Vancheri, A.; Krause, R.; Danani, A. *J. Comput. Phys.* **2015**, *300*, 592–604.

(58) Cao, Z.; Dama, J. F.; Lu, L.; Voth, G. A. *J. Chem. Theory Comput.* **2013**, *9*, 172–178.

(59) Cao, Z.; Voth, G. A. *J. Chem. Phys* **2015**, *143*.

Graphical TOC Entry