A Martini-based coarse-grained soil organic matter model derived from atomistic simulations

Lorenz F. Dettmann,[*,†] Oliver Kühn,[†,‡] and Ashour A. Ahmed[†]

†*University of Rostock, Institute of Physics, Albert-Einstein-Str. 23-24, D-18059 Rostock, Germany.*

‡*University of Rostock, Department of Life, Light and Matter (LLM), Albert-Einstein-Str. 25, D-18059 Rostock, Germany*

E-mail: lorenz.dettmann@uni-rostock.de

Abstract

The significance of soil organic matter (SOM) in environmental contexts, particularly its role in pollutant adsorption, has prompted increased utilization of molecular simulations to understand microscopic interactions. This study introduces a coarse-grained SOM model, parametrized within the framework of the versatile Martini 3 force field. Utilizing systems generated by the Vienna Soil Organic Matter Modeler 2, which constructs humic substance systems from a fragment database, we employed Swarm-CG to parametrize the fragments and subsequently assembled them into macromolecules. Direct Boltzmann Inversion (DBI) facilitated the determination of bonded parameters between fragments. The parametrization yielded favorable agreement in the radius of gyration and solvent-accessible surface area. Transfer free energies exhibited a strong correlation with hexadecane-water and chloroform-water values, albeit deviations were noted for octanol-water values. Comparing densities of modeled Leonardite Humic Acid (LHA) systems at coarse-grained and atomistic levels revealed promising agreement, particularly at higher water concentrations. The DBI approach effectively reproduced average values of bonded interactions between fragments. Radial distribution functions between carboxylate groups and calcium ions partially concurred, yet limitations arose in reproducing certain peaks due to fixed bead sizes. Detailed analysis of atomistic systems elucidated different configurations between groups, further explaining discrepancies. The present contribution provides a comprehensive insight into the properties, strengths, and weaknesses of the coarse-grained SOM model, serving as a foundation for future investigations encompassing pollutant interactions and varied SOM compositions.

1 Introduction

Soil is a complex, dynamic, and heterogeneous system, composed of mineral and organic matter, as well as water and air. It is of biogeochemical relevance because it provides a reservoir of resources, such as carbon, nitrogen, and water, which are essential for life. Soil organic matter (SOM) refers to the total amount of organic material in the soil. Although SOM represents only a small percentage of most top soils, it plays a central role in the global cycling of carbon[1].

The classification of SOM can be subdivided depending on the origin and degree of decomposition of plant and animal residues. The term "raw organic matter" is often used to describe the organic material that is found at the initial stages of decomposition. At the final stage of this process the so-called "humus" remains, which is the dead part of SOM[1,2]. Humic substances (HSs) can be categorized into fulvic

acids, humic acids (HAs), and humin. This definition is based on the extraction methods used for these substances.

From a chemical point of view, SOM is one of the most heterogeneous substances, containing a wide spectrum of molecules such as carbohydrates, proteins, lipids and others. This also results in the presence of a multitude of functional groups, often including aromatic and aliphatic ones, as well as functional groups composed of different elements such as oxygen, nitrogen or sulphur.

Because it has a relatively large surface area, SOM is an important sorbent for organic and inorganic substances. Mainly negatively charged functional groups increase the cation exchange capacity. Conversely, less-soluble hydrophobic substances can be bound to hydrophobic areas, which are also present in SOM. Furthermore, SOM is crucial for the development of a stable soil structure through various aggregation processes[1].

As an important part of ecosystems, soils fulfill many functions both for humans and the environment. The presence of chemical contaminants and physical soil degradation potentially endangers these soil functions[3]. Examples of contaminants include persistent organic pollutants (POPs)[4], micro- and nanoplastics[5,6], heavy metals[7], and antibiotics[8], among others. Ongoing research is being conducted to determine the influence of these various compounds on the soil. Different modeling approaches have been presented for undertaking such investigations[9–14].

Molecular dynamics (MD) simulations are a powerful method for revealing the dynamics and interactions that occur at the nanoscale level. It is of major interest to better understand sorption processes with SOM. For instance, interactions between carbon nanotubes and nanoplastic particles were previously investigated by our group, using the Martini force field[15]. Among other aspects, this allowed the cavity and trapping effects to be investigated, which are similar to the adsorption processes that occur in SOM.

A systematic modeling approach was introduced in the form of the Vienna Soil Organic Matter Modeler (VSOMM)[16,17]. This is a tool for producing condensed-phase models of SOM, more specifically HSs. It uses a fragment database to construct HS molecular structures based on input parameters. Thus, it provides the user with a series of distinct models, that are diverse in structure and composition, reflecting the heterogeneity and complexity of SOM. Numerous SOM systems have been investigated using this modeler[2,18–21], including interactions with proteins[22] and mineral surfaces[23], among others.

However, the applicability of MD simulations can be limited by the system size and time scales of the processes of interest. For instance, modeling supramolecular SOM structures, and their interactions with long-chained carbohydrates or complex proteins, would require to take into account a large number of atoms in the simulation. The range of adsorption processes and free energy calculations to be investigated is often limited by computational resources as well. In such cases, the efficiency of the MD method could be compromised.

One way to speed up simulations is to use a coarse-grained instead of an atomistic representation of the molecules[24]. Thereby, atoms are grouped into new particles, so-called beads, which effectively reduce the number of particles and interactions in the system. The link between the two representations is known as mapping. Furthermore, coarse-graining includes finding new interaction parameters for the beads. Over the last few decades, various coarse-graining procedures have been introduced, among them approaches leveraging machine learning methodologies[25,26]. Similarly to atomistic methods, force fields for coarse-grained simulations exist[27–29]. Of these, the Martini force field[30] is the most commonly used.

Martini uses a building block approach with pre-parametrized beads that are applicable to a broad range of chemical compounds. Recently, a new version, Martini 3, was published, which involved a complete reparametrization of the force field[31]. New types of beads were introduced, including three different applicable bead sizes. In addition, a variety of molecules that

have been parametrized for Martini are available. However, the absence of a SOM model that aligns with the latest version of the Martini force field, while adhering to the concept of parametrizing new molecular systems, hinders the simulation of various SOM compositions. Addressing this gap is the primary focus of this study. While the first version of the VSOMM has been transferred to a coarse-grained representation using Dissipative particle dynamics (DPD)[32], a generalized coarse-grained version of the VSOMM2 compatible with Martini 3 would benefit from integrating the already parametrized molecules. This integration would enable the investigation of their interactions with SOM. Thus, employing a coarse-grained representation would facilitate simulations of adsorption processes and free energy calculations for macromolecular SOM systems, enhancing computational efficiency.

Therefore, this study aims to present and test a coarse-grained representation of selected VSOMM2 models. This is done by parametrizing the fragments of the modeler using Martini 3 and with the Swarm-CG[33] tool. As an example, a coarse-grained model of the Leonardite Humic Acid (LHA) is created, which has been previously studied using MD simulations[2,18–23]. The accuracy of the model is tested against thermodynamic properties and atomistic simulations.

2 Methods

In this section, a general overview of the approach is given first, followed by a more detailed description of the parametrization approach and the computational details.

2.1 General overview of the approach

The coarse-grained SOM model is based on the VSOMM2[17]. This Modeler creates HS structures, building them from basic chemical fragments (see Figure S1 in the supporting information). The fragment combination is dependent on input parameters, such as the amount of certain functional groups that should occur in the system, or the Carbon/Nitrogen (C/N) ratio. The coarse-graining approach follows this method, by first coarse-graining the fragment molecules and then assembling them into a macromolecular structure.

The first part of this study deals with the parametrization of these fragments. The mapping and non-bonded interactions are then set according to the guidelines for parametrizing new molecules for the Martini 3 force field[31]. The parameters for bonded interactions are determined via Swarm-CG[33], using trajectories of isolated fragments in water in their atomistic representation as input. Transfer free energies, calculated using the COSMO-RS method[34–36] are used as references, to be compared with free energy values, calculated with the coarse-grained fragments.

In the second part, a coarse-grained SOM model is created, based on an atomistic reference from the VSOMM2. Thereby, the aforementioned coarse-grained fragments are used for the model construction. More specifically, coarse-grained models of LHA are investigated and compared to atomistic models. Bonded interactions between coarse-grained fragments are automatically determined via Direct Boltzmann Inversion (DBI), which is further explained below (see section 2.2.4). The modeling approach was based on previous works that examined LHA at the atomistic level[18,20].

2.2 Parametrization approach

2.2.1 Determination of mapping and non-bonded interactions

The mapping of each fragment was determined following the Martini guidelines for parametrizing new small molecules. This includes maximizing the degree of mapping, or in other words, using the least number of coarse-grained beads as possible, to describe a molecule. In addition, the symmetries of molecules should not be broken, and functional groups should be kept together as much as possible. Regarding the new Martini 3 force field, a center of geometry mapping should be used to repro-

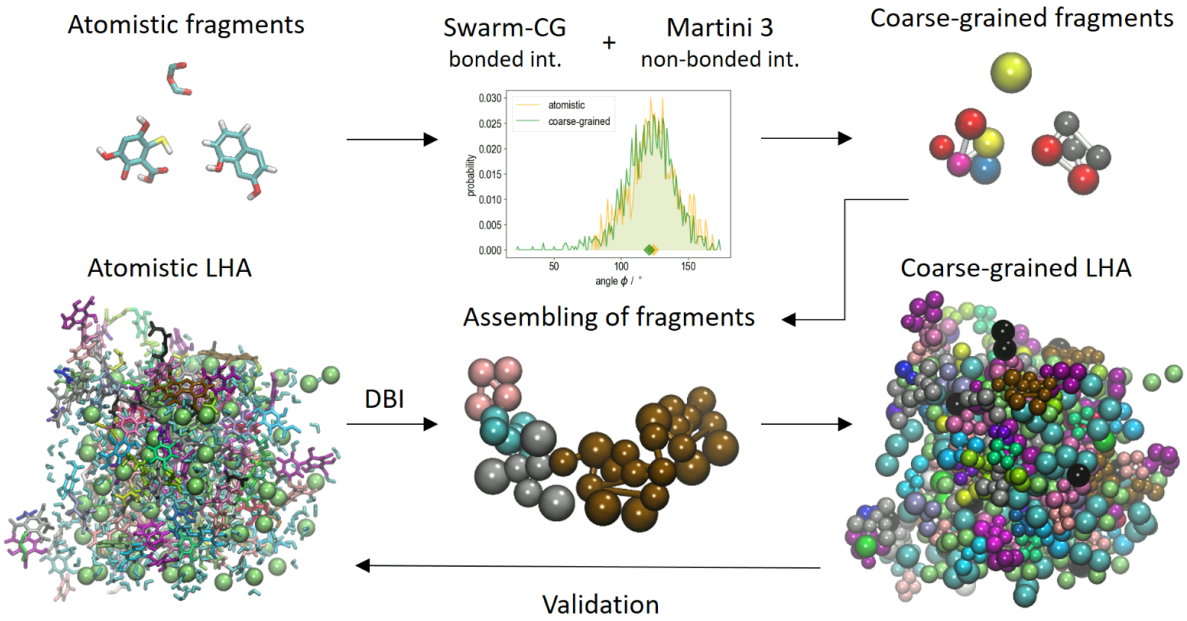

Figure 1: Overview of the procedure for creating the coarse-grained SOM models. Atomistic fragments are transformed into a coarse-grained representation using Swarm-CG and Martini 3. These fragments are assembled with parameters from Direct Boltzmann Inversion (DBI), determined from atomistic simulations, for bonded interactions between different fragments. This results in the coarse-grained SOM models. The properties of these models are then compared with the atomistic systems.

duce the molecular shape more accurately. This also includes giving preference to certain sizes of beads, for example for aromatic molecules (tiny beads) or aliphatic ring structures (small beads). The mapping, which was applied to the fragments is shown in Figure S1 in the supporting information.

An important part of the whole parametrization is the determination of non-bonded interactions. In Martini, a building block approach is used with pre-defined beads for certain atom groups. This means that all interactions are defined via a discrete set of possible interaction parameters. These parameters can be fine-tuned, based on thermodynamic properties such as transfer free energies or miscibility trends. In the present case, for most of the atom groups, standard beads were applied, which are beads that mostly reflect the behavior of the corresponding atom group in an isolated form. An explanation for every bead that was applied can be found in Table S1 in the supporting information.

Some of the main bead classes in Martini 3 are polar (P), intermediate (N), and apolar (C). Each of these classes contains 6 beads with different degrees of polarity, ranging from 1 (low polarity) to 6 (high polarity). In contrast to regular beads, small and tiny beads are represented by the first letter with S and T, respectively. For further information on the available bead types, labels and the nomenclature, please refer to the supplementary information of the main publication of Martini 3[31], sections A1 and A4.

2.2.2 Determination of bonded interactions of fragment molecules

Bonded interactions follow from the mapping of a molecule. Bonds were constrained for aromatic molecules, whose rings were each described by three TC5 beads. Otherwise, harmonic bond potentials were applied. Furthermore, harmonic angle potentials and improper dihedral angle potentials were used. The latter were applied mainly to relatively large fragments, like HS12, HS29, or HS34, to retain their

structure.
In general, to improve the numerical stability, restricted bending potentials can be used, especially, if an angle between constructing beads for a dihedral angle is close to 180°. There was no need to apply such potentials within the individual fragment molecules. However, these potentials played a role when connecting the individual fragments to a macromolecular structure.
For certain fragments such as HS2 or HS12, virtual sites (VSs) were introduced. Such particles improved the numerical stability of the respective molecules by avoiding complicated constraint networks, which would otherwise be necessary. Here again, VSs were similarly introduced like existing Martini parametrizations of small molecules, e.g. HS29, whose coarse-grained structure is similar to naphthalene. The parameters for all the VSs were determined from the reference trajectories of the isolated fragments in water in their atomistic representation.
The parameters for all the other bonded interactions (bonds, angles, and dihedral angles) were optimized with Swarm-CG[33]. This is a tool that is applied to perform an automated parametrization of bonded interactions of coarse-grained molecules, which can be used in combination with the Martini 3 force field. It uses particle swarm optimization to find an optimal solution for a given set of bonded interaction parameters, thereby minimizing the so-called earth mover's distance between distributions of bonded interactions of mapped and coarse-grained molecules. Overall, it is an iterative optimization procedure, whereby multiple simulations are sequentially conducted and bonded parameters adjusted according to the reference.
To facilitate the reproduction of structural properties of the fine-grained reference, the two nearest bonded neighbors were excluded from the non-bonded interactions, following a common practice in certain parametrizations in the frame of the Martini force field[37–39].

2.2.3 Determination of reference properties

As mentioned earlier, the transfer free energies ΔG are the main target property of the Martini parametrization. In this study, octanol-water, hexadecane-water, and chloroform-water transfer free energies were calculated from the respective partition coefficients P from COSMO-RS, using the equation $\Delta G = -RT \ln P$ Here, R is the universal gas constant and T is the temperature. Computational details regarding the calculation of ΔG are provided in section 2.3.3.
The solvent-accessible surface area (SASA) for coarse-grained and atomistic fragments was calculated, to compare their molecular shape. This was done with the GROMACS tool *gmx sasa*, using 4800 dots per sphere and a probe radius of 0.191 nm for atomistic and coarse-grained fragments. The van der Waals radii from the work of Roland and Taylor[40] were used for calculating the SASA of the atomistic fragments. In the case of the coarse-grained fragments, radii were chosen according to the bead sizes. It should be noted that, due to using the united atom representation of the GROMOS force field, experimentally determined radii were not available for the united atoms, which represent the CH_1, CH_2, and CH_3 groups. Instead, they were approximated using that of a single carbon atom. To provide other reference values for the SASA, values with van der Waals radii from GROMOS and from the mapped molecules were also calculated (see Figure S2).
Furthermore, the radius of gyration (R_g) was compared to the atomistic reference as a measure of the extent of a fragment, using the GROMACS tool *gmx gyrate*. For each fragment, both properties (R_g and SASA) were determined firstly from the atomistic reference trajectories of isolated fragments and secondly for the coarse-grained representation, using the trajectory with the converged model, which was part of the Swarm-CG parametrization. The time averages and corresponding standard deviations were determined from time evolutions using *gmx analyze*. This was done in the same

way for other quantities used in this research, that are connected to a time evolution, for example the SASA, densities, and distributions of bonded interactions, all of which are discussed later.

2.2.4 Construction of coarse-grained SOM models (LHA)

In order to simulate coarse-grained LHA, fragments have to be combined into a macromolecular structure. Based on the input parameters, the modeler creates macromolecules that generally differ in terms of the sequence of fragments. The challenge is then to determine the bonding parameters between the individual fragments for such an arbitrary sequence.
The procedure is as follows. It starts with an atomistic structure that provides information about the sequence of the fragments and the initial structure. The latter is obtained by mapping the atomistic molecules to their coarse-grained representation. Bonded and non-bonded interaction parameters of isolated fragments, determined previously, were assigned to the corresponding fragment parts of the macromolecular structure. Bonds, angles, and proper dihedral angles between fragments are automatically determined based on the connectivity of the fragments.
However, parameters for these bonded interactions are missing and have to be determined based on a reference. Therefore, atomistic simulations of the macromolecular structures were conducted. To minimize further parametrization effort, (non-iterative) DBI was used to estimate the parameters for bonded interactions between the fragments. For angles that were involved in a dihedral angle potential, restricted bending potentials were used, to avoid numerical instabilities.

2.3 Computational Details

All simulations were performed with the software package GROMACS[41] version 2019.4. The analysis was conducted using GROMACS tools and the Python 3 package MDAnalysis[42,43]. Other packages that were utilized included NumPy[44], SciPy[45], IPython[46] and Matplotlib[47]. Structures were visualized with VMD[48]. Unless otherwise stated, the solvation of molecules was performed using Packmol[49].

2.3.1 Atomistic isolated fragments

For the atomistic simulations, the GROMOS force field version 54A7[50,51] was used, together with the SPC water model[52]. Each fragment was solvated in a cubic water box with an edge length of 4 nm.
The following procedure was applied to each of the systems: First, an energy minimization was performed, followed by an NVT equilibration and an NPT equilibration of 100 ps each. Finally, a production run of 25 ns was conducted. For the equilibrations and the production run, the temperature of the system was held constant with the velocity-rescale thermostat[53] at 298.15 K. Furthermore, equations of motion were integrated with the leap-frog integrator, using a time step of 2 fs. For the NPT equilibration and the production run, the pressure of the system was modified by the Parrinello-Rahman barostat[54] with a coupling constant of $\tau_p = 1.0$ ps and a reference pressure of 1 bar. In the case of short-range non-bonded interactions, the Verlet cutoff scheme was applied with a cutoff radius of 1.4 nm. The reaction-field cutoff scheme was used for the Coulomb interactions with the same cutoff radius as for the short-range interactions and with a relative dielectric constant of $\epsilon_{rf} = 61.0$. Periodic boundary conditions were applied in each direction. All the bonds were constrained using the LINCS algorithm.

2.3.2 Atomistic LHA models

The LHA models were generated by the VSOMM2. For more information on the model construction, please refer to the original publication of the Modeler[17]. The input parameters, which were based on an experimentally determined[55] composition of LHA and the final composition of the constructed systems, are shown in Table 1. All the systems were modeled at a pH value of 7. Calcium was used as counter

Table 1: Experimentally (exp) determined composition of carbon and nitrogen, and functional groups, contained in LHA, and corresponding composition of the LHA models. Experimentally determined values were given as input parameters to the modeler.

fraction	exp	LHA1	LHA2	LHA3
Carbon	0.638	0.548	0.546	0.535
Nitrogen	0.012	0.014	0.014	0.015
Carbonyl	0.080	0.077	0.077	0.076
Carboxyl	0.150	0.143	0.143	0.143
Aromatic	0.580	0.585	0.586	0.592
Acetal	0.040	0.039	0.040	0.039
Heteroaliphatic	0.010	0.010	0.010	0.010
Aliphatic	0.140	0.141	0.140	0.138

ions because it is abundant in most soils and soil solutions. In total, 200 building blocks were used, with 5 building blocks per molecule.
Three compositions of LHA were considered, which were realized by using three different random seeds for the model construction. For each composition, different water contents were considered, which varied from 500 to 1500 water molecules per system, increasing gradually by 100 molecules. In total, this produced 33 model systems. The simulation parameters were the same as for the systems with isolated fragments in water, described in section 2.3.1. However, the NVT and NPT equilibrations, respectively, lasted 1 ns and the simulation time of each production run was set to 100 ns, where the final 50 ns were respectively used for evaluation.

2.3.3 Coarse-grained isolated fragments

Coarse-grained models were simulated in the framework of the Martini force field version 3. Initially, the structures of the coarse-grained fragments were generated by mapping the atomistic structures using the CGBuilder tool (`https://jbarnoud.github.io/cgbuilder`).
The fragments were then solvated in a cubic water box with an edge length of 4 nm. Subsequently, these systems were used as starting points for the iterative optimization by Swarm-CG for each fragment.
In each iteration, energy minimization, NPT equilibration, and the production run were carried out consecutively. The equilibration time was set to 1 ns. The production run lasted either 10 ns or 25 ns, depending on the respective optimization cycle. For more details about the optimization procedure used by Swarm-CG, see the corresponding study[33]. The equations of motions were integrated, using the leap-frog integrator, with a time step of 20 fs. For the equilibration, the Berendsen barostat[56] was used with a coupling constant of $\tau_p = 4.0$ ps, whereas for the production run, the Parrinello-Rahman barostat was used with a coupling constant of $\tau_p = 8.0$ ps. For both simulations, the reference pressure was set to 1 bar. The temperature was held constant with the velocity-rescale thermostat at 298.15 K.
The following simulation parameters were based on the recommended Martini run parameters[57]. The Verlet cutoff scheme with a straight cutoff of 1.1 nm was used. Coulomb interactions were treated with the reaction-field approach, using the same cutoff radius and a relative dielectric constant of $\epsilon_r = 15$. Constraints were included using the LINCS algorithm. Periodic boundary conditions were applied for all the coarse-grained simulations.
The optimized models produced by Swarm-CG were subsequently used for the free energy calculations. As in the Martini 3 main publication, only neutral molecules were considered for these calculations[31]. The coarse-grained fragments were respectively solvated in water, hexadecane, chloroform and water-saturated octanol, with a mole fraction of 0.25:0.75 for the latter. The cubic simulation box had an edge length of 5 nm in each case.
The same simulation parameters that had been used for the coarse-grained isolated fragments were used here, with the following exceptions. The equilibration procedure included an NVT and an NPT equilibration, lasting 1 ns, respectively. The Parrinello-Rahman barostat was used for the latter, the same as for the production run, which lasted 12 ns. Equations of motions were integrated with the leap-frog stochastic dynamics integrator. The interaction of the fragments with their environment

was gradually decoupled, with equidistant steps of $\Delta\lambda = 0.05$. The aforementioned equilibrium procedure was carried out for each individual lambda step. Since only neutral fragments were considered, van der Waals interactions were the only type of interaction present in the system. The corresponding potentials were modified with "soft-core" potentials, using a soft-core parameter of $\alpha = 0.5$. Finally, the transfer free energies were calculated from the difference between two solvation free energies: $\Delta G_{S1\to S2} = \Delta G_{S1\to\emptyset} - \Delta G_{S2\to\emptyset}$. The solvation-free energies were calculated with the help of the Bennett acceptance ratio method[58], using the GROMACS tool *gmx bar*.

2.3.4 Coarse-grained LHA models

The initial coordinates of the SOM structure and calcium ions were obtained by mapping the final output structure of the atomistic simulations to the coarse-grained representation. The structure was solvated, firstly with regular-sized water, and secondly with tiny-sized water, using the GROMACS tool *gmx solvate*. The number of water beads was chosen so that it corresponded to the same water concentration as in the respective atomistic model. This was done for each of the three different LHA models and for the different water contents, respectively. The equilibration procedures and simulation parameters were the same as for the isolated coarse-grained fragments, except that a time step of 10 fs and a pressure coupling constant of $\tau_p = 12.0$ ps were used for all the simulations. Furthermore, the simulation time of the production run was set to 500 ns.

3 Results & Discussion

3.1 Parametrization of isolated fragments

In Figure 2 the mean values of the radius of gyration, R_g, from the atomistic fragments are plotted against the mean values from their coarse-grained counterparts. The error bars are the corresponding standard deviations, which should be seen as a measure of the fluctuation

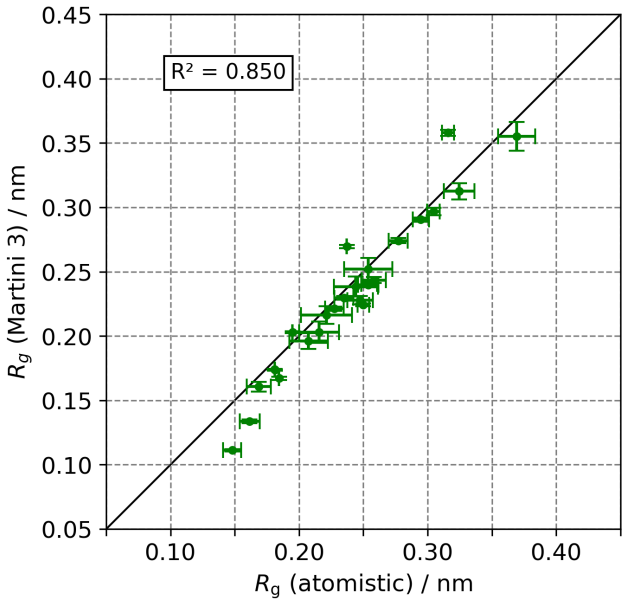

Figure 2: Comparison of time-averaged R_g of atomistic fragments with their coarse-grained counterparts (Martini 3); error bars according to standard deviations.

of R_g over time. Overall, a good correlation between the two quantities is visible, with minor deviations at the lower and higher range of values. In general, it is to be expected that lower values of R_g will deviate more from the reference, due to the relatively low number of beads for small molecules, thus reducing the accuracy of R_g compared to the reference. The results shown in Figure 2 partially validated the choice of mapping and the decision to conduct the parametrization using Swarm-CG, which resulted in a good reproduction of the overall size of each fragment.

In Figure 3, the mean SASA values of the atomistic fragments are compared to their coarse-grained counterparts, with the corresponding standard deviations. The coefficient of determination R^2 is close to 1, which indicates that there is a good correlation between the values of the atomistic and coarse-grained fragments, similar to the R_g case ($R^2 = 0.850$). A slight systematic overestimation of the SASA of the coarse-grained fragments is visible only among the higher values. The results indicate that, overall, there is a good reproduction of the fragment's molecular shape. This implies that although Swarm-CG does not use the SASA as

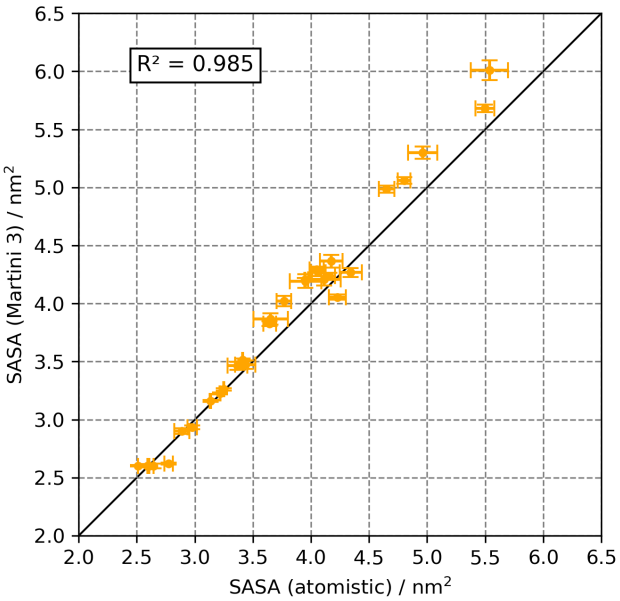

Figure 3: Comparison of time-averaged SASA of atomistic fragments with their coarse-grained counterparts (Martini 3); error bars according to the standard deviation.

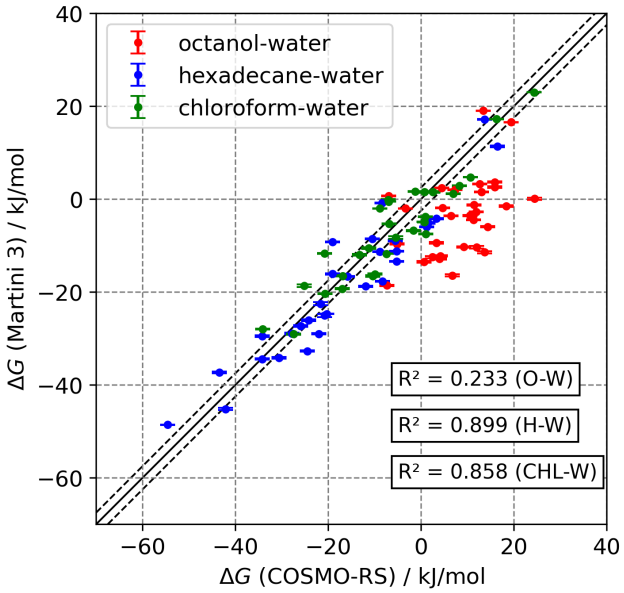

Figure 4: Comparison of octanol-water, hexadecane-water, and chloroform-water transfer free energies ΔG of coarse-grained fragments (Martini 3) with the reference. ΔG for coarse-grained fragments were determined from the difference between two solvation-free energies. Reference values were determined via partition coefficients from COSMO-RS.

a target for the parametrization, matching the distributions of bonded interactions can enable a good reproduction of this quantity. In contrast to these results, the parametrization of carbohydrates for the Martini 3 force field initially led to a systematic underestimation of the SASA[59]. This was resolved by scaling the bond distances by 15 %. In Figure S2 in the supporting information, the SASA values of the coarse-grained fragments are compared with values from the mapped atomistic fragments and the atomistic fragments, using van der Waals radii from the GROMOS force field. Again, very good correlations are visible for both cases.

In Figure 4, transfer free energies for each fragment are shown. The free energies from the coarse-grained fragments are plotted against reference values from COSMO-RS[34–36]. Here, dotted lines represent a free energy deviation from the reference of $k_BT \approx 2.479\,\mathrm{kJ\,mol^{-1}}$. Points between these lines represent a very good agreement between the calculated and reference-free energy values[60]. Matching transfer free energies is one important criterion for accurately parametrizing molecules for the Martini force field[31,59–61]. In principle, this could have been done with automated tools[37,38] using the previous version of the force field, which takes octanol-water transfer free energies as a target. Molecules that were parametrized in the main publication on Martini 3 were tested against transfer free energies between several solvents and water, namely octanol, hexadecane, and chloroform[31]. This improves the parametrization of a molecule with respect to its wide range of chemical properties. Following this approach and because experimental values were not available for each fragment, reference values were calculated by using COMSO-RS.

Overall, a good correlation is visible for the hexadecane-water and chloroform-water values, with a similar R^2 value. In contrast, the octanol-water values show a low correlation. To further investigate the octanol-water case, reference values were recalculated using ALOGPS[62], a neural network that can predict octanol-water partitioning coefficients. These values are shown in Figure S3 in the supporting information. Compared to the COSMO-RS val-

ues, they result in a slightly better correlation ($R^2 = 0.363$), but it is still relatively low in contrast to the hexadecane-water and chloroform-water values.

Of the three solvents considered for the transfer free energies, octanol is the least hydrophobic. As a result, the range of free energy values is narrower. In addition, many of the molecules considered are similar in terms of the functional groups present, which also results in a smaller range of free energy values. This could explain the lower correlation of the octanol-water values compared to the values with the other solvents. It should also be noted that methods such as COSMO-RS can only provide limited accuracy. They are suitable for analyzing general trends in the free energies for different solvents, but could also contribute to additional scattering of the values.

In general, the match with the reference could be improved by slightly adjusting the choice of beads. However, for the current set of molecules, this proves to be a time-consuming task. Also, not only the free energy should be taken into account, but also the exact properties of the functional groups, which can vary depending on the individual structure of a molecule (e.g. change of polarity of Ketone groups in caffeine[60]). It is assumed that, despite deviations in the free energy values, the standard assignment used in this work provides an adequate initial description of the molecules. Any adjustments regarding the assignment of the beads could be addressed in future studies.

3.2 Coarse-grained LHA model

3.2.1 LHA system densities

To investigate the performance of the parametrization approach with respect to particle densities, three different LHA models are considered, to take into account the effect of different compositions. The only difference in the input parameters of the VSOMM2 among the models was the random seed. The system's densities were calculated with the help of the GROMACS tool *gmx energy*. Because standard masses of Martini were used for the

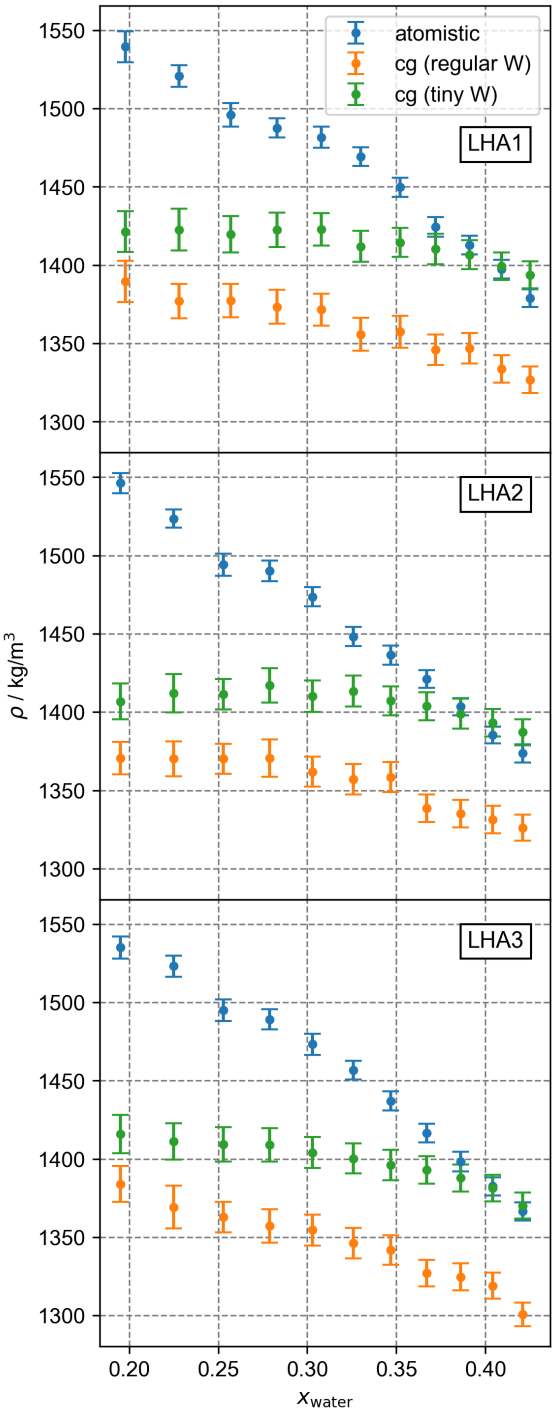

Figure 5: Time averaged system densities of the three LHA models, respectively, for different water contents. For each system, densities of the atomistic model (blue), coarse-grained model with regular water (orange), and coarse-grained model with tiny water (green) are plotted. Error bars indicate standard deviations of the respective values.

coarse-grained model, the densities were estimated with the time evolution of the system's volume, to enable a better comparison between the atomistic and coarse-grained systems. The time average of the densities was determined by dividing the total mass of the corresponding atomistic system by the time-averaged volume. Thus, an atomistic and a coarse-grained system with the same composition and the same volume would yield an equal density.

In Figure 5, the densities of the atomistic systems, coarse-grained systems with tiny water, and coarse-grained systems with regular water are plotted against x_{water}, which is the heavy atom fraction of water with respect to the whole system. For the atomistic systems, overall, the densities decreased with increasing water content. Changes in densities were less pronounced for the coarse-grained systems and relatively similar when the different compositions were compared. When comparing the densities of the atomistic and coarse-grained systems, some deviations were apparent and more pronounced in the case of lower water content. The largest discrepancies between the atomistic and coarse-grained systems were visible when using regular water, with a density deviation of around 11 %. This gap becomes reduced with increasing water content. For the water range considered, the coarse-grained systems with regular water came closest to the reference densities at the highest water content. In the case of tiny water, the densities approached atomistic density at a higher water content, achieving a good level of agreement, but slightly overestimating the density.

The atomistic systems with low water content are assumed to be more rigid. To evaluate the relaxation process of a structure, the time evolutions of the densities were calculated. The densities needed longer to equilibrate for systems with a lower water content. In the case of the LHA1 model with $x_{\text{water}} = 0.228$, the equilibration time had to be extended by 100 ns. This was done similarly for the LHA2 model with $x_{\text{water}} = 0.195$ and $x_{\text{water}} = 0.225$, whereby the equilibration time was extended by 400 ns. When comparisons were made between different compositions of the coarse-grained systems, the density curves had greater similarities. Presumably, due to the coarse-graining process, the finer details and characteristics of the different systems are averaged out, making the systems behave more similarly. Deviations between the atomistic and coarse-grained models are presumed to be caused by the inability of coarse-grained water to accurately fill in small voids, which are more prevalent with lower water contents. This could partially explain the smaller gap between the models with tiny water, although the overall density trend was not reproduced in this case. The tiny water model could also show a slight overestimation of the density when the water content was increased further. It is, however, expected that, when x_{water} approaches 1, the densities will match as well as the coarse-grained water model reproduces the density of atomistic bulk water. Therefore, the highest deviations are presumed to be visible in the range of x_{water} considered in this research, or when considering even drier SOM models.

Models with higher water contents should be considered, for instance when the coarse-grained model is parametrized to reproduce the atomistic system's density. To achieve this, the non-bonded interactions could be adjusted, as they may have a significant influence on the overall density (see, for instance, Lutsyk et al.[63]). However, the strong heterogeneity of the SOM systems could make this kind of matching a challenging task. It could therefore be investigated in future research.

Finally, it can be seen that the density of systems with higher water content was reproduced more accurately. It should also be noted that the systems with small and tiny water should be used with caution, as their transferability to different systems was not as extensively tested as for those with regular water[64]. Therefore, for the remainder of this study, systems with regular water are further considered.

3.2.2 DBI approach performance

For the other properties that were considered in the following, only one system (LHA1) was used to present the results, as there were no signifi-

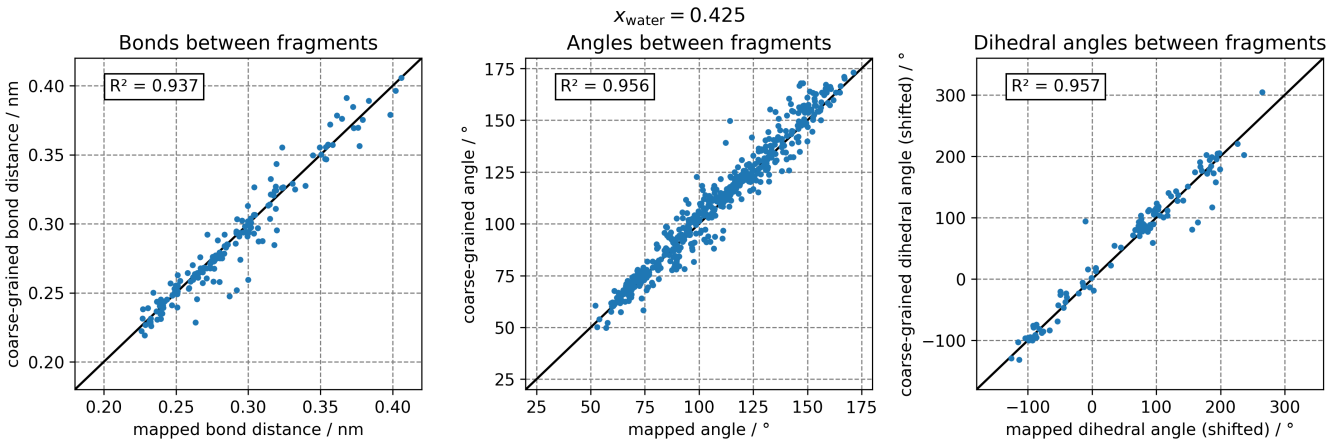

Figure 6: Comparison of average values of bonded interactions between a coarse-grained LHA structure and the mapped version of its atomistic counterpart. Here, the LHA1 model with $x_{\text{water}} = 0.425$ is considered. Bonded interactions are divided into bonds, angles, and dihedral angles between different fragments.

cant differences when the different compositions were compared. As mentioned earlier, parameters for bonded interactions between different fragments were calculated using DBI. Further details about this approach can be found in section S1 of the supporting information.

Figure 6 shows the average values of the time evolutions of the bonds, angles, and dihedral angles between fragments. For the water content $x_{\text{water}} = 0.425$, the average values for each bonded interaction are compared between the mapped and coarse-grained systems. The same plots for models with different water contents are shown in Figure S5. No significant differences or trends are visible when comparing models with different water contents. To achieve a better representation, the standard deviations are not shown. A good correlation is visible for all the bonded interactions. For some distributions of dihedral angles that run periodically over the $[-180°, 180°]$ interval, the deviations of the mean from the reference can be very large. Therefore, to achieve a better comparison, 360° were added to the negative dihedral angles in such cases.

Generally, the correlations show the applicability of DBI to these models, although only the average values are shown here, which does not yield information about the form of the potentials. Deviations could mainly be caused by cross-correlations between the bonded interactions, which are neglected using this approach. To achieve further fine-tuning, iterative matching could be applied, which would conversely increase the parametrization effort for each SOM model. In general, a trade-off between the parametrization effort, the simplicity of the model, and its performance has to be made. The current parametrization of bonded interactions can be regarded as achieving a good compromise in this regard. In this case, more accurate parametrized fragments are used as basic building blocks, with fixed parameters for the bonded interactions, and less accurate parametrized interactions between the fragments, which are relatively easy to obtain. Ideally, if they were available, generic bonded interaction parameters between the fragments would be applied, so that they would be completely independent of an atomistic trajectory. We are currently working on a tool that can convert atomistic VSOMM2 models into a Martini 3 representation using generic bonded interaction parameters, without the need for a reference trajectory.

3.2.3 Calcium-carboxylate configuration analysis

Reconsidering the non-bonded interactions, as mentioned previously, the bead types for the ions and charged groups are mostly fixed for

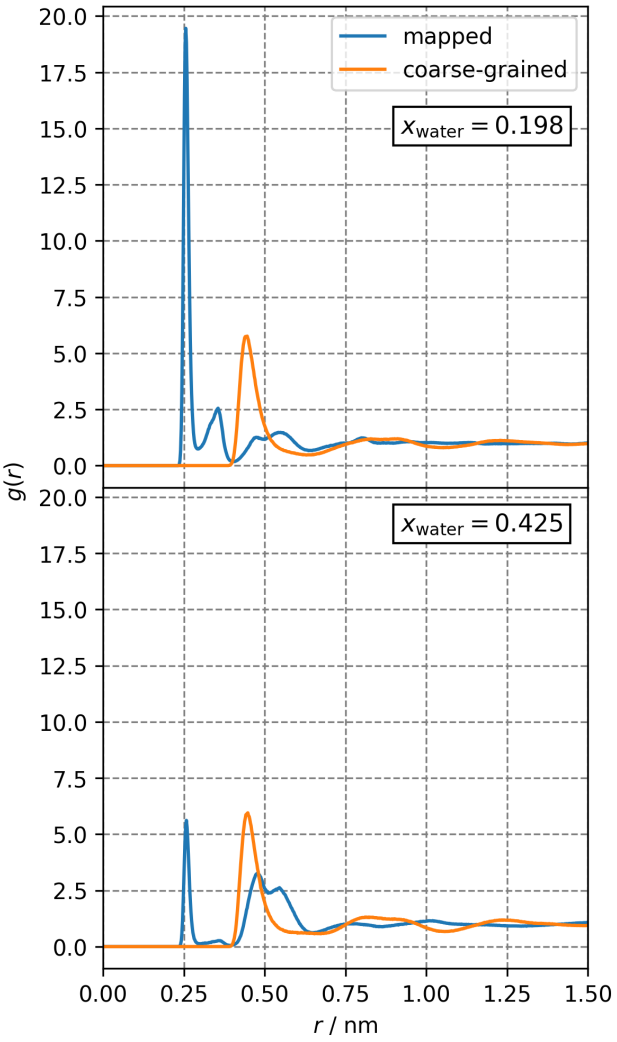

Figure 7: Comparison between mapped (blue) and coarse-grained (orange) RDFs, which were calculated between carboxylate groups and calcium ions. Here, the LHA1 model for $x_{water} = 0.198$ and for $x_{water} = 0.425$ is shown.

the Martini force field. For the SOM models, the calcium ions were represented by SD beads and the negatively charged carboxylate groups by SQ5n beads. These are the respective types of standard bead types.

The aim was to investigate how these parametrizations of ions reproduce properties of the reference. This was done by calculating the radial distribution functions (RDFs) between the carboxylate groups and calcium ions, which are shown in Figure 7. In this case, the SOM models with the lowest and highest water content were considered. The RDFs from the mapped trajectories (blue) show a peak at around 0.26 nm, which is more pronounced for the system with low water content. For the mapped system with high water content, a second broader peak is visible around 0.5 nm. The coarse-grained RDFs (orange) are similar for both cases, showing a peak at around 0.45 nm. However, a slightly better agreement between the mapped and coarse-grained RDF was observed for the case with high water concentration. Any peaks which are lower than 0.40 nm, are not captured by the coarse-grained model. The form of the RDFs is mainly determined by the potential, describing the interaction between both groups. In the present case, the coarse-grained potential is a combination of a Lennard-Jones and a Coulomb potential. Therefore, it is inherently not possible to reproduce a double peak as in the case with high water content. The peaks at lower distances cannot be captured by the coarse-grained model, due to the predefined particle size of the beads. In general, to accurately reproduce RDF curves, it would be necessary to apply tabulated potentials, describing the non-bonded interactions. One possible method, that involves conducting such matching, is the so-called Iterative Boltzmann Inversion[65]. However, applying this method is beyond the scope of this study. The position of the mapped RDF in the high water case was partially reproduced by the coarse-grained RDF's first peak. Concerning this, it could again be argued that the model captures the features of the reference at higher water content more effectively, similarly to the system density. It can also be seen that the

coarse-grained RDF is less affected by the water content as the reference. This might be an indication of the coarse-grained model's limited transferability, regarding different water concentrations in the system.

To better understand the form of the mapped RDFs, angle distributions of the calcium ions and oxygen atoms of carboxylate groups were calculated. The atomistic systems with the lowest and highest water content were again considered. The results of the angle calculations are shown as distributions in Figure 8(a). Most notably, three peaks can be identified, which have been labeled as peak I, II, and II respectively. These peaks are roughly visible in the range from 0° to 20°, from 20° to 30° and from 40° to 55°, respectively.

Additionally, the center of geometry distances between the calcium ions and carboxylate groups were calculated (see Figure 8(b)), and associated with each of the three peaks from the angle distributions. For instance, for configurations of calcium ions and carboxylate groups associated with peak III, the center of geometry distances were calculated for those configurations with an angle between 40° and 55° and plotted for both water contents. This was also done with the configurations corresponding to the other two peaks, using the corresponding angle ranges. This resulted in a total of six different curves.

From the angle distributions, it can be ascertained that the configurations associated with peak III are more often observed in the case of lower water content. This also holds for peak I configurations, but the difference to the system with the lower water content is less pronounced. In contrast, peak II configurations are more pronounced in the system with the higher water content. Peak III configurations can be associated with lower distances, visible in the peaks at around 0.26 nm. The calcium ions are very close to the oxygen atoms of the carboxylate group, which results in a configuration that is assumed to be highly stable.

Galicia-Andrés et al.[2] investigated the role of metal cations on the stability of SOM with atomistic simulations. From the results, the authors concluded that cations such as calcium

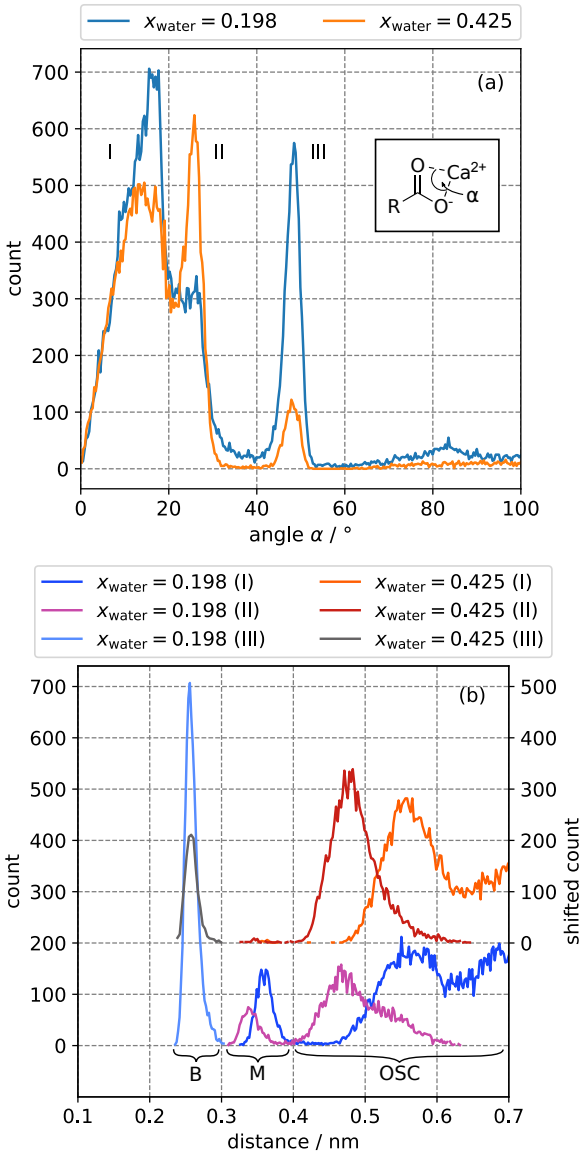

Figure 8: (a) Angle distributions between calcium ions and oxygen atoms of carboxylate groups, for lowest and highest water content considered. Visible peaks are divided into peaks I, II, and III. A cutoff radius of 0.7 nm between the calcium ions and the oxygen atoms was used for the angle calculations. (b) Center of geometry distance between calcium ions and carboxylate groups, associated with the different configurations and different water contents. Different peaks can be assigned to specific complexes, which are inner-sphere complexes (monodentate (M) and bidentate (B)) and outer-sphere complexes (OSC). Please note that, to achieve a better representation, the curves for $x_{water} = 0.425$ are shifted by a count of 200.

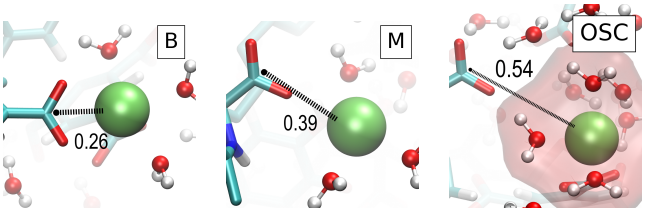

Figure 9: Trajectory snapshots for $x_{\text{water}} = 0.198$. Inner-sphere complexes (bidentate (B) and monodentate (M)), and an outer-sphere complex (OSC) are shown from left to right. Calcium ions are represented as green particles, LHA molecules as sticks, and water molecules as balls and sticks. Oxygen, carbon, nitrogen, and hydrogen atoms are shown in red, cyan, blue, and white, respectively. Center of geometry distances between carboxylate groups and calcium ions (black) are given in nanometers.

ions are important for the SOM agglomeration process. Due to the presence of calcium ions, more dense SOM phases are created, which increases the overall stability of the SOM system. Comparing further the results to the work of Galicia-Andrés et al.[2], monodentate (M) inner sphere complex and a bidentate (B) inner sphere complex were investigated there. In addition, an outer-sphere complex (OSC) was observed. In the present context, an M complex is formed when calcium interacts with the carboxylate groups using one oxygen as a donor site. Similarly, in a B complex, calcium ions interact with two oxygen atoms, which act as two donor sites. In an OSC, the calcium ions are surrounded by a monolayer of water, resulting in a greater distance between the ions and the carboxylate groups.

Projecting these observations to the present study, the complexes can be identified via the center of geometry distance in Figure 8(b). Peak III configurations can be uniquely assigned to the B complexes. The M complexes can be identified with the peaks in the region from 0.32 to 0.39 nm, which are only present for $x_{\text{water}} = 0.198$. Configurations with a center of geometry distances bigger than 0.40 nm can be assigned to OSCs. For the $x_{\text{water}} = 0.198$ case, complexes are shown as trajectory snapshots in Figure 9. For further details about the complexes, please refer to Galicia-Andrés et al.'s study.

In light of these complexities, it is possible to understand the limitations of the coarse-grained model. The substructure is averaged out, leaving no possibility of reproducing these configurations with a single pair potential. The fixed bead size, which results in an overestimation of the distance between the calcium ions and carboxylate groups, could contribute to the underestimation of the coarse-grained SOM density. The aggregations may play a less important role for systems with higher water content, which could, on the other hand, explain the lesser discrepancy between the densities in these cases. It is natural for a coarse-grained model to lose features that are more connected to many-body interactions. When using the coarse-grained SOM model, it is essential to clarify whether reproducing such features is critical when observing the quantities of interest.

3.2.4 Power spectrum analysis

As previously mentioned, a time step of 10 fs was used, when simulating the coarse-grained LHA models. This was done because the occurrence of instabilities at 20 fs, which is commonly used for Martini simulations, and higher time steps, made it impossible to observe a trajectory with a sufficient simulation time, such as 500 ns. A more detailed investigation with a high output of frames from the trajectory revealed the parts of the system that are most likely to be responsible for this, for instance, the ones that were connected to the HS30 fragment. These parts showed a tendency towards instability, due to an extreme rocking of the beads within a relatively short time. It can be regarded as a relatively rare event but makes it impractical to run long simulations at a larger time step.

To get closer to the source of the problem, power spectra for isolated fragments in water were calculated from the velocity autocorrelation function. By doing so, emerging frequencies within one fragment become visible. As

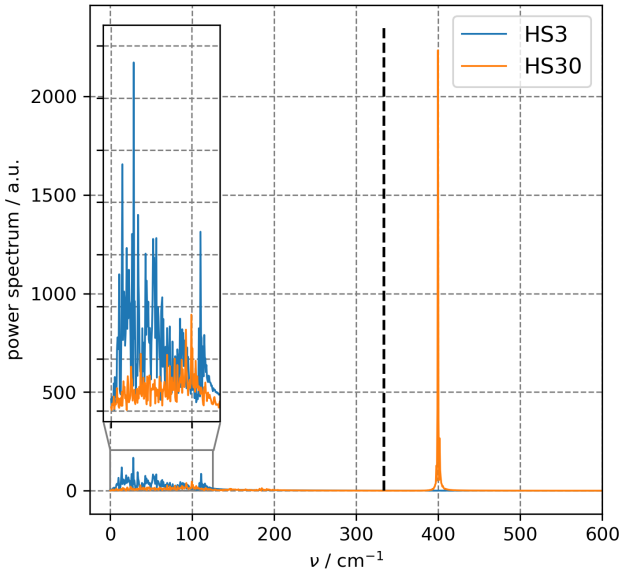

Figure 10: Power spectra of isolated fragments in water, calculated via the velocity autocorrelation function. As examples, the spectra for the fragments HS3 and HS30 are plotted. The dashed line denotes a frequency, with a time period of 5 times 20 fs. This frequency can be used as a guideline to verify which frequencies can be adequately simulated by a time step of 20 fs.

examples, power spectra of the fragments HS3 and HS30 are shown in Figure 10. The frequency range from 0 to 125 cm^{-1} is also shown zoomed-in. For the HS30 fragment, a high peak at around 400 cm^{-1} is visible. The dashed vertical line corresponds to a frequency with a time period being 5 times 20 fs long, which is further explained below. Frequencies in the zoomed-in region can be identified as noise, which looks very similar to other fragments. Due to the coarse-graining, the characteristic vibrational lines of the molecules are averaged out, leaving a noisy signal behind, which is determined by the parameters of the bonded interactions. The peak of the HS30 fragment, however, is one of the strongest ones to emerge for all the fragments, as well as the one with the highest frequency. For proper integration with the leap-frog integrator, the time step should be at least 5 times shorter than the period of the fastest oscillation in the system[66]. Following this guideline, the dashed line therefore represents the maximum frequency that can be resolved with a time step of 20 fs, which is lower than the maximum frequency of HS30. Due to the correlations with instabilities with the HS30 fragment, it is therefore assumed that the high frequencies play an important role. The interplay with the bonded interactions between fragments should also not be neglected. Generally, to allow the use of higher time steps (at least 20 fs) and therefore boost performance the bonded interactions of the fragments should be recalibrated, to lower the highest emerging frequencies. This is most likely connected to the interplay between the bond and angle parameters within the fragment. However, here the parameters were not changed, to ensure a systematic parametrization procedure. As part of the current development of the VSOMM2 conversion tool, the force constants of the bonded interactions are being revised to reduce such high frequencies.

3.2.5 Hydrogen bond analysis

In addition to electrostatic and weak intermolecular interactions, hydrogen bonding interactions are an important characteristic of HS

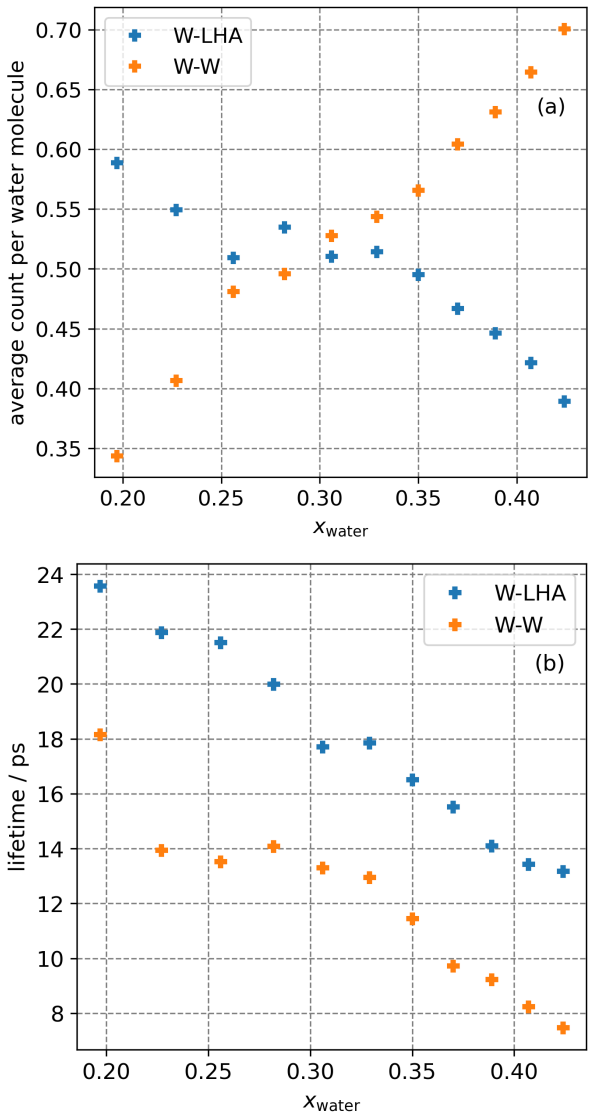

Figure 11: (a) Average number of hydrogen bonds per water molecule in the system, depending on the water content. Distinctions were made between hydrogen bonds between the LHA molecules and water (W-LHA), and hydrogen bonds between the water molecules themselves (W-W). (b) Hydrogen bond lifetime, fitted from an exponential fit of the hydrogen bond autocorrelation function. Again, different water contents are considered, as well as being differentiated into W-LHA and W-W hydrogen bonding.

systems, as they determine their properties and play a role in stabilizing them[67,68]. In the case of the LHA models, hydrogen bonding was previously analyzed via atomistic simulations[18,20]. Here, hydrogen bonds between water and LHA molecules (W-LHA), as well as water-water hydrogen bonding (W-W) are considered. Hydrogen bonds were analyzed with the hydrogen bond analysis tool of MDAnalysis. The criteria for an existing hydrogen bond were a maximum of 3 Å in donor-acceptor distance and a minimum angle of 150° between the donor, hydrogen, and acceptor atoms.

In Figure 11(a), the average number of hydrogen bonds per water molecule is plotted for both cases. Values were calculated by taking the time average of all the hydrogen bonds that occurred between the two groups considered and dividing it by the total number of water molecules in the system. The total number of hydrogen bonds averaged over time is shown in Figure S6 in the supporting information. Regarding the W-LHA hydrogen bonding in Figure 11(a), the average count per water molecule decreased with increasing water content. The opposite trend is apparent for the W-W case. The trends for the average counts behave in an expected way. More water molecules in the system decrease the relative amount of those molecules, which contribute to the hydrogen bonding to the LHA molecules. Conversely, the amount of hydrogen bonds between the water molecules increases. The results show a similar trend to that observed in Petrov et al.'s[18] investigations.

The hydrogen bond lifetime was calculated via the hydrogen bond autocorrelation function. Similar lifetime calculations have previously been undertaken in Smith et al.'s[69] work, examining the adsorption of hyaluronic acid by lipid membranes. First, the correlation functions of the hydrogen bonds were calculated. A distinction was again made between W-LHA and W-W. The correlation function curve was then fitted using a biexponential function. The lifetime was finally obtained from the integral over the fit function. More details about such calculations can be found in the aforementioned study.

Regarding Figure 11(b), overall, the hydrogen bond lifetime decreases with increasing water content. Comparing the lowest and highest water content on the x_{water} scale, the lifetime is nearly halved. These results suggest that hydrogen bonds have a greater effect on the molecule's behavior at lower water content. This is a feature which presumably cannot be reproduced by the coarse-grained model. Therefore, it could prove difficult to achieve the transferability of the coarse-grained model to all different water contents with equal accuracy. Especially at lower water contents, hydrogen bonding might become more important.
Hydrogen bonding is captured in the Martini force field with labels, denoting donor or acceptor behavior of atom groups, resulting in more or less attractive interactions. However, the orientational dependence of such interactions is not reproduced, as this is a natural feature, resulting from the coarse-graining process. Regarding the current state of the model, it is recommended to work with water contents for which the coarse-grained system already performs relatively well at reproducing the reference properties.

4 Conclusions and Outlook

This work has presented an approach to provide a coarse-grained SOM model using the Martini 3 force field. The coarse-grained models were based on atomistic models of LHA generated by the VSOMM2. The VSOMM2 builds up models of HS by combining fragment molecules into a macromolecular structure.
In the first part, the fragment molecules were converted into a coarse-grained representation. This was realized by applying the standard Martini bead assignment to the fragment molecules, for the non-bonded interactions. The Swarm-CG tool was used to determine the bonded interaction parameters. Each fragment was put into water boxes, and parametrized, using the trajectories of atomistic, isolated fragment molecules as a reference.
From these simulations, the R_{g} and the SASA were compared with those of the atomistic fragment molecules. Furthermore, the transfer free energies of the coarse-grained fragment molecules were calculated using free energy calculations and compared with the values determined by the COSMO-RS method. The R_{g} and the SASA showed an overall good agreement with the reference, indicating a sufficient reproduction of the shape and size of the molecules. The transfer free energies showed a good reproduction for the hexadecane-water and the chloroform-water cases.
In the second part, coarse-grained models of LHA were constructed and compared with atomistic reference systems, provided by the VSOMM2. Three models were analyzed, each with varying water content, and using calcium as counter ions. Bonded interaction parameters between the fragment molecules were obtained by the DBI method, using the trajectories of the reference systems.
The densities of the coarse-grained models were in good agreement with the reference systems, especially at higher water contents. The automatic assignment of the bonded interactions via DBI reproduced the average values of the bonded interactions with sufficient accuracy, which is not dependent the on water content, at least as far as those considered in this research.
RDFs between carboxylate groups and calcium ions were calculated. As a reference, RDFs between these two compounds were also determined from the mapped atomistic systems. A better agreement was observed for the higher water content case. Regions with small distances were not reproduced by the coarse-grained model, due to the fixed size of the beads involved. The angles between the oxygen atoms of the carboxylate groups and the calcium ions were analyzed and plotted as distributions. This showed that the calcium ions formed different complexes with the LHA molecules. This is in agreement with the observations of a similar work by Galicia-Andrés et al.[2]. Hydrogen bonds were analyzed for the atomistic models. This showed, among other results, an increased hydrogen bond lifetime for systems with a lower water content.
For further investigations, the reference of the

SASA could be replaced with one, calculated from simulations with an all-atom representation of the fragments. The overall correlation of the transfer free energy values with the reference could be improved by changing the choice of beads for each fragment. This could be particularly useful for the octanol-water case.
Concerning the DBI approach, trajectories are required to generate a coarse-grained model. To make such models more easily available, we are working on a tool that generates systems of coarse-grained SOM models and does not require a trajectory. A frequency analysis of the individual fragments also showed that relatively high frequencies occur in the current parametrization, which limits the use of a longer time step. By revising the force constants, these frequencies can be reduced. Such a procedure will be part of the development of the aforementioned tool.
The hydrogen bond analysis revealed additional features of the atomistic model, which may be more significant at lower water contents. In general, the accuracy of the model, concerning the properties such as density, RDF, and hydrogen bonds is lower, compared to higher water contents. Also, different complexes observed in the atomistic model were not inherently captured by the coarse-grained model. Therefore, when using this model, it should be ensured that the observables of interest are not directly connected to these quantities.
In conclusion, this study has provided a comprehensive understanding of the properties, strengths, and weaknesses inherent in the coarse-grained SOM model. Serving as a robust foundation for further research, this model offers opportunities to explore the diverse compositions of SOM. Moreover, it sets the stage for investigating interactions with soil substances and pollutants within the Martini 3 force field framework. This comprehensive analysis marks a significant step forward in our understanding of SOM dynamics and its broader environmental implications.

References

(1) Blume, H.-P.; Brümmer, G. W.; Fleige, H.; Horn, R.; Kandeler, E.; Kögel-Knabner, I.; Kretzschmar, R.; Stahr, K.; Wilke, B.-M. *Scheffer/SchachtschabelSoil Science*; Springer Berlin Heidelberg, 2015; pp 55–86.

(2) Galicia-Andrés, E.; Escalona, Y.; Oostenbrink, C.; Tunega, D.; Gerzabek, M. H. Soil organic matter stabilization at molecular scale: The role of metal cations and hydrogen bonds. *Geoderma* **2021**, *401*, 115237.

(3) Blume, H.-P.; Brümmer, G. W.; Fleige, H.; Horn, R.; Kandeler, E.; Kögel-Knabner, I.; Kretzschmar, R.; Stahr, K.; Wilke, B.-M. *Scheffer/SchachtschabelSoil Science*; Springer Berlin Heidelberg, 2015; pp 485–559.

(4) Jones, K.; de Voogt, P. Persistent organic pollutants (POPs): state of the science. *Environmental Pollution* **1999**, *100*, 209–221.

(5) Guo, J.-J.; Huang, X.-P.; Xiang, L.; Wang, Y.-Z.; Li, Y.-W.; Li, H.; Cai, Q.-Y.; Mo, C.-H.; Wong, M.-H. Source, migration and toxicology of microplastics in soil. *Environment International* **2020**, *137*, 105263.

(6) Pérez-Reverón, R.; Álvarez-Méndez, S. J.; González-Sálamo, J.; Socas-Hernández, C.; Díaz-Peña, F. J.; Hernández-Sánchez, C.; Hernández-Borges, J. Nanoplastics in the soil environment: Analytical methods, occurrence, fate and ecological implications. *Environmental Pollution* **2023**, *317*, 120788.

(7) Zwolak, A.; Sarzyńska, M.; Szpyrka, E.; Stawarczyk, K. Sources of Soil Pollution by Heavy Metals and Their Accumulation in Vegetables: a Review. *Water, Air, & Soil Pollution* **2019**, *230*.

(8) Cycoń, M.; Mrozik, A.; Piotrowska-Seget, Z. Antibiotics in the Soil Environment—Degradation and Their Impact on Microbial Activity and Diversity. *Frontiers in Microbiology* **2019**, *10*.

(9) Kubicki, J.; Apitz, S. Models of natural organic matter and interactions with organic contaminants. *Organic Geochemistry* **1999**, *30*, 911–927.

(10) Schulten, H.-R.; Thomsen, M.; Carlsen, L. Humic complexes of diethyl phthalate: molecular modelling of the sorption process. *Chemosphere* **2001**, *45*, 357–369.

(11) Aquino, A. J. A.; Tunega, D.; Haberhauer, G.; Gerzabek, M. H.; Lischka, H. Interaction of the 2,4-dichlorophenoxyacetic acid herbicide with soil organic matter moieties: a theoretical study. *European Journal of Soil Science* **2007**, *58*, 889–899.

(12) Ahmed, A. A.; Thiele-Bruhn, S.; Aziz, S. G.; Hilal, R. H.; Elroby, S. A.; Al-Youbi, A. O.; Leinweber, P.; Kühn, O. Interaction of polar and nonpolar organic pollutants with soil organic matter: Sorption experiments and molecular dynamics simulation. *Science of The Total Environment* **2015**, *508*, 276–287.

(13) Gros, P.; Ahmed, A.; Kühn, O.; Leinweber, P. Glyphosate binding in soil as revealed by sorption experiments and quantum-chemical modeling. *Science of The Total Environment* **2017**, *586*, 527–535.

(14) Böhm, L.; Grančič, P.; Scholtzová, E.; Heyde, B. J.; Düring, R.-A.; Siemens, J.; Gerzabek, M. H.; Tunega, D. Adsorption of the hydrophobic organic pollutant hexachlorobenzene to phyllosilicate minerals. *Environmental Science and Pollution Research* **2022**, *30*, 36824–36837.

(15) Dettmann, L. F.; Kühn, O.; Ahmed, A. A. Coarse-grained molecular dynamics simulations of nanoplastics interacting with a hydrophobic environment in aqueous solution. *RSC Advances* **2021**, *11*, 27734–27744.

(16) Sündermann, A.; Solc, R.; Tunega, D.; Haberhauer, G.; Gerzabek, M. H.; Oostenbrink, C. Vienna Soil-Organic-Matter Modeler—Generating condensed-phase models of humic substances. *Journal of Molecular Graphics and Modelling* **2015**, *62*, 253–261.

(17) Escalona, Y.; Petrov, D.; Oostenbrink, C. Vienna soil organic matter modeler 2 (VSOMM2). *Journal of Molecular Graphics and Modelling* **2021**, *103*, 107817.

(18) Petrov, D.; Tunega, D.; Gerzabek, M. H.; Oostenbrink, C. Molecular Dynamics Simulations of the Standard Leonardite Humic Acid: Microscopic Analysis of the Structure and Dynamics. *Environmental Science & Technology* **2017**, *51*, 5414–5424.

(19) Petrov, D.; Tunega, D.; Gerzabek, M. H.; Oostenbrink, C. Molecular modelling of sorption processes of a range of diverse small organic molecules in Leonardite humic acid. *European Journal of Soil Science* **2019**, *71*, 831–844.

(20) Escalona, Y.; Petrov, D.; Oostenbrink, C. Modeling soil organic matter: Changes in macroscopic properties due to microscopic changes. *Geochimica et Cosmochimica Acta* **2021**, *307*, 228–241.

(21) Escalona, Y.; Petrov, D.; Galicia-Andrés, E.; Oostenbrink, C. Exploring the Macroscopic Properties of Humic Substances Using Modeling and Molecular Simulations. *Agronomy* **2023**, *13*, 1044.

(22) Gotsmy, M.; Escalona, Y.; Oostenbrink, C.; Petrov, D. Exploring the structure and dynamics of proteins in soil organic matter. *Proteins: Structure, Function, and Bioinformatics* **2021**, *89*, 925–936.

(23) Galicia-Andrés, E.; Oostenbrink, C.; Gerzabek, M. H.; Tunega, D. On the Adsorption Mechanism of Humic Substances on Kaolinite and Their Microscopic Structure. *Minerals* **2021**, *11*, 1138.

(24) Harmandaris, V. A.; Adhikari, N. P.; van der Vegt, N. F. A.; Kremer, K. Hierarchical Modeling of Polystyrene: From Atomistic to Coarse-Grained Simulations. *Macromolecules* **2006**, *39*, 6708–6719.

(25) Joshi, S. Y.; Deshmukh, S. A. A review of advancements in coarse-grained molecular dynamics simulations. *Molecular Simulation* **2020**, *47*, 786–803.

(26) Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine Learning for Molecular Simulation. *Annual Review of Physical Chemistry* **2020**, *71*, 361–390.

(27) Darré, L.; Machado, M. R.; Brandner, A. F.; González, H. C.; Ferreira, S.; Pantano, S. SIRAH: A Structurally Unbiased Coarse-Grained Force Field for Proteins with Aqueous Solvation and Long-Range Electrostatics. *Journal of Chemical Theory and Computation* **2015**, *11*, 723–739.

(28) Machado, M. R.; Barrera, E. E.; Klein, F.; Sóñora, M.; Silva, S.; Pantano, S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *Journal of Chemical Theory and Computation* **2019**, *15*, 2719–2733.

(29) Orsi, M.; Essex, J. W. The ELBA Force Field for Coarse-Grain Modeling of Lipid Membranes. *PLoS ONE* **2011**, *6*, e28637.

(30) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B* **2007**, *111*, 7812–7824.

(31) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature Methods* **2021**, *18*, 382–388.

(32) Feng, H.; Zhang, H.; Cao, H.; Sun, Y.; Zhang, A.; Fu, J. Application of a Novel Coarse-Grained Soil Organic Matter Model in the Environment. *Environmental Science & Technology* **2018**, *52*, 14228–14234.

(33) Empereur-Mot, C.; Pesce, L.; Doni, G.; Bochicchio, D.; Capelli, R.; Perego, C.; Pavan, G. M. Swarm-CG: Automatic Parametrization of Bonded Terms in MARTINI-Based Coarse-Grained Models of Simple to Complex Molecules via Fuzzy Self-Tuning Particle Swarm Optimization. *ACS Omega* **2020**, *5*, 32823–32843.

(34) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *The Journal of Physical Chemistry* **1995**, *99*, 2224–2235.

(35) Pye, C. C.; Ziegler, T. An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package. *Theor. Chem. Acc.* **1999**, *101*, 396–408.

(36) Louwen, J.; Pye, C.; van Lenthe, E.; Austin, N.; McGarrity, E.; Xiong, R.; Sandler, S.; Burnett, R. AMS 2023.1 COSMO-RS. SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, 2023.

(37) Bereau, T.; Kremer, K. Automated Parametrization of the Coarse-Grained Martini Force Field for Small Organic Molecules. *Journal of Chemical Theory and Computation* **2015**, *11*, 2783–2791.

(38) Potter, T. D.; Barrett, E. L.; Miller, M. A. Automated Coarse-Grained Mapping Algorithm for the Martini Force Field and Benchmarks for Membrane–Water Partitioning. *Journal of Chemical Theory and Computation* **2021**, *17*, 5777–5791.

(39) Potter, T. D.; Haywood, N.; Teixeira, A.; Hodges, G.; Barrett, E. L.; Miller, M. A. Partitioning into phosphatidylcholine-cholesterol membranes: liposome measurements, coarse-grained simulations, and implications for bioaccumulation. *Environ. Sci. Process. Impacts* **2023**, *25*, 1082–1093.

(40) Rowland, R. S.; Taylor, R. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii. *The Journal of Physical Chemistry* **1996**, *100*, 7384–7391.

(41) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(42) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the Python in Science Conference. 2016.

(43) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **2011**, *32*, 2319–2327.

(44) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array programming with NumPy. *Nature* **2020**, *585*, 357–362.

(45) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **2020**, *17*, 261–272.

(46) Pérez, F.; Granger, B. E. IPython: a System for Interactive Scientific Computing. *Computing in Science and Engineering* **2007**, *9*, 21–29.

(47) Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **2007**, *9*, 90–95.

(48) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(49) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **2009**, *30*, 2157–2164.

(50) Poger, D.; Gunsteren, W. F. V.; Mark, A. E. A new force field for simulating phosphatidylcholine bilayers. *Journal of Computational Chemistry* **2009**, *31*, 1117–1125.

(51) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal* **2011**, *40*, 843–856.

(52) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *The Jerusalem Symposia on Quantum Chemistry and Biochemistry*; Springer Netherlands, 1981; pp 331–342.

(53) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007**, *126*.

(54) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, *52*, 7182–7190.

(55) Thorn, K. A.; Folan, D. W.; MacCarthy, P. Characterization of the International Humic Substances Society standard and reference fulvic and humic acids by solution state carbon-13 (13C) and hydrogen-1 (1H) nuclear magnetic resonance spectrometry. *Water-Resources Investigations Report* **1989**,

(56) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.

(57) de Jong, D. H.; Baoukina, S.; Ingólfsson, H. I.; Marrink, S. J. Martini straight: Boosting performance using a shorter cutoff and GPUs. *Computer Physics Communications* **2016**, *199*, 1–7.

(58) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **1976**, *22*, 245–268.

(59) Grünewald, F.; Punt, M. H.; Jefferys, E. E.; Vainikka, P. A.; König, M.; Virtanen, V.; Meyer, T. A.; Pezeshkian, W.; Gormley, A. J.; Karonen, M.; Sansom, M. S. P.; Souza, P. C. T.; Marrink, S. J. Martini 3 Coarse-Grained Force Field for Carbohydrates. *Journal of Chemical Theory and Computation* **2022**, *18*, 7555–7569.

(60) Alessandri, R.; Thallmair, S.; Herrero, C. G.; Mera-Adasme, R.; Marrink, S. J.; Souza, P. C. T. *A Practical Guide to Recent Advances in Multiscale Modeling and Simulation of Biomolecules*; AIP Publishing LLCMelville, New York, 2023; pp 1–1–1–34.

(61) Alessandri, R.; Barnoud, J.; Gertsen, A. S.; Patmanidis, I.; de Vries, A. H.; Souza, P. C. T.; Marrink, S. J. Martini 3 Coarse-Grained Force Field: Small Molecules. *Advanced Theory and Simulations* **2021**, *5*, 2100391.

(62) Tetko, I. V.; Tanchuk, V. Y. Application of Associative Neural Networks for Prediction of Lipophilicity in ALOGPS 2.1 Program. *Journal of Chemical Information and Computer Sciences* **2002**, *42*, 1136–1145.

(63) Lutsyk, V.; Wolski, P.; Plazinski, W. Extending the Martini 3 Coarse-Grained Force Field to Carbohydrates. *Journal of Chemical Theory and Computation* **2022**, *18*, 5089–5107.

(64) Marrink, S. J.; Monticelli, L.; Melo, M. N.; Alessandri, R.; Tieleman, D. P.; Souza, P. C. T. Two decades of Martini: Better beads, broader scope. *WIREs Computational Molecular Science* **2022**, *13*.

(65) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry* **2003**, *24*, 1624–1636.

(66) Apol, E.; Apostolov, R.; Berendsen, H. J.; Aldert Van Buuren,; P¨Ar Bjelkmar,; Drunen, R. V.; Feenstra, A.; Fritsch, S.;

Groenhof, G.; Junghans, C.; Hub, J.; Kasson, P.; Kutzner, C.; Lambeth, B.; Larsson, P.; Lemkul, J. A.; Marklund, E.; Peiter Meulenhoff,; Murtola, T.; Szil´Ard P´All,; Pronk, S.; Schulz, R.; Shirts, M.; Sijbers, A.; Tieleman, P.; Wennberg, C.; Abraham, M. W. M.; Hess, B.; Spoel, D. V. D.; Lindahl, E. *GROMACS USER MANUAL (Version 5.0-rc1)*; Royal Institute of Technology and Uppsala University, Sweden., 2014.

(67) Cozzolino, A.; Conte, P.; Piccolo, A. Conformational changes of humic substances induced by some hydroxy-, keto-, and sulfonic acids. *Soil Biology and Biochemistry* **2001**, *33*, 563–571.

(68) Schaumann, G. E. Soil organic matter beyond molecular structure Part I: Macromolecular and supramolecular characteristics. *Journal of Plant Nutrition and Soil Science* **2006**, *169*, 145–156.

(69) Smith, P.; Ziolek, R. M.; Gazzarrini, E.; Owen, D. M.; Lorenz, C. D. On the interaction of hyaluronic acid with synovial fluid lipid membranes. *Physical Chemistry Chemical Physics* **2019**, *21*, 9845–9857.