An Automated Parametrization Approach for Coarse-Graining Soil Organic Matter Molecules

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

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

[‡]Leibniz Institute for Catalysis (LIKAT), Albert-Einstein-Str. 29a, D-18059 Rostock, Germany.

E-mail: ashour.ahmed@uni-rostock.de

Abstract

Investigating the molecular structure of soil organic matter (SOM), along with its intramolecular interactions and interactions with other soil components and xenobiotics, is essential due to its ecological importance. However, the complexity and heterogeneity of SOM present significant challenges for systematic studies. While experimental methods are commonly employed, atomistic simulations provide a complementary approach to exploring molecular-level processes. The Vienna Soil Organic Matter Modeler 2 (VSOMM2) facilitates the construction of molecular models of SOM systems with various compositions at the atomistic scale, which can then be examined through molecular dynamics (MD) simulations. This study introduces a parameterization strategy that enables the conversion of VSOMM2-generated structures into a coarse-grained representation, thus allowing larger time and length scales to be explored. By employing a conformer search technique, direct construction and analysis of coarse-grained SOM models with diverse compositions were made possible, eliminating the need for atomistic MD simulations. To demonstrate this approach, coarse-grained SOM models were created based on selected samples from the International Humic Substances Society, considering different water content levels for each model. Comprehensive analyses, including density and potential energy profile calculations, revealed a partial correlation with the SOM compositions and demonstrated that electrostatic interactions govern the structural packing. Moreover, a local phase separation process, particularly the formation of SOM voids, was observed over several microseconds, underscoring the advantages of the coarse-graining technique.

1 Introduction

Soil is an important source of life on earth. It provides a reservoir of plant nutrients and contains the largest carbon pool that is biogeochemically active[1–3]. The functioning and composition of terrestrial ecosystems, as well as long-term climate regulation, are affected by the soil. Leveraging the soil's potential to mitigate climate change requires an understanding of its geochemical processes[4–6].

The organic fraction, also known as Soil Organic Matter (SOM) is a diverse and complex mixture of organic substances, formed through the interplay between various factors such as plant and microbial inputs, the nature of the parent material, and the climatic conditions[1,7,8]. Humic substances (HSs) are major components of SOM formed from the decomposition of plant and microbial residues. They are complex mixtures composed of numerous molecules, many of which contain an aromatic

nucleus with phenolic and carboxylic groups as substituents[9]. HSs contain two main fractions: humic acid (HA) and fulvic acid (FA), which differ in their chemical composition, solubility, molecular weight, color, and biological activity. HA is a larger, darker, and less soluble molecular system that mainly enhances soil structure and water retention, while FA is a smaller, lighter, and more soluble molecular system with a higher concentration of carboxylic groups that improves nutrient absorption and bioavailability[9].

These compounds have been standardized by the International Humic Substances Society (IHSS)[10]. The IHSS furnishes standard HAs and FAs from diverse sources. In addition, information is provided about the composition of the samples, which includes the elemental composition, or the ^{13}C NMR Carbon distribution.

Along with advancements in experimental methods and research into soil science, molecular-level simulations have become an indispensable tool, driven by increasingly sophisticated numerical algorithms and improved computer hardware[11]. Molecular modeling and simulations can offer valuable insights into soil processes at the molecular level, including molecular geometries and dynamics, thermodynamic and kinetic properties, and various physical phenomena. Due to their complexity and heterogeneity in chemical composition[12], it is challenging to model SOM substances. Various approaches and hypotheses have been introduced to describe SOM via molecular simulations.

Schulten and colleagues developed a polymeric SOM model, based on bio- and geochemical, NMR-spectroscopic and mass spectrometric analyses[13–16]. It consisted of a network of long-chain alkyl structures interspersed with aromatic rings[17]. Subsequent refinements incorporated carbohydrate and protein units[14,18]. It had the advantage of providing a multitude of interaction sites and functionalities, to represent an average HS molecular structure. This enabled the study of the adsorption of xenobiotic compounds onto HS[15,19,20]. However, the idea that SOM is composed of high-weight molecular species is now considered obsolete, even though the models do not imply the existence of these specific structures[21–24].

Another approach that has been employed is the modeling of individual SOM functional groups[25,26]. This reduces the complexity of the systems, which in turn enables the application of more accurate methods, such as those based on quantum mechanics. However, this approach could be compromised by the number of functional groups that have to be considered to sufficiently capture the characteristics of SOM. Ahmed and colleagues addressed this challenge by utilizing a large test set comprising over 30 separate and distinct organic molecules that represent the most common SOM functional groups. These molecules were identified through extensive characterization and analysis of numerous soil samples using pyrolysis field-ionization mass spectrometry (Py-FIMS) and x-ray absorption near-edge spectroscopy (XANES)[27–33]. This approach involved simulating various SOM building blocks, including carbohydrates, phenols/lignin monomers, lignin dimers, lipids, alkylaromatics, non-amide nitrogen-containing compounds, amides, and sterols. It effectively captured the binding/interactions between organic contaminants and SOM on three levels: individual SOM functional groups (small organic molecules); SOM compound classes (or building blocks); and SOM as a whole. By applying quantum mechanical calculations and quantitative structure-activity relationship (QSAR) models, Ahmed and colleagues identified key parameters driving the specific binding mechanism. Overall, this molecular modeling approach showed good agreement with experimental results and provided insights into molecular-level interactions that are beyond the reach of current experimental methods.

An alternative approach based on the experimental characterization of HS is the Vienna Soil Organic Matter Modeler (VSOMM), introduced by Sündermann et al.[34]. It assumes that HS is composed of a mixture of smaller molecules (such as those previously introduced by Ahmed et al.) that are mainly held together by dispersion and electrostatic forces, with covalent bonds playing a minor role[35–39].

2

The VSOMM generates atomistic models of SOM composed of HS molecules, which are constructed from a fragment database, matching the elemental composition and fractions of certain functional groups. This enables the investigation of HS structures that are versatile in composition. Together with the release of a second version[6], which included optimizations such as an improved algorithm for matching the organic composition to the input parameters, this has made it possible to carry out numerous investigations to better understand HSs[24,40–45].

Simulations of complex and relatively large systems, which are often necessary for modeling SOM, often require considerable computing resources. One approach that can be used to speed up molecular simulations is the so-called coarse-graining technique[46], whereby several atoms or even entire molecules are grouped into new particles, often known as beads. This reduces the number of particles and interactions that need to be evaluated, which improves performance. The coarse-graining technique inherently involves a trade-off, as it results in a loss of atomistic detail. However, this limitation is not a significant drawback when applied to SOM molecules, as their precise structural details are largely unknown. Different coarse-grained force fields exist[47–49], including the widely used Martini force field[50].

Several approaches have been used to achieve a coarse-grained representation of VSOMM models. Dissipative particle dynamics was used to simulate models from the initial version of the Modeler[51]. Recently, Xue et al.[52] used the Martini force field to convert a set of HS molecules from the VSOMM2 into a coarse-grained representation. They investigated the dynamics and interactions of the molecules with lipids, peptides, carbohydrates, and lignin. We have developed a coarse-grained VSOMM2 model by utilizing the recently released Martini 3 version[53], by coarse-graining the entire fragment database[54]. This enabled the generation and investigation of different HS structures based on a general conversion scheme.

While the coarse-graining approach used in our previous work improved the simulation of the HS molecules, the parametrization of the bonded interactions relied on atomistic simulations of the reference systems, which could be computationally expensive. This approach may become less efficient when different compositions of SOM need to be investigated, as it would require atomistic simulations for each reference system. Readily available coarse-grained SOM models could facilitate long-term and large-scale simulations of SOM systems of different compositions, which would be advantageous, for example, in the investigation of adsorption processes.

Therefore, this study aims to build on the approach used in our recent work[54] by introducing an improved parameterization protocol. Instead of using atomistic simulations of SOM structures of VSOMM2 for the parameterization of the bonded interactions, we employ a conformer generation method known as Experimental-Torsion Basic Knowledge Distance Geometry (ETKDG)[55]. Our new parameterization approach was benchmarked by constructing coarse-grained SOM systems representing various abundant HSs with compositions characterized by differing degrees of acidity, aromaticity, and common functional groups found in soils, thereby reflecting a general and realistic SOM. The rationale behind selecting these SOM systems is to capture the diversity and complexity of natural SOM, by providing realistic models that mimic its behavior under different environmental conditions. Simulations for each molecular model system with a specific chemical composition were conducted with varying levels of water content, enabling the evaluation of corresponding density and energetic properties and providing insights into molecular-level interactions and processes that occur in soil environments.

2 Methods

2.1 Parametrization Approach

We recently introduced SOM models that are compatible with the Martini 3 force field[54]. These models were derived from structures generated by the VSOMM2, which constructs SOM

3

structures using a set of fragment molecules[6,54]. The parametrization approach utilized Swarm-CG[56] to optimize the bonded interaction parameters of the isolated fragments and to generate the corresponding coarse-grained structures. These fragments were then assembled into larger macromolecules[1], using potential energy parameters for the connection between fragments, determined through the direct Boltzmann inversion (DBI) method.

While this work presents a robust pipeline for constructing coarse-grained SOM models with arbitrary compositions, the DBI-based parametrization requires MD simulations of the respective atomistic SOM reference systems. Conducting these reference simulations for each case can be time-consuming, especially when investigating coarse-grained SOM systems with varying compositions. Furthermore, the previous investigation demonstrated that a time step of 10 fs is necessary to maintain the stability of the equations of motion of the systems. Although the coarse-grained representation has already extended simulation times, further optimizing the interaction parameters to allow for a larger time step could significantly enhance performance.

To address the issues mentioned above, we implemented two major modifications to our previous parametrization approach[54]. First, to improve numerical stability, the individual fragments were reparametrized using Swarm-CG, restricting the maximum value for the angle force constants to $350 \, \mathrm{kJ \, mol^{-1} \, rad^{-2}}$. For further details about the reparametrization, please refer to Section S1 of the Supporting Information. As will be demonstrated later, this modification effectively reduces high-frequency components in the vibrational power spectra of the fragments, indicating a reduction in high force gradients. This, in turn, enabled the use of a time step of 20 fs for the integration of the equations of motion for the SOM systems.

Second, to construct coarse-grained models independently of atomistic MD simulations,

equilibrium distances and angles were determined using the conformer search method, specifically the ETKDG approach[55]. This approach has been used for the parametrization of Martini models in the work of Potter et al.[57,58]. For further information about the ETKDG approach, please refer to Section S2 of the Supporting Information. The main advantage of using this method is its ability to provide interaction parameters between different fragments much faster than performing MD simulations.

The workflow of the modified parametrization approach is illustrated in Figure 1. This procedure has been implemented into a Python tool, which is available on Zenodo[59].Using this tool, topology files obtained from the VSOMM2 can be directly converted into those for the coarse-grained structures, using a parametrization step. As a result, there is no need to perform or prepare reference simulations, as the parameterization can be carried out via a single command from the user.

To benchmark the modified parametrization approach, the following analyses were conducted. First, the vibrational power spectra for the fragments were calculated with the GROMACS tool *gmx velacc*. These spectra were calculated using both parameterization approaches: the previous one described above[54], and the current modified approach with restricted force constants. Moreover, the influence of the reparametrization on the agreement between the atomistic and coarse-grained distributions was assessed using a mismatch score derived from the Swarm-CG tool. Finally, the optimal number of conformers needed to achieve sufficient convergence of equilibrium distances and angles was determined.

2.2 Simulation of SOM Compositions

The chemical heterogeneity of SOM is one of its characteristic features. To explore this diversity, SOM systems with five different compositions were investigated. Escalona et al.[45] analyzed and sorted HSs from the International Humic Substances Society (IHSS) via a principal component (PC) analysis, using aromaticity

---

[1]The term "macromolecule" refers to molecules composed of multiple fragments. It should not be confused with the polymeric SOM model introduced by Schulten and colleagues.

Read in atomistic topology files

Perform conformer search

Add Swarm-CG parameters

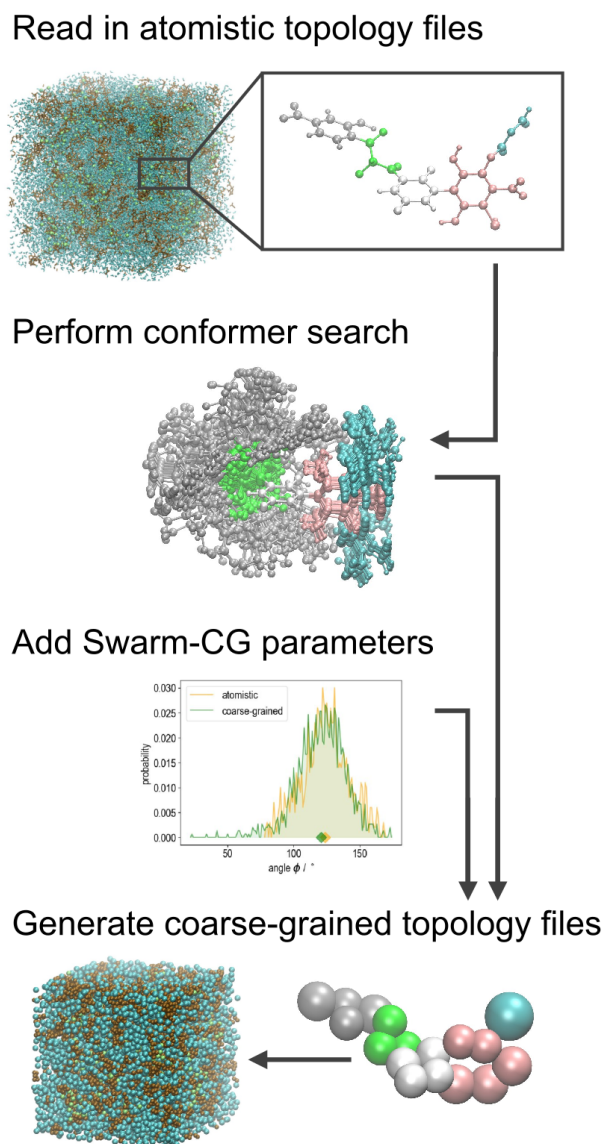Generate coarse-grained topology files

Figure 1: Graphical illustration of the modified parameterization approach. The tool that was developed reads the atomistic topology files. Subsequently, equilibrium values are determined using the conformer search (ETKDG) and augmented with values obtained from the Swarm-CG parameterization. Finally, the coarse-grained topology files are output with the mapped structure, which can subsequently be solvated in water.

and acidity as PCs (please refer to Figure 1 in their paper). Following this categorization, certain HS types, including humic acid (HA) and fulvic acid (FA), were investigated to represent characteristic compositions within the chemical space (see Table 1). For more information about the different SOM types, please see Table S1 in the Supporting Information, which provides the elemental and organic fractions of the SOM systems. In addition, Figures S2 and S3 provide detailed insights into the heterogeneity and distribution of fragments within the SOM compositions.

The atomistic structures were generated using the VSOMM2, wherein each SOM model system comprised 1600 fragments in total, with each macromolecule consisting of 5 connected fragments. The pH was maintained at 5, and calcium ions were introduced to achieve charge neutrality. The water content was varied for each SOM model system (please refer to Table S2 in the Supporting Information). To ensure consistency with previous studies, the water content is reported as the heavy atom fraction of water, x_{water}, determined from the corresponding atomistic system. These structures were subsequently converted into coarse-grained models using our newly developed tool. As the SOM composition remains constant across different water contents for each SOM type (model system), the parameterization procedure only needs to be performed once (i.e., for a single water content). The resulting topology files can then be used for the cases with other levels of water content. Additionally, the new tool allows to produce the mapped coarse-grained structure without the need for the parameterization step. These mapped structures were generated using the center of geometry mapping and subsequently solvated with *gmx solvate*.

Following the coarse-grained MD simulations, detailed analyses of the structures' properties were carried out. In particular, densities and potential energies were calculated, using the GROMACS tool *gmx energy*. Furthermore, a phase separation process was investigated and evaluated using partial densities, which were determined utilizing the *gmx density* tool.

5

Table 1: HS samples from the IHSS, which are modeled and analyzed in this work.

Identifier	HS sample	Characteristics
SOM I	Elliot Soil I (FA)	high carboxyl and carbonyl content
SOM II	Elliot Soil IV (HA)	low carboxyl and carbonyl content
SOM III	Leonardite (HA)	high aromatic content
SOM IV	Pahokee Peat II (FA)	intermediate carboxyl and aromatic content
SOM V	Suwannee River II (FA)	high hetero-aliphatic & aliphatic content

2.3 Computational Details

The MD simulations were performed using the GROMACS simulation package version 2019.4[60]. Python3 packages, including NumPy[61], Matplotlib[62], MDAnalysis[63,64], RDKit (https://www.rdkit.org), and IPython[65], were employed to generate the coarse-grained topology files and structures. The VMD software was used for visualization[66].

Each SOM model system was first relaxed utilizing an energy minimization. An NPT equilibration was then carried out, which was followed by the production run. The equilibration time was set to 10 ns. The production run lasted 20 μs, where the last 3 μs were used for the analysis of densities and potential energies. The leap-frog integrator was applied, using a time step of 20 fs for all MD simulations. The temperature was held constant at $T = 298.15\,$K, using the velocity-rescale thermostat[67] with a coupling constant of $\tau_T = 1\,$ps. For the NPT equilibration, the Berendsen barostat[68] with a coupling constant of $\tau_p = 8\,$ps was applied, whereas the Parrinello-Rahman barostat[69] with a coupling constant of 12 ps was utilized for the production run. The reference pressure was set to 1 bar. For the Lennard-Jones interactions, the Verlet cutoff scheme with a straight cutoff of 1.1 nm was used. Reaction-field electrostatics were utilized to evaluate the Coulomb interactions. A cut-off of 1.1 nm with a dielectric constant of $\epsilon_r = 15$ was applied. Constraints were incorporated by employing the LINCS algorithm. All simulations were conducted using periodic boundary conditions.

As a reference for the Swarm-CG reparametrization, the atomistic simulation setup from our previous study[54] was used, where isolated atomistic fragments were solvated in water and simulated for 25 ns. The GROMOS force field version 54A7[70,71] was applied in the united-atom representation, following the description of the VSOMM2 atomistic models. Moreover, the identical setup for the Swarm-CG parametrization was employed, encompassing the simulation of the coarse-grained fragments that were solvated in water. One difference to the parametrization from our previous study is that the angle force constants were restricted, as described above. For further computational details about the atomistic simulations and the Swarm-CG parametrization, please refer to Dettmann et al.[54], Section 2.3.1 and 2.3.3, respectively.

Regarding the parameterization of the equilibrium distances and angles of the macromolecules using ETKDG, 1000 conformers per molecule were generated. This number was sufficient to achieve convergence of the distance and angle values, as discussed in Section 3.1.2.

3 Results and Discussion

This section is organized into three subsections that address the benchmarking of the modified parameterization approach (Section 3.1); the development of coarse-grained SOM systems with varying compositions (Section 3.2); and the evaluation of numerical performance (Section 3.3). The first subsection provides an analysis of the revised force constants for the SOM fragment molecules (Section 3.1.1) and details of the ETKDG approach (Section 3.1.2).

6

3.1 Benchmarking the Modified Parametrization Approach
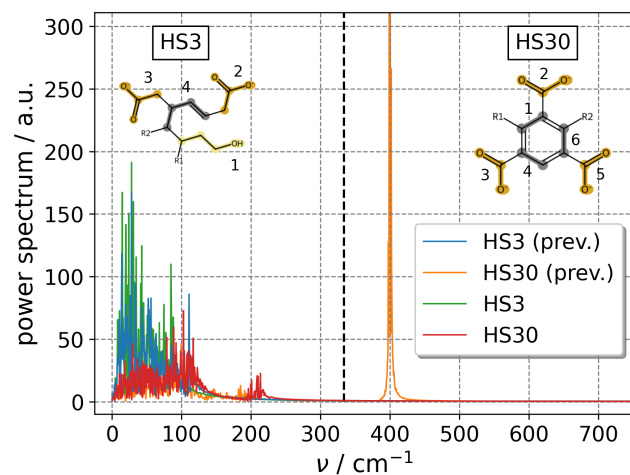
3.1.1 Revised Force Constants

Figure 2: Power spectra of the velocity auto-correlation function for HS3 and HS30 fragments, obtained using the previous (prev.) parametrization approach (blue and orange) and the current restricted parametrization approach employed in this work (green and red). The colored groups of fragment molecules HS3 and HS30 represent their bead assignment.

The revision of the force constants of the SOM fragment molecules from our previous work[54] aimed to enable a larger time step in the integration of the equations of motion. Our previous findings suggested that high vibrational frequencies could compromise the numerical stability of simulations. To illustrate this point, we calculated the vibrational power spectra for two SOM fragments, HS3 and HS30, using the revised force constants and compared them to power spectra obtained with the parameters from our previous parameter set (see Figure 2). The fragment numbering follows the VSOMM2 database. HS30, known for its high frequencies, was a primary contributor to simulation crashes. The adjusted force constants caused a significant spectral shift in HS30, with the most intense peak moving from around $400\,\mathrm{cm}^{-1}$ (orange) to approximately $200\,\mathrm{cm}^{-1}$ (red), accompanied by a notable reduction in intensity. Conversely, HS3 displayed no significant changes,

likely due to its inherently low force constants being unaffected by the restriction.

It should be noted that the dashed line in Figure 2 represents the maximum frequency that can be resolved with a time step of $20\,\mathrm{fs}$. With the revised force constants, all fragment frequencies fall below this limit, indicating that the equations of motion can be effectively integrated with the larger time step. The final force constants can be found in the Python script used to create coarse-grained SOM structures (available in the Zenodo repository[59]). From a general perspective, the choice of the time step influences the accuracy of the integrator (in this case, the second-order leap-frog algorithm) and directly impacts the conservation of key quantities such as energy, temperature, and pressure[72]. A detailed investigation into the potential effects of further increasing the time step on these quantities could be considered a topic for future research.

We also determined the parameters of the bonded interactions for the set of fragments using the recently released "Bartender" tool[73] which employs MD simulations at the semiempirical quantum mechanical level of theory as a reference. However, this approach resulted in even higher force constants than those obtained from the previous parametrization using Swarm-CG. This finding is consistent with the work of Pereira et al.[73], which demonstrated that the molecules described using semiempirical quantum mechanics are usually more rigid than those described by atomistic force fields. In certain cases, the force constants would also need to be manually scaled down, which would probably result in a similar parameter set.

The limitation of force constants results in a narrower range of feasible solutions for the optimal parameters. In Swarm-CG, the match between coarse-grained and atomistic distributions is evaluated via the so-called Earth Mover's distance (EMD)[74]. The similarity of the distributions can serve as an indicator of how accurately the coarse-grained model reproduces the dynamics of the grouped (mapped) atoms. In other words, a closer alignment between the two distributions corresponds to a smaller EMD value (for details about the EMD,
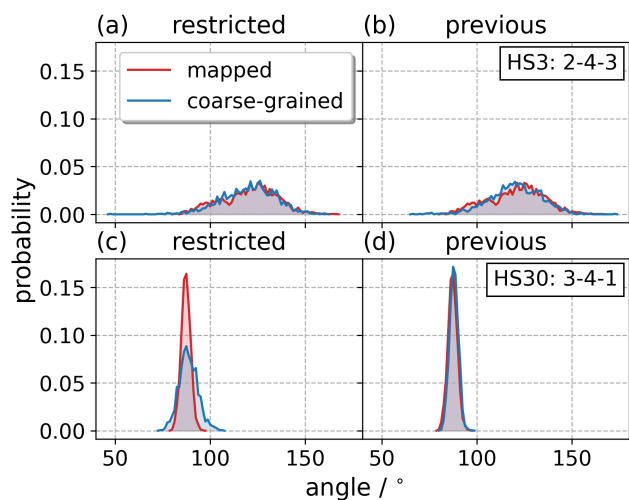
Figure 3: Comparison of the angle distributions of mapped (red) and coarse-grained (blue) models. The distributions of the angle between the beads 2, 4 and 3 (2-4-3) of the HS3 fragment are shown for the restricted parametrization (a) and previous parametrization (b). Similarly, the angle distributions of the 3-4-1 angle of the HS30 fragment are shown in (c) and (d). The molecules with the bead indices are shown in Figure 2.

please refer to Section S5 in the Supporting Information).

To ascertain the potential impact of the restricted parametrization on accuracy, we compared the agreement of the bonded distributions of the fragments from our previous parameterization approach[54] with those of the present work. Figure 3 illustrates the angular distributions of a specific angle of the HS3 and HS30 fragments, respectively, for both approaches. In each diagram, the distribution of the optimized coarse-grained model is compared with the mapped reference. The complete set of distributions for these two fragments is given in the Supporting Information, Figure S4.

No significant difference is observed for the HS3 fragment when comparing the two approaches, as shown in Figure 3 (a) and (b). This is consistent with the results obtained from the power spectrum calculation, which also showed no significant change for the HS3 fragment. In contrast, the HS30 fragment exhibits a deviation in the match of the coarse-grained distri-

bution to the reference for the restricted model, as observed in Figure 3 (c). This behavior is also evident in other angular distributions of this fragment (see Figure S4). Due to the restriction of the angle force constants, the model does not fully reproduce the reference distribution. However, the deviation remains moderate, and the mean value is accurately captured.

Swarm-CG employs a sequential optimization approach by conducting multiple coarse-grained simulations to match the distributions of the coarse-grained molecules to the reference. The optimal solution is identified within a multidimensional solution space using the particle swarm algorithm. This means that the solution obtained by Swarm-CG may vary with each calculation, even under identical conditions. Therefore, attention should be directed toward systematic changes in the distributions rather than minor deviations.

In general, the stronger restriction of the angular force constants can be expected to result in a lower degree of accuracy in describing the angular interactions. This illustrates the compromise that must be made to improve the numerical stability and thus the performance of the models. However, the extent of the accuracy loss depends on the nature of the particular angular potentials, and is therefore assumed to be different for different fragments, as suggested by Figure 3.

To quantitatively investigate the impact of the restricted parameterization, the total mismatch score (TMS) of the individual fragments for the previous and current models was compared (see Figure 4). The TMS of a fragment is calculated from the individual EMDs of each bonded term, including bonds, angles, and dihedral angles. For further details about the TMS calculation, please refer to Section S5 in the Supporting Information.

The individual score components of the TMS are not normalized by the number of bonded interactions; therefore, TMS values should only be compared for the same molecule and under the same definition of bonded interactions. Consequently, when interpreting Figure 4, attention should be paid to the position of each point rather than comparing different points. A
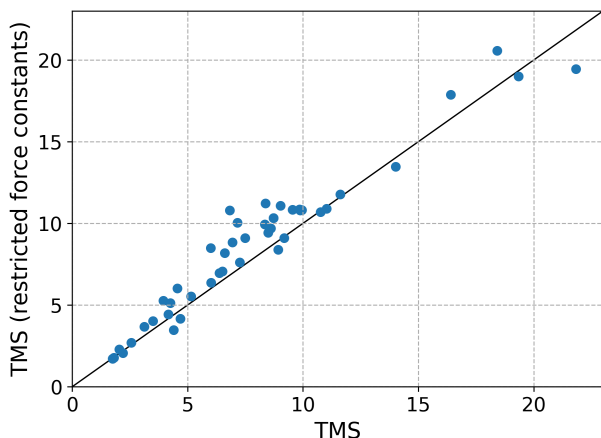
8

Figure 4: Comparison of the total mismatch scores (TMSs) of the fragments from the Swarm-CG parametrization. The values from the previous parameterization (x-axis) are plotted against those from the modified parameterization in this study (y-axis). Each point represents a comparison of the two EMD values for a specific fragment.

higher position in the upper half of the graph indicates a larger TMS value for the restricted parameterization compared to the previous one, which suggests that restricting the force constants results in a less accurate reproduction of the distributions. Most TMS values remain almost unchanged, implying that the current model retains a similar level of accuracy in reproducing bonded distributions as the previous model.

3.1.2 ETKDG Approach

The conformer search approach is capable of providing structural information about molecules relatively quickly, which is the motivation for its use in the modified parametrization approach. Equilibrium distances and angles are determined by mapping the individual conformers. The mapped representation is then averaged over the entire set of configurations to obtain the equilibrium values. However, the accuracy of the calculation is affected by the number of conformers included. Therefore, it was necessary to determine the number of conformers required to ensure sufficient convergence of the average bond and angle values.
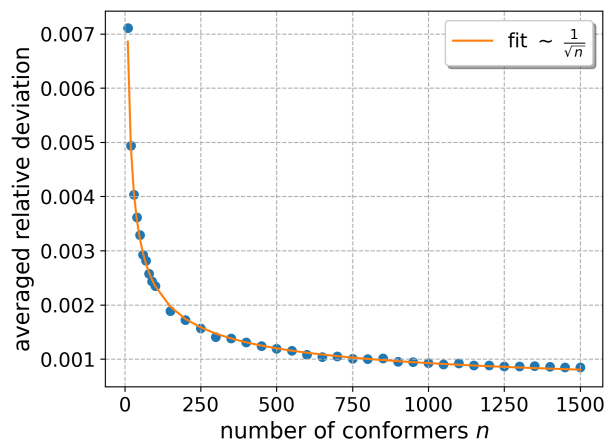

Figure 5: Convergence of the bond values for the "SOM III" model system as a function of the number of conformers. The plot depicts the relative deviation from the mean value, averaged across all bonded terms within the model system.

Further details can be found in the Supporting Information, Section S2.

Figure 5 illustrates an average convergence behavior for the "SOM III" model system. For a more detailed explanation with additional illustrations, please refer to the Supporting Information (Figure S5). The convergence behavior results in a relatively low deviation for a high number of conformers. For example, the relative deviation from the mean for 1000 conformers averaged over all bonds in the system is around $1\,\%_0$ (0.001). Overall, the data points follow a convergence behavior proportional to $1/\sqrt{n}$. The angle values (please see Figure S6 in the Supporting Information) show slightly higher deviations, which can be attributed to the associated higher number of degrees of freedom compared to the bonds. Based on this analysis, it is concluded that 1000 conformers are sufficient. This number was utilized for the parametrization of all the other SOM systems, which is discussed in Section 3.2.

Potter et al.[57] conducted a similar investigation for a dodecane molecule, focusing on how average bond lengths and angles vary with the number of conformers. Given the possible heterogeneity of bonded interactions in the SOM systems and the multiple options for construct-

9

ing the models, the convergence behavior may be contingent on the particular system considered. For example, the number of bonded terms present within a particular macromolecule and, therefore, its number of fragments, could influence the results. Thus, an investigation of this nature should be revisited, when altering the number of fragments per macromolecule.

As shown, it is possible to determine the equilibrium values of bonds and angles by averaging over different conformers generated by ETKDG for all bonded interactions within one molecule. The parametrization based on Swarm-CG provides these values solely for the bonded interactions within individual fragments, as not every macromolecule can be sampled through MD simulations in an appropriate timeframe. Hence, the subsequent approach was employed to combine these two sets of parameters. The equilibrium values of bonds and angles, as well as dihedral angles are taken from ETKDG, to prevent a mixing of these values with those from Swarm-CG. The force constants obtained from Swarm-CG are employed for the bonded interactions within the fragments, whereas standard force constants are utilized for the interactions between different fragments. A comparison of the equilibrium values generated by Swarm-CG and ETKDG can be found in Section S6 of the Supporting Information.

3.2 SOM Systems with Different Compositions

The densities yield information about the compactification behavior of SOM. The five different compositions of SOM presented in Table 1 were simulated. The system densities, respectively averaged over the last $3\,\mu s$ of the production runs and calculated for the different levels of water contents in the system, denoted by x_{water}, are shown in Figure 6. Please note that due to the utilization of Martini standard masses, the densities were rescaled, using the ratio of the mass of the respective atomistic to that of the coarse-grained system.

The overall trends for the densities of the different SOM types are similar. As discussed previously[54], coarse-grained Leonardite (HA)
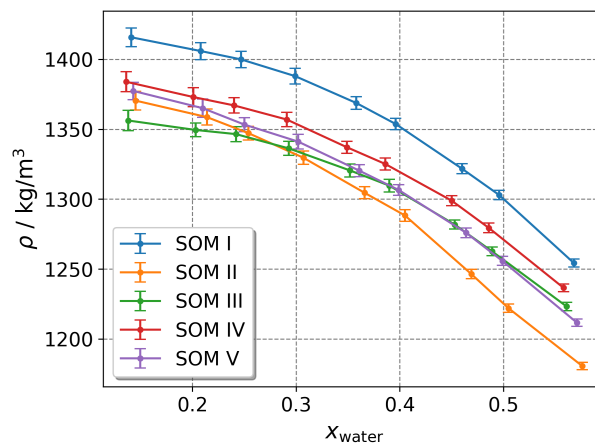

Figure 6: Average of system densities for the SOM types calculated for different water contents x_{water} in the systems; error bars denote standard deviations.

systems tend to underestimate the density, especially for lower water contents, due to the calcium ions having fixed bead sizes. This results in the coarse-grained model being unable to form distinct calcium-carboxylate complexes. In particular, the so-called bidentate complexes, which contribute to a tight packing at lower water contents, cannot be reproduced by the coarse-grained model, which leads to an underestimation of the density for these cases. However, for Leonardite (HA), the deviations were still acceptable (maximum of around $10\,\%$). The use of smaller-sized water beads (tiny water) facilitated a denser packing of the SOM molecules, thereby reducing the discrepancy between atomistic and coarse-grained densities. Nonetheless, at lower water concentrations, this discrepancy remained significant, emphasizing the critical influence of calcium-carboxylate complexes on the system's density.

By comparing the different compositions, the densities can potentially be related to the abundance of certain functional groups, such as carboxylate groups. As mentioned previously, the interaction between these groups and calcium ions contributes to the SOM packing, which is especially important at lower water contents. In the case of the SOM systems with a similar aromatic content that were investigated, the densities are higher for the compositions with a

10

higher acidity. From high to low acidity, the respective SOM types are: SOM I (blue), SOM IV (red) and SOM II (orange). Therefore, this could again be explained by the higher number of available carboxylate-calcium complexes that might contribute to structures that are denser in overall. This behavior has also been observed in studies investigating humic substances using atomistic simulations[45]. Consequently, this is a feature that can be reproduced with the coarse-grained model as well.

Due to the associated computational efforts, simulations of atomistic structures were not considered in the present study. Especially, considering the system sizes (320 macromolecules each) and the equilibration time of such structures, particularly at lower water contents[45,54], modeling at an atomistic resolution becomes unfeasible. The relatively high number of molecules was used to achieve a higher degree of heterogeneity, reflecting an important property of SOM. However, given the size of the systems and the simulation of several microseconds, the densities could be obtained relatively easily using the coarse-grained approach.

To measure the strengths of the interaction types in the system, the potential energy was normalized with the total number of beads in the system. The different parts of the total potential energy were calculated, which includes the potentials of bonded and non-bonded interactions. It should be noted that for each potential type, the same number of beads for the respective SOM type was used for the total potential energy so that the individual potential types can be regarded as a true division of the total potential energy.

Concerning the total potential energy in Figure 7 (a), a relatively constant value for lower water contents is visible. The energy values become less negative with increasing water content. Thus, the potential energy has the minimum value, when the SOM systems are more compact, namely at lower water contents. Based on the aforementioned considerations regarding the stability of SOM structures, it is assumed that the non-bonded interactions, and especially the Coulomb interactions are responsible for the lower potential energy. To verify

this, the potential energy was subdivided into its individual components.

The potential energies of the bonded terms, namely bonds (Figure 7 (b)), angles (Figure 7 (c)), and improper dihedral angles (Figure 7 (d)) follow a similar trend. The higher potential energies at lower water contents could be explained by the bonded terms of the molecules being in a less energetically preferable configuration compared to other water contents, due to their tighter packing. These trends are opposite to those for the total potential energy, which are, however, not noticeably affected, as the fraction of potential energies of the bonded terms is relatively small.

The non-bonded interactions are primarily described by the Lennard-Jones potentials in the Martini force field. Usually, partial charges are not applied to the beads, in contrast to the atoms used in atomistic simulations. Because of this, the Lennard-Jones interactions in Figure 7 (e) account for a significant proportion of the total potential energy and follow a very similar trend, except for slightly higher values at lower water contents. From the perspective of this interaction type, the system is in an energetically preferable state when water contents are around $x_{\text{water}} = 0.3$ or higher, depending on the SOM type.

The reason why the total potential energy is relatively constant at lower water contents is the negative contribution of the Coulomb energies (Figure 7 (f)), which is also evident from the sum of the Lennard-Jones and Coulomb energies (Figure S7 in the Supporting Information). The lower water content enables a stronger interaction between the ions and the charged carboxylate groups, leading to a stronger Coulomb energy. The curves of the Coulomb energies are consistent with the notion that the Coulomb interaction is the driving force behind the compactification of the systems. These interactions are unique in that their strength increases as water content decreases. This phenomenon can be understood as a dilution or shielding effect: as the water content rises, the shielding by water molecules increases, and conversely, it decreases with lower water content. Furthermore, it is pos-
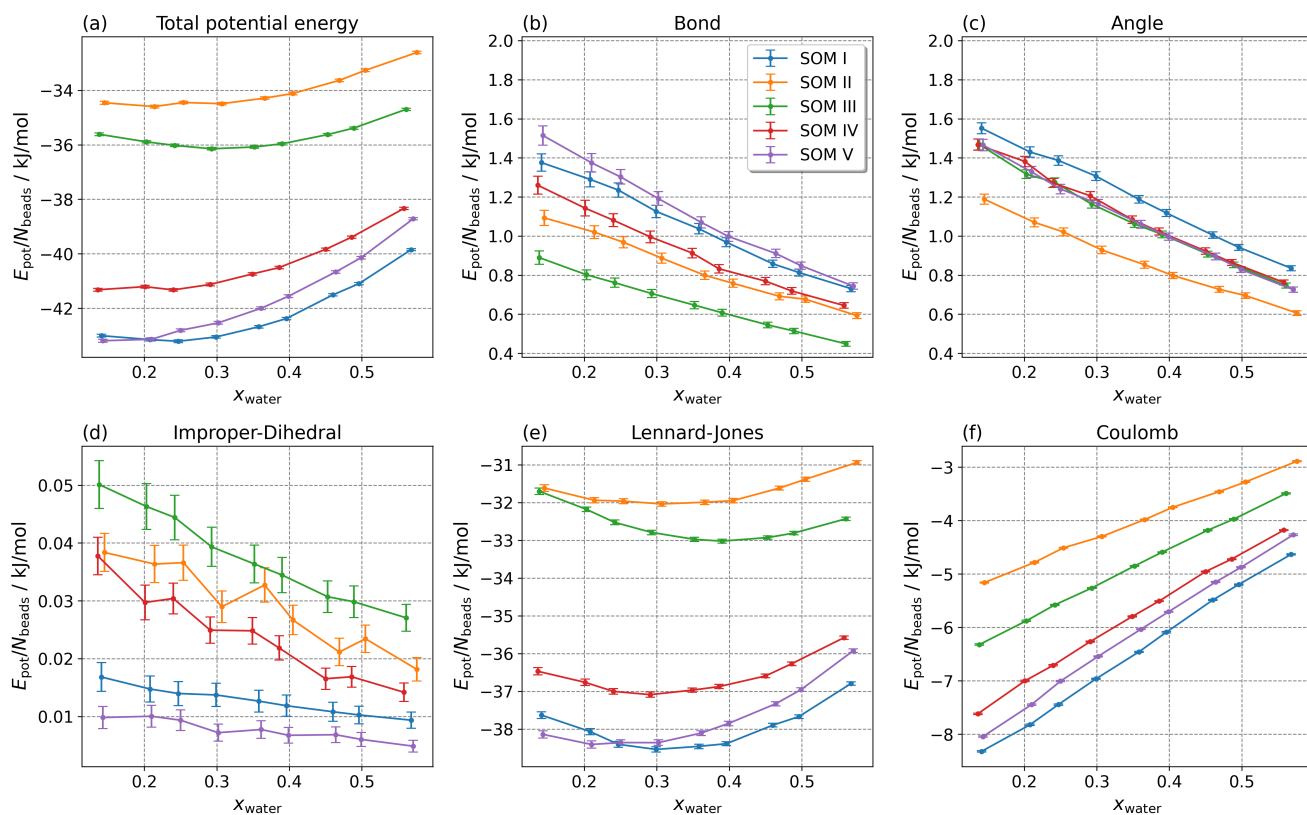
11

Figure 7: Time averaged total potential energies (a), and potential energies of bonded ((b)-(d)) and non-bonded interactions ((e)-(f)) for the different SOM models; error bars denote standard deviations. Please note the different scales of the y-axes.

sible to establish a link between the Coulomb interaction strength and the acidity of different SOM types, as already assumed in the discussion about the densities. A higher acidity implies a larger number of charged carboxylate groups in the system, which contribute to the Coulomb potential. For example, SOM I (high acidity) has the strongest Coulomb interaction, whereas the opposite is true for SOM II (low acidity).

Interestingly, the curves for the total potential energy and the Lennard-Jones potential energies follow the same order as those of the Coulomb potential energies.

To obtain further information and also to analyze the behavior of the Lennard-Jones energies, these were subdivided into their individual components, as shown in Figure 8. The subdivision of the Coulomb interactions can be seen in Figure S8 of the Supporting Information. The characteristic order of the different curves is most apparent in the HS-calcium (HS-CA) in (Figure 8 (c)) and water-calcium (W-CA) Lennard-Jones interactions (Figure 8 (e)). This is reasonable because the same order was observed in the Coulomb interactions, where the calcium ions were involved, and they are also involved in these two types of interaction. Furthermore, it is evident that the Lennard-Jones interactions between the calcium ions and the rest of the system account for a significant proportion of the total Lennard-Jones potential energy.

The curves of the HS-HS (Figure 8 (a)) and HS-CA (Figure 8 (c)) interactions exhibit a similar trend, that is, the interaction strength decreases as the water content increases. The opposite trend was observed for the W-W case (Figure 8 (d)). These trends can be explained by the increasing number of water molecules in the system with increasing x_{water}. As the water concentration increases, the HS molecules and calcium ions become more solvated, which increases the contact area and consequently the interaction with water. The replacement of more interaction sites with water results in a decrease in the number of HS molecules interacting with calcium and among themselves, thereby reducing the strength of the

Lennard-Jones interactions. In the W-W case, a higher water concentration results in more water molecules interacting with each other, which leads to a decrease in the corresponding potential energy.

HS-W (Figure 8 (b)) and W-CA Lennard-Jones interactions (Figure 8 (e)) follow a similar trend, whereby the potential energy decreases first with increasing water content. After reaching an extremum at around $x_{water} = 0.4$, the energies increase again. This can be explained in a similar way. The presence of more water beads in the system results in an increased interaction strength with the calcium ions and HS molecules. However, when a certain concentration is reached, the number of interaction sites becomes saturated, which leads to a weaker increase in the interaction. Keeping in mind that the per bead Lennard-Jones potential energies are taken into consideration, the number of beads in the system N_{beads} increases with higher water concentrations, leading to an increase in the per bead potential energies. This saturation effect is also evident in the Lennard-Jones interaction energies, which were not normalized by the number of beads (please refer to Figure S9 in the Supporting Information).

Escalona et al.[43] studied Leonardite (HA) models from the VSOMM2 using atomistic simulations, and calculated the solvation free energy of water molecules for different water contents in the Leonardite (HA) systems. They observed that the solvation energy increased until $x_{water} = 0.17$, when the system reached a fully hydrated state. The free energy remained constant at the free energy of inserting a single water molecule into bulk water until it increased again when it reached $x_{water} = 0.43$. This indicated that a phase separation between HS molecules and water had begun.

Although the free energy also contains entropic information, an analogy can be made with the Lennard-Jones potential energies from this study. As mentioned above, the weakening of the Lennard-Jones energies for the HS-W and W-CA cases suggests that more water molecules start interacting with each other than with HS molecules or calcium ions. The point at which the per bead potential energies are
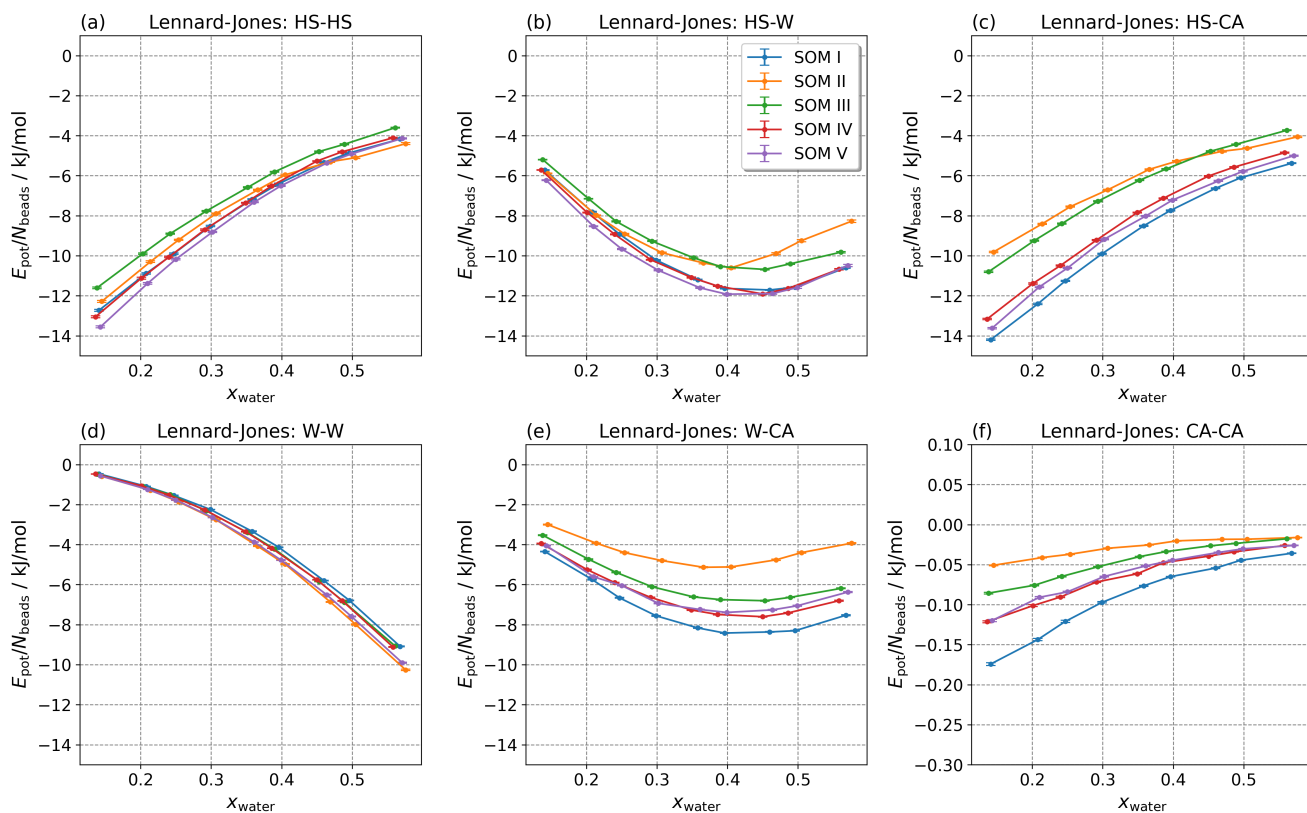
Figure 8: Time averaged Lennard-Jones potential energies for the different pairs of interactions between HS molecules, water (W) and calcium (CA); error bars denote standard deviations.

strongest, indicating that the interaction sites are saturated with water, is comparable to the x_{water} value obtained by Escalona et al.[43], at which phase separation occurs. However, the Lennard-Jones potential energies alone do not provide sufficient information about whether or not a phase separation occurred.

In this regard, the formation of larger regions filled with water was observed, which can be interpreted as a local phase separation between SOM and water, particularly evident for SOM II at the highest water content. The phase separation process is depicted quantitatively in Figure 9. It shows the overlap between the normalized partial densities of the HS molecules and water as a function of the simulation time of the production part of the trajectory. For additional information on the overlap calculation, please refer to Section S7 in the Supporting Information. In addition, Figure 10 presents two snapshots of the trajectory of the aforementioned system, one at the beginning of the production run (a) and one at the end (b).
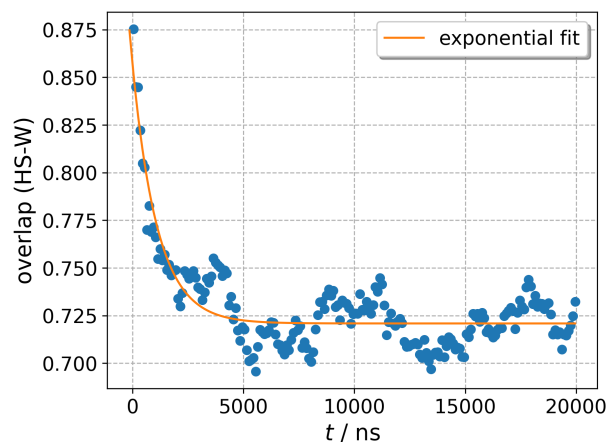

Figure 9: Overlap between normalized partial densities of HS and water over simulation time, in y-direction. SOM II with the highest water content is considered. Please note that the exponential fit was applied to obtain the average value, which the partial density oscillates around.
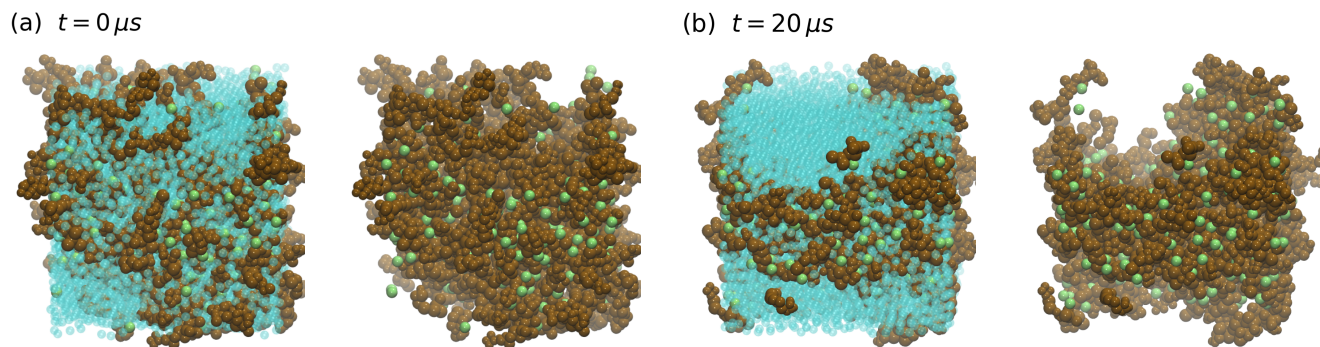
Figure 10: Snapshots from the trajectory of SOM II with the highest water content, at $0\,\mu s$ (a) and after $20\,\mu s$ (b). HS (brown), water (cyan) and calcium (green) are depicted. Each snapshot is shown with and without water, respectively.

Figure 9 shows that the overlap is decreasing exponentially, approaching a value of around 0.721. This indicates that the initially more mixed structure, with a greater overlap between partial densities, gradually becomes less mixed, resulting in a lower overlap. The strength of the demixing can be further quantified using the parameters of the exponential fits. These can be found in Table S3 of the Supporting Information. For the other SOM systems at the highest water concentration, the overlap is exclusively less than one but higher than for SOM II. A certain amount of demixing is to be expected, as the structures formed by the HS molecules will not result in an ideal mixture.

A possible explanation for, why this system, in particular, shows a clear phase separation is, first, that the high water content provides sufficient water molecules in the system to facilitate the formation of a distinct water phase. Second, SOM II is distinguished by its low acidity, which may be attributed to the particular hydrophobicity of the HS molecules, in contrast to the other SOM types. This could contribute to the formation of a separate phase in a hydrophilic solvent, such as water. Additionally, the accessible time scales are within the microsecond range, which supports the observation of such processes.
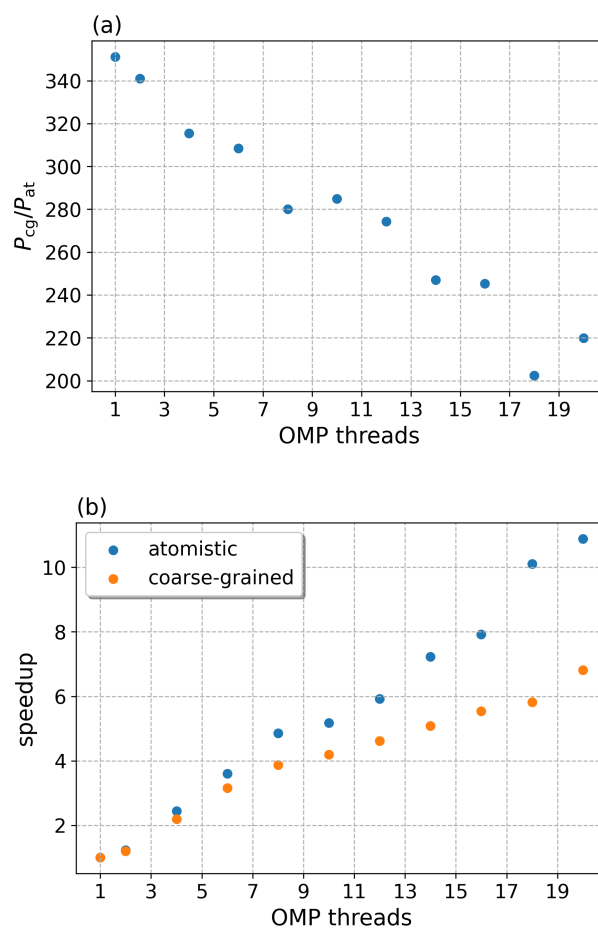

Figure 11: Performance calculations for SOM III with the highest water content, comparing atomistic and coarse-grained representation. The ratio of coarse-grained to atomistic performance is shown in (a). The speedup for different OMP Threads is depicted in (b).

15

3.3 Numerical Performance

To assess the improvement in numerical performance, measured as the length of trajectory produced within a given computational time, benchmarking simulations were performed using both atomistic and coarse-grained approaches. The performance ratio between coarse-grained and atomistic models, P_{cg}/P_{at} (Figure 11 (a)), was therefore calculated, along with the speedup of the corresponding systems (Figure 11 (b)). The speedup is defined as the ratio of serial to parallel runtime.

In this case, SOM III with the highest water content was considered. The performance, depending on the number of OpenMP (OMP) threads, is shown for the atomistic and coarse-grained systems in the Supporting Information, Figure S12, including additional information regarding the benchmark.

In Figure 11 (a), it can be seen that the performance improvement reduces as the number of OMP threads rises. This behavior can be elucidated through the speedup values in Figure 11 (b), whereby the atomistic system exhibits a higher speedup compared to the coarse-grained one, which is particularly evident with a higher number of threads. This indicates that the coarse-grained system is less efficiently parallelizable through OMP parallelization, leading to a lower enhancement in performance for more threads. This could be explained by the fact that the number of particles and therefore the number of interactions in the coarse-grained system is lower, which means that splitting the processing operations into different threads is less efficient. It is worth noting that the performance behavior may be different for MPI or hybrid parallelization methods.

However, when looking at the scale of the performance increase, the decisive advantage of the coarse-grained approach becomes apparent. It lies in the range from approximately 200 to 350, i.e., an improvement in performance in the range of two orders of magnitude can be observed. This performance improvement can be primarily attributed to two factors. First, the timestep in coarse-grained simulations is ten times larger than that in atomistic simulations. Second, for the system under consideration, each coarse-grained bead represents, on average, approximately seven atoms, including hydrogens. This reduction in the number of particles (N) decreases the computational complexity, leading to a performance improvement by a factor of approximately 49, assuming the $O(N^2)$ scaling of classical MD simulations. When combined with the tenfold increase in timestep, the total theoretical performance improvement is around a factor of 490. It is important to note, however, that performance scaling often improves slightly better than N^2 due to algorithmic optimizations, which accounts for the slightly lower performance improvement observed in this study. As demonstrated by this and by the simulations presented in this paper, the utilization of these SOM models for long-time simulations can be accomplished with reasonable effort. This enables the investigation of diverse adsorption processes on longer time scales as well as the interactions between biomacromolecules and SOM.

4 Conclusions and Outlook

The ecological importance of SOM has driven ongoing efforts to develop and refine SOM models that provide deeper insights into its processes at the molecular scale. In this paper, we introduced an improved parametrization scheme for coarse-grained SOM models, which extends our previous work[54]. This new approach enables direct conversion of atomistic VSOMM2 models to their coarse-grained counterparts, which eliminates the need for atomistic MD simulations. In addition, it provides the coarse-grained topology files in a more efficient and user-friendly manner.

We first assessed how the modified parameterization impacted the models' properties. The revised force constants reduced high vibrational frequencies within the fragment molecules, as evidenced by the power spectra, which led to higher simulation stability. Consequently, the time step could be extended to 20 fs, enhancing the performance. Importantly, the reparametrization did not significantly increase

the deviation between the bonded distributions of atomistic and coarse-grained fragments, as measured by the TMS.

To demonstrate the capability of the new protocol, we analyzed coarse-grained SOM models with varying compositions. The calculation of the potential energies of the non-bonded interactions indicated that the Coulomb interactions acted as the driving force for the packing of the SOM structures, which correlated with the acidity of the SOM systems. A subdivision of the Lennard-Jones interactions provided insight into how interactions changed with increasing water content. Exploring the larger length scales and longer simulation times revealed a local phase separation process, notably the formation of SOM voids, demonstrating the advantages of the coarse-graining approach.

For future improvements, force constants for bonds and angles could be refined by using an all-atom representation of the reference molecules, improving parameter transferability between Swarm-CG and ETKDG. Moreover, it is feasible to optimize the determination of bonded interaction parameters through the conformer search, particularly for a higher number of fragments per molecule. For example, repetitive partial sequences could be identified and considered separately during the conformer search. As the molecules are composed of a relatively large but limited number of fragments (please refer to Table S2 of the Supplementary Information of Escalona et al. [6] or Figure S1 in the Supporting Information of Dettmann et al. [54]), one could also consider sampling all possible pairs of fragments, e.g., through MD simulations, to construct a comprehensive database of equilibration constants and force constants. This would make it possible to skip the parameterization step completely. Furthermore, instead of mapping the atomistic structures, the algorithm could be extended by constructing the coarse-grained structures from the outset, to act as an independent modeler.

The present methodology allows for versatile applications of coarse-grained SOM models in simulations, including long-time simulations to dynamically study pollutant interactions and adsorption on nanoplastic particles. Moreover, model realism could be improved by incorporating biologically relevant molecules such as lipids, peptides, and carbohydrates.

Associated Content

Data Availability Statement

All topology files required to reproduce the simulations, as well as the Python tool for generating the coarse-grained SOM structures, are available in Zenodo at DOI: 10.5281/zenodo.12942690.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org.
It includes details about the modified parametrization approach and conformer search, as well as additional information about the SOM compositions and further benchmark results.

References

(1) Masoom, H.; Courtier-Murias, D.; Farooq, H.; Soong, R.; Kelleher, B. P.; Zhang, C.; Maas, W. E.; Fey, M.; Kumar, R.; Monette, M.; Stronks, H. J.; Simpson, M. J.; Simpson, A. J. Soil Organic Matter in Its Native State: Unravelling the Most Complex Biomaterial on Earth. *Environmental Science & Technology* **2016**, *50*, 1670–1680.

(2) Jackson, R. B.; Lajtha, K.; Crow, S. E.; Hugelius, G.; Kramer, M. G.; Piñeiro, G. The Ecology of Soil Carbon: Pools, Vulnerabilities, and Biotic and Abiotic Controls. *Annual Review of Ecology, Evolution, and Systematics* **2017**, *48*, 419–445.

(3) Scharlemann, J. P.; Tanner, E. V.; Hiederer, R.; Kapos, V. Global soil carbon: understanding and managing the largest terrestrial carbon pool. *Carbon Management* **2014**, *5*, 81–91.

(4) Paustian, K.; Lehmann, J.; Ogle, S.; Reay, D.; Robertson, G. P.; Smith, P. Climate-smart soils. *Nature* **2016**, *532*, 49–57.

(5) Kibblewhite, M.; Ritz, K.; Swift, M. Soil health in agricultural systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2007**, *363*, 685–701.

(6) Escalona, Y.; Petrov, D.; Oostenbrink, C. Vienna soil organic matter modeler 2 (VSOMM2). *J. Mol. Graph. Model.* **2021**, *103*, 107817.

(7) Paul, E. A. The nature and dynamics of soil organic matter: Plant inputs, microbial transformations, and organic matter stabilization. *Soil Biology and Biochemistry* **2016**, *98*, 109–126.

(8) Chen, X. D.; Dunfield, K. E.; Fraser, T. D.; Wakelin, S. A.; Richardson, A. E.; Condron, L. M. Soil biodiversity and biogeochemical function in managed ecosystems. *Soil Research* **2020**, *58*, 1.

(9) Stevenson, F. J. *Humus chemistry*, 2nd ed.; John Wiley & Sons: Nashville, TN, 1994.

(10) International Humic Substances Society. https://humic-substances.org.

(11) Kubicki, J. D., Ed. *Molecular modeling of geochemical reactions*; John Wiley & Sons: Nashville, TN, 2016.

(12) Senesi, N.; Xing, B.; Huang, P. M. *Biophysico-chemical processes involving natural nonliving organic matter in environmental systems*; Wiley Series Sponsored by IUPAC in Biophysico-Chemical Processes in Environmental Systems; Wiley-Interscience: Newy York, 2009.

(13) Schulten, H. R.; Schnitzer, M. Three-dimensional models for humic acids and soil organic matter. *Naturwissenschaften* **1995**, *82*, 487–498.

(14) Schulten, H.-R.; Schnitzer, M. Chemical Model Structures for Soil Organic Matter and Soils. *Soil Science* **1997**, *162*, 115–130.

(15) Schulten, H.-R.; Leinweber, P. New insights into organic-mineral particles: composition, properties and models of molecular structure. *Biology and Fertility of Soils* **2000**, *30*, 399–432.

(16) Schulten, H.-R. *Soil Mineral-Organic Matter-Microorganism Interactions and Ecosystem Health, Dynamics, Mobility and Transformation of Pollutants and Nutrients*; Elsevier, 2002; p 351–381.

(17) Schulten, H. R.; Plage, B.; Schnitzer, M. A chemical structure for humic substances. *The Science of Nature* **1991**, *78*, 311–312.

(18) Schulten, H. R.; Schnitzer, M. A state of the art structural concept for humic substances. *Naturwissenschaften* **1993**, *80*, 29–30.

(19) Thiele-Bruhn, S.; Seibicke, T.; Schulten, H.; Leinweber, P. Sorption of Sulfonamide Pharmaceutical Antibiotics on Whole Soils and Particle-Size Fractions. *Journal of Environmental Quality* **2004**, *33*, 1331–1342.

(20) Schwarz, J.; Thiele-Bruhn, S.; Eckhardt, K.-U.; Schulten, H.-R. Sorption of Sulfonamide Antibiotics to Soil Organic Sorbents: Batch Experiments with Model Compounds and Computational Chemistry. *ISRN Soil Science* **2012**, *2012*, 1–10.

(21) Schulten, H.-R. Analytical pyrolysis and computational chemistry of aquatic humic substances and dissolved organic matter. *Journal of Analytical and Applied Pyrolysis* **1999**, *49*, 385–415.

(22) Schulten, H. Interactions of dissolved organic matter with xenobiotic compounds: Molecular modeling in water. *Environmental Toxicology and Chemistry* **1999**, *18*, 1643–1655.

18

(23) Sutton, R.; Sposito, G.; Diallo, M. S.; Schulten, H. Molecular simulation of a model of dissolved organic matter. *Environmental Toxicology and Chemistry* **2005**, *24*, 1902–1911.

(24) 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.

(25) 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.

(26) Aquino, A. J. A.; Tunega, D.; Schaumann, G. E.; Haberhauer, G.; Gerzabek, M. H.; Lischka, H. Stabilizing Capacity of Water Bridges in Nanopore Segments of Humic Substances: A Theoretical Investigation. *The Journal of Physical Chemistry C* **2009**, *113*, 16468–16475.

(27) Ahmed, A. A.; Kühn, O.; Leinweber, P. Controlled experimental soil organic matter modification for study of organic pollutant interactions in soil. *Science of The Total Environment* **2012**, *441*, 151–158.

(28) Ahmed, A. A.; Kühn, O.; Aziz, S. G.; Hilal, R. H.; Leinweber, P. How soil organic matter composition controls hexachlorobenzene–soil-interactions: Adsorption isotherms and quantum chemical modeling. *Science of The Total Environment* **2014**, *476–477*, 98–106.

(29) Ahmed, A. A.; Leinweber, P.; Kühn, O. A computational study of hexachlorobenzene-soil organic matter-interactions. *Journal of Theoretical and Computational Chemistry* **2014**, *13*, 1450009.

(30) 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. *Sci. Total Environ.* **2015**, *508C*, 276–287.

(31) Ahmed, A. A.; Thiele-Bruhn, S.; Leinweber, P.; Kühn, O. Towards a molecular level understanding of the sulfanilamide-soil organic matter-interaction. *Science of The Total Environment* **2016**, *559*, 347–355.

(32) Gros, P.; Ahmed, A.; Kühn, O.; Leinweber, P. Glyphosate Binding in Soil as Revealed by Sorption Experiments and Quantum-Chemical Modeling. *Sci. Total Environ.* **2017**, *586*, 527–535.

(33) Ahmed, A. A.; Gros, P.; Kühn, O.; Leinweber, P. Molecular Level Investigation of the Role of Peptide Interactions in the Glyphosate Analytics. *Chemosphere* **2018**, *196*, 129–134.

(34) 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.

(35) Piccolo, A. The supramolecular structure of humic substances. *Soil Science* **2001**, *166*, 810–832.

(36) Piccolo, A. *Advances in Agronomy*; Elsevier, 2002; p 57–134.

(37) Aquino, A. J.; Tunega, D.; Pašalić, H.; Schaumann, G. E.; Haberhauer, G.; Gerzabek, M. H.; Lischka, H. Study of solvent effect on the stability of water bridge-linked carboxyl groups in humic acid models. *Geoderma* **2011**, *169*, 20–26.

(38) Schaumann, G. E.; Gildemeister, D.; Kunhi Mouvenchery, Y.; Spielvogel, S.; Diehl, D. Interactions between cations and

19

water molecule bridges in soil organic matter. *Journal of Soils and Sediments* **2013**, *13*, 1579–1588.

(39) Gerzabek, M. H.; Aquino, A. J. A.; Balboa, Y. I. E.; Galicia-Andrés, E.; Grančič, P.; Oostenbrink, C.; Petrov, D.; Tunega, D. A contribution of molecular modeling to supramolecular structures in soil organic matter. *Journal of Plant Nutrition and Soil Science* **2022**, *185*, 44–59.

(40) 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.

(41) 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.

(42) 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.

(43) 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.

(44) 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.

(45) Escalona, Y.; Petrov, D.; Galicia-Andrés, E.; Oostenbrink, C. Exploring the macroscopic properties of humic substances using modeling and molecular simulations. *Agronomy (Basel)* **2023**, *13*, 1044.

(46) 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.

(47) 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.

(48) 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.

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

(50) 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.

(51) 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.

(52) Xue, Q.; Jiao, Z.; Liu, X.; Pan, W.; Fu, J.; Zhang, A. Dynamic Behavior and Interaction Mechanism of Soil Organic Matter in Water Systems: A Coarse-Grained Molecular Dynamics Study. *Environmental Science & Technology* **2023**, *58*, 1531–1540.

(53) 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.

(54) Dettmann, L. F.; Kühn, O.; Ahmed, A. A. Martini-Based Coarse-Grained Soil Organic Matter Model Derived from Atomistic Simulations. *Journal of Chemical Theory and Computation* **2024**, *20*, 5291–5305.

(55) Riniker, S.; Landrum, G. A. Better informed distance geometry: Using what we know to improve conformation generation. *J. Chem. Inf. Model.* **2015**, *55*, 2562–2574.

(56) 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.

(57) 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. *J. Chem. Theory Comput.* **2021**, *17*, 5777–5791.

(58) 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.

(59) Dettmann, L. F.; Kühn, O.; A. Ahmed, A. Topology files and Python tool for "An automated parametrization approach for coarse-graining soil organic matter molecules". 2024; https://zenodo.org/doi/10.5281/zenodo.12942690.

(60) 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.

(61) 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.

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

(63) 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.

(64) 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.

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

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

21

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

(68) 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.

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

(70) 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.

(71) 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.

(72) Winger, M.; Trzesniak, D.; Baron, R.; van Gunsteren, W. F. On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models. *Physical Chemistry Chemical Physics* **2009**,

(73) Pereira, G. P.; Alessandri, R.; Domínguez, M.; Araya-Osorio, R.; Grünewald, L.; Borges-Araújo, L.; Wu, S.; Marrink, S. J.; Souza, P. C. T.; Mera-Adasme, R. Bartender: Martini 3 Bonded Terms via Quantum Mechanics-Based Molecular Dynamics. *Journal of Chemical Theory and Computation* **2024**, *20*, 5763–5773.

(74) Rubner, Y.; Tomasi, C.; Guibas, L. J. A metric for distributions with applications to image databases. Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271). 2002.